universite
PARIS-SACLAY

Bias correction for drift and volatility estimation of a jump diffusion and non-parametric adaptive estimation of the invariant measure

Thèse de doctorat de l'Université Paris-Saclay

Ecole Doctorale de Mathématique Hadamard (EDMH) n° 574
Spécialité de doctorat : Mathématiques appliquées
Unité de recherche : Université Paris-Saclay, CNRS, Univ Evry, Laboratoire de Mathématiques et Modélisation d'Evry, 91037, Evry-Courcouronnes, France.
Référent : Université d'Evry-Val d'Essonne

Thèse présentée et soutenue à Évry, le 2 Juillet 2020, par

Chiara AMORINO

Composition du jury :

Mathieu ROSENBAUM Professeur, École Polytechnique	Président
Alexandre BROUSTE Professeur, Le Mans Université	Rapporteur & Examinateur
Eulalia NUALART Professeur, Universitat Pompeu Fabra	Rapporteur & Examinatrice
Fabienne COMTE Professeur, Université Paris Descartes	Examinatrice
Agathe GUILLOUX Professeur, Université d'Évry	Examinatrice
Fabien PANLOUP Professeur, Université d'Angers	Examinateur
Arnaud GLOTER Professeur, Université d'Évry	Directeur de Thèse

Thèse de doctorat

NNT : 2020UPASE006

université
evry
val-d'essonne

LaMME

Fondation mathématique
FMJH
Jacques Hadamard

Remerciements

Je profite de ces premières pages pour remercier les personnes sans qui ce manuscrit n'aurait pas été possible.

Je voudrais tout d'abord remercier mon directeur de thèse Arnaud Gloter pour m'avoir constamment aidée et soutenue pendant ces trois années. Il a toujours été extrêmement gentil: toujours prêt à expliquer, à m'écouter et à me donner beaucoup de conseils (pas seulement à propos des maths). Je ne sais pas comment j'aurais fait sans lui: j'ai vraiment eu de la chance à avoir ce brillant mathématicien comme directeur de thèse.

Je suis très reconnaissante aux rapporteurs, Alexandre Brouste et Eulalia Nualart, d'avoir consacré un temps certain à la lecture de mon (long) manuscrit.

Je remercie également Fabienne Comte, Agathe Guilloux, Fabien Panloup et Mathieu Rosenbaum qui m'ont fait l'honneur d'accepter de faire partie de mon jury.

Mes années doctorales ont aussi été l'occasion de diriger des TD. Je remercie Christophe Ambroise, Dasha Loukianova, Marco Pascucci, Abass Sagne, Marie-Luce Taupin et Anne-Sophie Tocquet de m'avoir aidé d'un point de vue pédagogique.

Lors de ce doctorat, j'ai fait beaucoup de belles rencontres au LaMME. Je souhaite remercier, en particulier, Valérie Gontier-Picot pour sa disponibilité et sa gentillesse et El Maouloud Ould Baba pour sa bonne humeur (contagieux)!

At the lab I also met some amazing PhD students (past and present) with whom I shared my coffee addiction. Thank you Oscar, Elizabeth, Pedro, Antonello, Edmond, Xavier, Ludivine, Halaleh and Elisabetta: I really enjoyed the coffee breaks and the lunches we had all together! I want to thank Lorenzo as well for his (many) advices about both mathematical and personal stuff and for all the good time we had together around the world.

I want to thank Kevin and Myra. Sometimes it is impossible to see each other for a longtime, but then we organize some amazing weekend somewhere and it seems like time has never passed. Hanging out with you is always wonderful.

Cambio lingua di nuovo per ringraziare i miei amici di vecchia data: coloro che, nonostante la distanza, sono sempre stati presenti.
Grazie a Jaco, Grazia, Dario e Fra per tutte le risate fatte insieme: passare del tempo con voi e non spaccarsi dal ridere è impossibile. Anche perché se Gra, a casa

sua, si mette a ridere... Si sente anche qui a Paris, facile.

Grazie alle mie ragazze (Alice, Giulia e Silvia) e a Penghino. Essere in università senza di voi non é la stessa cosa! Grazie per tutti i consigli e grazie per la vostra vicinanza, nonostante i km di distanza.

Grazie alla Ila, perché uscire con lei é una garanzia. Inizi ridendo, continui parlando di qualsiasi cosa e finisci ridendo di nuovo, dopo aver aperto mille frasi subordinate e aver perso 42 volte il filo del discorso.

Ragazzi, vi voglio bene!

Grazie alla mia famiglia, a mio padre in particolare, per aver sempre capito e appoggiato le mie scelte nonostante spesso non fosse affatto semplice.

Grazie alla mia famiglia acquisita: Pasquale, Davide, Lucia e Franci. Grazie, per avermi sempre trattato incredibilmente bene, per avermi sempre messo completamente a mio agio. Mi fate davvero sentire parte integrante della vostra fantastica famiglia.

Grazie al mio amore, perché senza di lui niente di tutto questo sarebbe stato possibile. Grazie per essere sempre stato presente, attento, divertente, comprensivo. Non hai idea di quanto tu mi renda felice. Ti amo da morire.

Contents

Introduction

Dans toutes les situations liées à des événements incertains nous prenons des décisions qui sont basées sur l'inférence statistique, même si nous n'en sommes pas toujours conscients. Du moment qu'une telle inférence est faite a partir des observations du même phénomène dans le passé, on peut recueillir des données et construire un modèle qui représente le phénomène que l'on considère, l'utiliser après pour estimer ce qu'on ne connaît pas et arriver comme cela a prendre des décisions plus pondérées. Dans cette procédure, plus sont nombreuses les données à notre disposition et plus soigneusement notre modèle peut prédire le futur et nous aider a faire les bons choix. L'inférence asymptotique est le domaine qui étudie les procédures d'inférence et les propriétés des estimateurs quand la taille de l'échantillon tend vers l'infini.
Nous nous intéressons, dans cette thèse, à l'inférence asymptotique pour des processus stochastiques qui suivent des équations différentielles stochastiques avec sauts.

Dans l'histoire de l'inférence des processus stochastiques, les équations différentielles stochastiques sans sauts ont retenu l'attention de beaucoup de statisticiens car elles ont été largement utilisées comme modèle pour les applications. Par exemple, elles modélisent les prix des obligations dans les marchés financièrs dans [69], [1], [74], [85] et [40]. On trouve aussi des applications aux modèles de risque en assurance dans [37], [28], [30]; à l'hydrologie dans [16] et aux modèles des populations dans [43] et [41].

Récemment, les EDS avec sauts sont devenues un outil également puissant pour la modélisation de divers phénomènes stochastiques dans de nombreux domaines comme la physique, la biologie, les sciences médicales, sociales et économiques. En finance, les processus à sauts ont été introduits pour modéliser la dynamique des taux de change dans [11], des prix des actifs dans [70],[56] et des processus de volatilité dans [8],[31]. Des utilisations des processus à sauts dans la neuroscience peuvent être trouvés par exemple dans [26]. En conséquence, l'étude de ce modèle à partir de différentes types des données est un problème qui attire, à présent, beaucoup d'attention.

Cette thèse se compose de trois parties qui portent sur l'étude du modèle de diffusion à sauts suivant:

$$X_t = X_0+\int_0^t b(X_s)ds+\int_0^t a(X_s)dW_s+\int_0^t \int_{\mathbb{R}^d\backslash\{0\}} \gamma(X_{s^-})z(\mu-\bar{\mu})(ds,dz), \quad t \in [0,T], \tag{1}$$

où $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ et $\gamma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$; $W = (W_t, t \geq 0)$ est un mouvement Brownien d- dimensionnel et μ est une mesure ponctuelle de Poisson sur $(0,\infty) \times \mathbb{R}^d$ associée au processus de Lévy $L = (L_t)_{t\geq 0}$, avec $L_t := \int_0^t \int_{\mathbb{R}^d} z(\mu(ds,dz) - \bar{\mu}(ds,dz))$. Les conditions sur la mesure de Lévy seront données dans la suite et seront spécifiques dans chacune des trois parties dont ce travail est composé.

On rappelle la définition d'un processus de Lévy; pour des détails supplémentaires on peut se reporter aux travaux de Sato [82] ou Applebaum [7].

Definition 1. *Un processus stochastique càdlàg $L = (L_t)_{t\geq 0}$ défini sur un espace de probabilité filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ est dit être un processus de Lévy si il a les propriétés suivantes:*

1. *$L_0 = 0$ presque sûrement.*

2. *Les accroissement de L sont indépendants. Cela veut dire que que, pour chaque choix de $n \geq 1$ et $0 \leq t_0 < t_1 < \ldots < t_n$, $(L_{t_j} - L_{t_{j-1}}, 1 \leq j \leq n)$ sont indépendants.*

3. *Les accroissements de L sont stationnaires. C'est à dire que, pour chaque $0 \leq s < t < \infty$, la distribution de $L_t - L_s$ est égal à L_{t-s}.*

Le processus de Lévy est donc un processus avec accroissements indépendants et stationnaires dont les trajectoires sont continues à droite et limitées à gauche. La stationnarité et l'indépendance des accroissements impliquent que le processus de Lévy est aussi un processus de Markov homogène.
Dans la suite, grande importance sera aussi donnée à α, l'indice d'activité des sauts ou paramètre indice d'un processus de Lévy, qui est un paramètre non - aléatoire qui ne dépend pas du temps.

Definition 2.

$$\alpha := \inf \left\{ r \in [0,2] : \int_{|x|\leq 1} |x|^r F(dx) < \infty \right\},$$

où $F(dx)$ est la mesure du Lévy.

L'indice d'activité des sauts est l'indice de Blumenthal -Getoor qui avait justement été introduit d'abord par Blumenthal et Getoor dans [15]. L'intérêt dans l'identification de α réside dans le fait que il classifie le processus de Lévy selon le degrée d'activité des sauts: quand α augmente de 0 à 2, le petits sauts tendent à être de plus en plus fréquents.

Avant de commencer notre étude il nous semble convenable de donner des conditions pour garantir l'existence et l'unicité d'un processus qui soit solution de (1). À ce sujet, on cite les Théorèmes 6.2.9 et 6.4.6 dans [7] selon lesquels il est suffisant que les coefficients soient globalement Lipschitz pour garantir l'existence et l'unicité d'une solution càdlàg adaptée qui possède la propriété de Markov forte.
En conséquence, on demandera toujours dans la suite la Lipschitizianité globale des fonctions susmentionnées et quasiment toujours des conditions qui assurent que le processus X soit ergodique. L'ergodicité du processus, d'ailleurs, joue en inférence stochastique un rôle essentiel dans l'étude du comportement asymptotique des estimateurs.
En général les théorèmes ergodiques pour processus de Markov sont décrits grâce à l'existence d'une limite en probabilité de la moyenne dans le temps: $\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t)dt$, où X est un processus stochastique et f est une fonction mesurable.

Précisons les notions d'ergodicité et stationnarité du processus. Soit X un processus solution de (1) et $p_t(x, A)$ la probabilité de transition définie comme

$$p_t(x, A) := \mathbb{P}(X_{s+t} \in A | X_s = x),$$

où A est un ensemble de Borel dans $\mathbb{R}^d$ et $t, s \geq 0$. Par l'homogénéité de X la probabilité $p_t(x, A)$ ici dessus ne dépend pas de s et du coup la distribution de X est déterminée uniquement par le semi-groupe $(p_t)_{t\geq 0}$ et la loi de X_0. En particulier, si π est la distribution de probabilité de la valeur initiale X_0, alors $\int_{\mathbb{R}^d} p_t(x, A)\pi(dx)$ est la distribution de probabilité de X_t pour chaque $t > 0$.

La distribution π est dite invariante si et seulement si on a $\pi(A) = \int_{\mathbb{R}} p_t(x, A)\pi(dx)$ pour n'importe quel $t \in \mathbb{R}^+$ et A ensemble borélien. De plus, on remarque que si la distribution initiale π est invariante, alors la distribution de X_t pour $t > 0$ est toujours π, cela veut dire que X est stationnaire. Dans la suite nous utiliserons toujours le mot « stationnaire » dans ce sens. Une telle distribution de probabilité π est aussi dite distribution stationnaire.

L'existence d'une mesure invariante pour un processus de Markov est essentielle dans la théorie ergodique puisque la définition d'ergodicité est la suivante:

Definition 3. *Soit X un processus solution de* (1)*. Le processus X est ergodique si et seulement si il existe une mesure de probabilité invariante π telle que*

$$\frac{1}{T}\int_0^T f(X_t)dt \xrightarrow{\mathbb{P}} \int_{\mathbb{R}^d} f(x)\pi(dx) \qquad \textit{pour } T \to \infty,$$

pour chaque fonction f définie sur $\mathbb{R}^d$ et integrable par rapport a π.

Il est normalement difficile de vérifier l'ergodicité d'un processus avec sauts. Cependant, on peut trouver des conditions suffisantes pour que le processus soit ergodique.

Meyn et Tweedie ont donné dans [71] et [72] une théorie ergodique pour des processus de Markov généraux; en appliquant leur théorie aux équations différentielles stochastiques avec sauts, Masuda a fourni dans [66], [67] des conditions plus explicites pour obtenir l'ergodicité. Par exemple, l'irreducibilité, le critère de Foster - Lyapunov, la stationnarité et des conditions sur les moments par rapport à la mesure invariante π nous donnent l'ergodicité exponentielle, qui est une propriété plus forte de celle définie dans la Définition 3.
En utilisant les critères introduits par Masuda, il est possible de donner des exemples de diffusion avec sauts qui soient ergodiques. Parmi eux, il y a le processus d'Ornstein-Uhlenbeck qui est l'un des principaux modèles pour les applications. Ce modèle sera par ailleurs employé pour illustrer numériquement les résultats de la première partie de cette thèse, qui concerne l'estimation paramétrique des paramètres de dérive et de volatilité en utilisant une fonction de contraste.

0.1 Un bref résumé de la thèse

Dans la première partie de la thèse, en particulier, on suppose observer le processus $(X_{t_i})_{i=0,\dots n}$, où le pas $\Delta_n := \max_i(t_{i+1} - t_i)$ converge vers 0; X est le processus solution de l'EDS avec sauts (1) pour $d = 1$ avec les coefficients de dérive et volatilité

qui dépendent de deux paramètres inconnus que nous dénoterons μ et σ, respectivement.
La première question posée dans la thèse porte sur l'amélioration des résultats existants en littérature, en proposant de fonctions de contraste qui puissent enlever les conditions restrictives présentes sur le pas d'observation de la trajectoire.
Shimizu utilisait dans [87] une fonction de contraste issue du schéma d'Euler pour l'estimation des paramètres de dérive, de volatilité et des sauts. La normalité asymptotique des estimateurs était obtenu sous des conditions reliant la vitesse à laquelle $\Delta_n \to 0$ à l'intensité des sauts au voisinage de 0. Ces conditions sur Δ_n étaient de plus en plus restrictives lorsque l'intensité des sauts en zéro était haute. Dans la situation la plus favorable, correspondante à une intensité des sauts finie, la condition était $n\Delta_n^2 \to 0$ et, lorsque α s'approchait de 1, cela finissait par être $n\Delta_n \to 0$, en contradiction avec $n\Delta_n \to \infty$.
Dans [38] la condition sur le pas est affaiblie, pour l'estimation de la dérive seule, et devient en particulier $n\Delta_n^3 \to 0$ en intensité finie.
Ces deux références font l'hypothèse de sauts sommables ($\alpha \leq 1$). D'ailleurs les conditions exigées sur le pas sont restrictives et, pour $\alpha > 1$, rentrent formellement en contradiction avec $n\Delta_n \to \infty$.
En considérant un modèle sans saut pour obtenir des conditions moins restrictives que $n\Delta_n^3 \to 0$ il est nécessaire d'introduire des corrections de la fonction de contraste issue du schéma d'Euler comme cela est fait dans Kessler [51].

Dans le premier chapitre il est proposé une correction du contraste de Shimizu [87] qui permet d'estimer le paramètre de dérive, sans nécessiter de conditions sur la vitesse à laquelle Δ_n converge vers 0. Nous étendons aussi les résultats de [88] et [38] en supprimant l'hypothèse de ces deux articles qui impose que la mesure μ ait des sauts sommables au voisinage de 0 (i. e. $\alpha < 1$). Dans le cas où l'intensité de saut est finie, nous sommes capables de proposer une correction explicite du contraste de Shimizu et la relions à la correction de Kessler.

Le second chapitre est dédié à l'estimation jointe du paramètre de diffusion et dérive dans un cadre similaire et sous la condition d'une intensité de saut finie. L'estimation jointe des deux paramètres introduit des difficultés notables : en particulier comme les deux paramètres ne s'estiment pas à la même vitesse, l'étude asymptotique de la fonction de contraste implique deux régimes asymptotiques différents. Par rapport aux résultats antérieurs (voir [88]), nous montrons qu'il est possible d'estimer conjointement les paramètres μ et σ sans condition de vitesse sur la décroissance du pas d'observation. Nous traitons aussi le cas d'observations non régulièrement espacées ce qui, à notre connaissance, n'avait jamais été fait pour l'estimation jointe de la dérive et volatilité d'une diffusion.

La seconde partie de la thèse étudie l'estimation de la volatilité intégrée, qui est un problème important en finance. Lorsque le modèle a des sauts, une des méthodes utilisée est de considérer la variation quadratique, où l'on supprime les accroissements au-dessus d'un seuil dont on pense qu'il est significatif de la présence d'un saut macroscopique. Dans le cas où la partie à saut de l'EDS admet un indice de Blumenthal Getoor $\alpha > 1$, avec α défini par (2), il est montré dans [47] que la vitesse d'estimation se dégrade et d'autres méthodes d'estimations que la variation quadratique tronquée sont proposées (e.g. [49]).
Dans ce chapitre nous considérons le cas d'EDS du type (1) avec $d = 1$ dont la partie saut est de type Stable avec indice α. En utilisant le calcul de Malliavin

nous étendons les travaux de [65] en caractérisant précisément le biais introduit par la présence des sauts dans la variation quadratique tronquée. Nous sommes alors capables de modifier l'estimateur pour réduire ce biais et démontrons que la vitesse d'estimation ne se dégrade plus toujours pour $\alpha > 1$.
Sur des simulations, nous montrons que notre méthode permet effectivement de réduire considérablement les biais et que nos estimateurs de la volatilité intégrée fonctionnent même pour des indices d'activité des sauts supérieurs à 1.

La troisième partie traite de l'estimation adaptative de la mesure stationnaire. Nous considérons la solution X de l'EDS multidimensionnelle avec sauts proposée dans (1), avec une unique mesure de probabilité invariante et une densité associée. Nous supposons qu'un enregistrement continu des observations $X^T = (X_t)_{0 \le t \le T}$ est disponible.
Dalalyan et Reiss en [25] et Strauch en [90] ont caractérisé la vitesse minimax pour l'estimation de la loi stationnaire d'une diffusion continue en dimension d dans les cas isotropique et anisotropique, respectivement. Cette vitesse dépend de la dimension d et de la régularité de la mesure stationnaire.
Nous étendons ces travaux en obtenant, dans le cadre d'un processus avec sauts, des estimateurs qui ont la même vitesse que dans le cas d'une diffusion continue pour $d \ge 2$ et une vitesse qui dépend de l'intensité des sauts α dans le cas 1 - dimensionnel.
Nous proposons par ailleurs une procédure de sélection de la fenêtre pour un estimateur à noyau basée sur le méthode introduit par Goldenshluger et Lepski dans [39], qui nous conduit à un estimateur non -paramétrique et adaptatif de la densité stationnaire de la diffusion multivariée avec sauts X.

0.2 Première partie: correction de fonctions de contraste pour l'estimation paramétrique des coefficients.

Dans la première partie de la thèse nous nous intéressons à l'estimations paramétrique de $\theta = (\mu, \sigma)$ à partir d'un échantillonage discret du processus X^θ solution de l'équation différentielle stochastique avec sauts suivante:

$$X_t^\theta = X_0^\theta + \int_0^t b(X_s, \mu)ds + \int_0^t a(X_s, \sigma)dW_s + \int_0^t \int_{\mathbb{R}} \gamma(X_{s^-})z\tilde{\mu}(ds, dz) \quad t \ge 0; \quad (2)$$

où W est un mouvement brownien 1 - dimensionnel et $\tilde{\mu} := \mu - \bar{\mu}$ est une mesure ponctuelle de Poisson sur $(0, \infty) \times \mathbb{R}$ associée au processus de Lévy $L = (L_t)_{t \ge 0}$.
Nous supposons que le processus est échantillonné à des instants t_i^n, avec i qui va de 0 jusqu'à n, où le pas de discrétisation $\Delta_n := \sup_{i=0,...,n-1} t_{i+1}^n - t_i^n$ va à 0; dans la suite on appellera ces instants de temps simplement t_i en n'explicitant plus la dépendance en n pour alléger la notation.
Pour les applications, un point crucial dans le cas de l'observation d'haute fréquence est d'imposer des conditions sur Δ_n qui soient les minimales possibles; un des objectifs principaux de la première partie de la thèse sera effectivement d'estimer θ avec un pas de discrétisation qui peut être irrégulier et sous des conditions beaucoup moins restrictives par rapport à celles existantes dans la littérature précédente.

Il est connu que, comme conséquence de la présence d'une composante Gaussienne, il est impossible d'estimer le paramètre μ sur un horizon temporel fini; nous supposons donc que $t_n \to \infty$ pour n qui tend vers l'infini et que le processus X^θ est ergodique (voir Définition 3).

En considérant le cas continu, où le processus X est solution de

$$X_t^\theta = X_0^\theta + \int_0^t b(X_s, \mu)ds + \int_0^t a(X_s, \sigma)dW_s, \tag{3}$$

le problème de l'estimation de θ a déjà été étudié par des nombreux auteurs comme Florens - Zmirou, qui a introduit dans [33] un estimateur pour les deux paramètres μ et σ sous la condition restrictive $n\Delta_n^2 \to 0$, Prakasa - Rao ([78], [79]) et Yoshida ([95]).

Une des difficultés principales est que la densité de transition du processus X est inconnue et, en conséquence, on ne dispose pas de la fonction de vraisemblance non plus. Donc l'estimateur de maximum vraisemblance, qui possède les bonnes propriétés habituelles (voir Dachuna - Castelle et Florens - Zmirou [24]), n'est pas une solution en pratique.

Une voie commune pour surmonter cette difficulté pour l'estimation de μ est de baser l'inférence sur la discrétisation de la fonction de vraisemblance continue, voir par exemple Yoshida [95] et Genon - Catalot [35].

Supposons que σ soit connu dans (3) et dénotons avec Q_σ la loi du processus solution de $dY_t = a(Y_t, \sigma)dW_t$; si nous disposons de l'entière trajectoire du processus jusqu'au temps t_n, alors la fonction de log - vraisemblable continue avec mesure de référence Q_σ est

$$\int_0^{t_n} \frac{b(X_t, \mu)}{a^2(X_t, \sigma)} dX_t - \frac{1}{2} \int_0^{t_n} \frac{b^2(X_t, \mu)}{a^2(X_t, \sigma)} dt.$$

Une discrétisation de l'équation ci - dessus nous donne la fonction de contraste

$$\sum_{i=0}^{n-1} [\frac{b(X_{t_i}, \mu)}{a^2(X_{t_i}, \sigma)}(X_{t_{i+1}} - X_{t_i}) - \frac{1}{2} \frac{b^2(X_{t_i}, \mu)(t_{i+1} - t_i)}{a^2(X_{t_i}, \sigma)}]. \tag{4}$$

Une autre manière pour construire une fonction contraste et contourner la méconnaissance de la fonction de vraisemblance a été introduit par Florens - Zmirou dans [33]. Il consiste dans l'utilisation d'une schéma d'approximation en temps discret connu comme schéma d'Euler - Maruyama. Florens - Zmirou considère le cas où $a(., \sigma) = \sigma$ et elle approxime le processus

$$X_{t_{i+1}} - X_{t_i} = \int_{t_i}^{t_{i+1}} b(X_s, \mu)ds + \sigma \int_{t_i}^{t_{i+1}} dW_s$$

avec le modèle obtenu en le discrétisant:

$$b(X_{t_i}, \mu)\Delta_{n,i} + \sigma(W_{t_{i+1}} - W_{t_i}),$$

où nous avons noté $\Delta_{n,i} := (t_{i+1} - t_i)$.

Cette approximation conduit l'auteur à considérer une approximation localement Gaussienne de la densité de transition, i. e. la loi de $\mathcal{L}(X_{t_{i+1}}|X_{t_i})$ est approchée par $N(b(X_{t_i}, \mu)\Delta_{n,i}, \sigma^2\Delta_{n,i})$ et la fonction de vraisemblance approximée de l'échantillon $(X_{t_i})_{i=0,...,n}$, appelée quasi-vraisemblance, devient

$$\frac{1}{2} \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - X_{t_i} - b(X_{t_i}, \mu)\Delta_{n,i})^2}{\sigma^2 \Delta_{n,i}} + n \log(\sigma^2). \tag{5}$$

Nous remarquons que la quasi- log vraisemblance donnée ci-dessus coincide, pour $a(.,\sigma) = \sigma$, à la discrétisation de la fonction de log - vraisemblable continue donnée dans (4), à une constante multiplicative et une variable aléatoire additive qui ne dépend pas de μ près. De plus, l'estimateur qui minimise la fonction de quasi- log vraisemblance proposée dans (5) avait aussi été étudiée par Prakasa -Rao dans [78] comme estimateur des moindres carrés pour μ.
L'estimation jointe des paramètres (μ, σ), toujours dans le cas continu, a été étudiée autant par Yoshida [95] dans le cadre d- dimensionnel que par Florens -Zmirou [33]; dans les deux cas le coefficient $a(.,\sigma)$ était multiplicatif, c'est à dire que $a(x,\sigma) = \sigma a(x)$.
Cela leur a permis de proposer un estimateur de μ trouvé en minimisant la fonction de contraste (5) et un estimateur de σ fondé sur la variation quadratique. Ensuite ils ont dû imposer des conditions plutôt restrictives sur la vitesse à laquelle le pas de discrétisation devait aller à zéro: dans Florens -Zmirou [33] Δ_n devait satisfaire $n\Delta_n^2 \to 0$ alors que Yoshida, à travers des corrections du contraste (5) a changé cette condition dans la moins restrictive $n\Delta_n^3 \to 0$. Sous ces conditions les estimateurs qu'ils proposent pour μ sont asymptotiquement efficaces.
Kessler présente en [51] une fonction contraste pour l'estimation jointe des paramètres μ et σ. Pour la construire il veut utiliser, comme dans Florens -Zmirou [33], une approximation localement Gaussienne de la densité de transition; la plus naturelle est obtenue en choisissant comme moyenne et variance de la Gaussienne la moyenne et la variance de la densité de transition. C'est a dire que, après avoir défini

$$m(\mu, \sigma, x) := \mathbb{E}[X_{t_{i+1}}^\theta | X_{t_i}^\theta = x] \quad \text{et} \tag{6}$$

$$m_2(\mu, \sigma, x) := \mathbb{E}[(X_{t_{i+1}}^\theta - m(\mu, \sigma, X_{t_i}^\theta))^2 | X_{t_i}^\theta = x],$$

Kessler approxime la densité de transition avec la densité de $N(m(\mu, \sigma, x), m_2(\mu, \sigma, x))$ et du coup il considère le contraste

$$\sum_{i=0}^{n-1} [\frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{m_2(\mu, \sigma, X_{t_i})} + \log(m_2(\mu, \sigma, X_{t_i}))]. \tag{7}$$

Il montre, sans ajouter des conditions sur la forme du coefficient a, que l'estimateur trouvé en minimisant telle fonction de contraste est asymptotiquement efficace sous la condition générale $n\Delta_n^p \to 0$ pour un nombre entier arbitraire $p \geq 2$.
Les quantités (6) n'étant pas explicites, Kessler propose aussi un développement explicite à l'ordre $\Delta_{n,i}^q$, avec q nombre entier arbitraire, telle que l'approximation de la fonction contraste (7) conduit à un estimateur efficace toujours sous la conditions générale $n\Delta_n^p \to 0$ pour un nombre entier arbitraire $p \geq 2$.

Quand une composante de sauts est ajoutée, moins des résultats sont connus. Shimizu étudie en [87] l'estimation paramétrique des trois coefficients: dérive, volatilité et sauts en montrant la normalité asymptotique des estimateurs sous des conditions explicitement liées au pas de discrétisation et à l'intensité des sauts du processus. Plus l'intensité des sauts en zéro est haute, plus ces conditions sur Δ_n sont restrictives; dans le cas d'intensité finie la condition présente dans [87] devient $n\Delta_n^2 \to 0$.
Dans [38] la condition sur le pas de discrétisation est relâchée et devient par exemple $n\Delta_n^3 \to 0$ pour une intensité de sauts finie et pour la seule estimation du paramètre de dérive μ. Aucun des deux travaux [87] [38] ne traite le cas d'intensité $\alpha > 1$. Pour

obtenir des conditions moins restrictives que $n\Delta_n^3 \to 0$ il est nécessaire d'introduire des corrections de la fonction du contraste issue du schéma d'Euler, comme cela est fait dans Kessler [51].

Dans ce but, en remarquant d'ailleurs qu'en présence de sauts les fonctions de contraste proposées dans [38], [87] et [88] sont toujours obtenues à partir d'une procédure de filtrage qui a comme objectif de supprimer la contribution des sauts et de récupérer la partie continue du processus, nous introduisons la suivante fonction de contraste:

$$U_n(\mu,\sigma) := \sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2}{m_2(\mu,\sigma,X_{t_i})} + log(\frac{m_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}, \tag{8}$$

où la fonction φ est une version lisse de la fonction indicatrice qui s'annule quand les accroissements des données sont trop grands comparés aux accroissements typiques d'une diffusion continue et peut donc être utilisée pour filtrer la contribution des sauts.

L'idée à la base est la suivante: dès que les données observées sont discrètes, il faut décider si des sauts ont eu lieu ou pas dans un particulier interval $[t_i, t_{i+1}]$ en observant seulement l'accroissement $X_{t_{i+1}} - X_{t_i}$, même s'il s'agit en réalité d'une décision stochastique qui pourrait inclure parfois des erreurs de jugement. Ce critère devrait dépendre de n et être tel que plus grand est n et meilleure est la précision de l'évaluation concernant la présence de sauts dans l'interval considéré.

Le critère que nous considérerons est le suivant: pour $\beta \in [0, \frac{1}{2})$, si l'accroissement dépasse Δ_n^β alors nous estimons qu'au moins un saut s'y est produit, autrement nous jugeons l'interval concerné comme un interval dans lequel on n'a pas eu des sauts. Le motif est que l'accroissement d'une diffusion continue dépasse Δ_n^β avec une probabilité petite alors que un accroissement d'une diffusion avec même un seul saut dépasse Δ_n^β avec une grande probabilité. Bien que cela soit un raisonnement intuitif, ce critère est justifié par les Lemmes 3.2 et 3.3 dans Shimizu [88], où est calculé la probabilité d'avoir 0, 1 et 2 ou plus sauts dans les deux cas $X_{t_{i+1}} - X_{t_i} \le \Delta_n^\beta$ et $X_{t_{i+1}} - X_{t_i} > \Delta_n^\beta$.

La valeur β doit être choisie avec attention. Si par exemple β est trop grand (et du coup Δ_n^β trop petit), la probabilité d'obtenir un accroissement supérieur à Δ_n^β pour une diffusion continue ne peut pas être ignoré. Par contre si β est trop petit (et donc Δ_n^β trop grand), la probabilité d'obtenir un accroissement qui soit plus petit que Δ_n^β même si un saut s'est produit n'est pas négligeable.

La présence de la dernière fonction indicatrice dans (2.2) est technique. Le but est d'éviter la possibilité que $|X_{t_i}|$ soit trop grand. La constante k est positive et sera choisi dans la suite, en relation au développements de m et m_2 qui ne sont obtenus que pour $|x| \le \Delta_{n,i}^{-k}$.

Les quantités m et m_2 qui apparaissent dans (2.2) sont l'extension naturelle des quantités proposées par Kessler dans [51].

Nous les avons en effet défini de la façon suivante:

$$m(\mu,\sigma,x) := \frac{\mathbb{E}[X_{t_{i+1}}^\theta \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} \quad \text{et}$$

$$m_2(\mu,\sigma,x) := \frac{\mathbb{E}[(X_{t_{i+1}}^\theta - m(\mu,\sigma,X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}.$$

Du moment que la densité de transition n'est pas connue, on ne dispose pas d'expression pour m et m_2 et alors la fonction de contraste (2.2) n'est pas explicite non plus. L'étude de la fonction de contraste (2.2) et de ses dérivées par rapport aux paramètres, nécessaire à l'étude du comportement asymptotique de l'estimateur, reposera alors sur des approximations explicites de m, m_2 et de leurs dérivées jusqu'au troisième ordre.

En particulier dans le premier chapitre dont est composé cette partie de la thèse nous ne nous intéressons que à l'estimation du paramètre de dérive et nous nous concentrons sur les développements de m et de ses dérivées par rapport au seul paramètre de dérive.
Le processus X que nous considérons est solution de l'EDS suivante:

$$X_t^\mu = X_0^\mu + \int_0^t b(\mu, X_s^\mu) ds + \int_0^t a(X_s^\mu) dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}^\mu) z \tilde{\mu}(ds, dz), \quad t \in \mathbb{R}_+, \tag{9}$$

avec une mesure de Lévy F qui est telle que $\exists c > 0$ tel que, pour chaque $z \in \mathbb{R}$, $F(z) \le \frac{c}{|z|^{1+\alpha}}$, avec $\alpha \in (0, 2)$.
D'ailleurs la fonction de contraste que nous utilisons dans ce chapitre est plus simple que (2.2) et est la suivante:

$$U_n(\mu) := \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m(\mu, X_{t_i}))^2}{a^2(X_{t_i}) \Delta_{n,i}} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}. \tag{10}$$

A partir de cette fonction de contraste nous définissons l'estimateur $\hat{\mu}_n$ de μ_0, la vraie valeur du paramètre, comme $\hat{\mu}_n := arg \min_\mu U_n(\mu)$.
Les résultats principaux de ce chapitre sont les suivants:

Résultat 1. *L'estimateur $\hat{\mu}_n$ est consistent en probabilité:*

$$\hat{\mu}_n \xrightarrow{\mathbb{P}} \mu_0, \qquad n \to \infty.$$

Nous rappelons par ailleurs que l'information de Fisher est définie par $I(\mu) = \int_{\mathbb{R}} \frac{(\dot{b}(\mu,x))^2}{a^2(x)} \pi^\mu(dx)$. Nous pouvons maintenant donner un autre résultat principal de ce chapitre:

Résultat 2. *L'estimateur $\hat{\mu}_n$ est asymptotiquement normal:*

$$\sqrt{t_n}(\hat{\mu}_n - \mu_0) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\mu_0)), \qquad n \to \infty.$$

De plus, l'estimateur $\hat{\mu}_n$ est asymptotiquement efficace dans le sens du théorème de convolution de Hájek-Le Cam. Le théorème de convolution de Hájek-Le Cam dit que n'importe quel estimateur régulier dans un modèle paramétrique qui satisfait la propreté LAN est asymptotiquement équivalent à la somme de deux variables aléatoires indépendantes, dont une est gaussienne avec variance égale à l'inverse de l'information de Fisher et l'autre possède une distribution arbitraire. Les estimateurs efficaces sont ceux qui ont la seconde composante égale à zéro.

Le modèle (9) est LAN avec information de Fisher $I(\mu) = \int_{\mathbb{R}} \frac{(\dot{b}(\mu,x))^2}{a^2(x)} \pi^{\mu}(dx)$ (voir [54]) et donc $\hat{\mu}_n$ est efficace.

Donc, les résultats principaux de ce chapitre sont la consistence et l'efficacité asymptotique de l'estimateur que nous montrons sans devoir ajouter des conditions supplémentaires sur le pas de discrétisation. En comparaison avec la littérature précédente (voir [87] [88] et [38]), le pas d'observation $\Delta_{n,i}$ peut être irrégulier, nous n'avons pas besoin des conditions sur la vitesse à laquelle $\Delta_n \to 0$ et nous avons supprimé l'hypothèse de ces articles qui imposaient que la mesure μ eût des sauts sommables au voisinage de 0. Nous soulignons que, quand l'activité des sauts est haute au point que les sauts ne sont plus sommables, nous devons choisir $\beta < \frac{1}{3}$ (voir Assumption $A\beta$ dans la suite).

Par contre notre fonction de contraste repose sur le quantité m qui n'est pas explicite en générale. Cependant, nous trouvons des développements asymptotiques pour m (voir Résultat 3 ci-dessous).

Dans la suite, pour $\delta \geq 0$, nous dénoterons comme $R(\mu, \Delta_{n,i}^{\delta}, x)$ n'importe quelle fonction $R(\mu, \Delta_{n,i}^{\delta}, x) = R_{i,n}(\mu, x)$ où $R_{i,n}$ est telle que

$$\exists c > 0 \qquad |R_{i,n}(\mu, x)| \leq c(1 + |x|^c)\Delta_{n,i}^{\delta} \tag{11}$$

uniformément en μ et avec c indépendant de n.

La fonction R représente le terme de reste. Les cas $\alpha < 1$ et $\alpha \geq 1$ nous donnent deux magnitudes différentes pour le terme de reste dans les développements de m.

Résultat 3. • *Soit $\alpha \in (0, 1)$. Il existe $k_0 > 0$ tel que, pour $|x| \leq \Delta_{n,i}^{-k_0}$,*

$$m(\mu, x) = x + \Delta_{n,i}\, b(x, \mu) + \tag{12}$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x) \left[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)\right] F(z)dz \; + \; R(\mu, \Delta_{n,i}^{2-2\beta}, x).$$

• *Soit $\alpha \in [1, 2)$. Il existe $k_0 > 0$ tel que, pour $|x| \leq \Delta_{n,i}^{-k_0}$,*

$$m(\mu, x) = x + \Delta_{n,i}\, b(x, \mu) +$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x) \left[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)\right] F(z)dz \; + \; R(\mu, \Delta_{n,i}^{2-3\beta}, x).$$

La constante k dans la définition de la fonction de contraste (10) peut être prise dans l'intervalle $(0, k_0]$. De cette façon $\Delta_{n,i}^{-k} \leq \Delta_{n,i}^{-k_0}$ et donc les deux développements ci-dessus sont vérifiés si $|x| = |X_{t_i}| \leq \Delta_{n,i}^{-k}$. Si cela n'est pas le cas, la contribution de l'observation X_{t_i} dans la fonction de contraste est simplement zéro. Cependant nous verrons que la suppression de la contribution de trop grands $|X_{t_i}|$ n'affecte pas l'efficacité de notre estimateur.

Nous remarquons que la contribution des sauts $\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x) \left[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)\right] F(z)dz$ présente dans les deux développements ne dépend pas de μ et donc il n'apparaît pas dans la différence $m(\mu, x) - m(\mu_0, x)$ mais elle n'est pas négligeable comparé à $\Delta_{n,i}\, b(x, \mu)$ car son ordre et $\Delta_{n,i}$ si $\alpha \in (0, 1)$ et au plus $\Delta_{n,i}^{\frac{1}{2}}$ si $\alpha \in [1, 2)$.

De plus, dans le cadre où l'intensité des sauts est finie et avec une choix spécifique de la fonction φ qui est une fonction oscillante, nous montrons qu'il est possible d'approximer notre fonction de contraste (10) avec une fonction complètement explicite, exactement comme dans le cas de Kessler. Cela nous donne un estimateur

efficace sous la condition $n\Delta_n^k \to 0$, où $k \geq 2$ est lié aux propriétés de la fonction oscillante φ.
En particulier nous trouvons le résultat suivant.

Résultat 4. *Soit φ une fonction $\mathcal{C}^\infty$ avec support compact et telle que $\varphi \equiv 1$ on $[-1,1]$ et $\forall k \in \{0,...,M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx = 0$ pour $M \geq 0$. Alors, pour $|x| \leq \Delta_{n,i}^{-k_0}$ avec $k_0 > 0$,*

$$m(\mu, x) = x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_K^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} + R(\mu, \Delta_{n,i}^{\beta(M+2)}, x), \tag{13}$$

où $A_K^{(k)}(x) = \bar{A}_c^k(g)(x)$, avec $g(y) = (y-x)$ et $\bar{A}_c(f) = \bar{b}f' + \frac{1}{2}a^2 f''$; $\bar{b}(\mu, y) = b(\mu, y) - \int_{\mathbb{R}} \gamma(y) z F(z) dz$.

Pour dire que l'équation (1.23) est utilisable, nous devons montrer l'existence d'une fonction φ avec support compact telle que $\varphi \equiv 1$ sur $[-1,1]$ et, $\forall k \in \{0,...,M\}$, $\int_{\mathbb{R}} x^k \varphi(x) dx$. Nous la construisons à travers ψ, une fonction avec support compact, $\mathcal{C}^\infty$ et telle que $\psi|_{[-1,1]}(x) = \frac{x^M}{M!}$. Nous définissons après $\varphi(x) := \frac{\partial^M}{\partial x^M}\psi(x)$.
Nous avons ainsi $\varphi \equiv 1$ sur $[-1,1]$, φ est $\mathcal{C}^\infty$, avec support compact et telle que pour chaque $l \in \{0,...M\}$, en utilisant l'intégration par parties, $\int_{\mathbb{R}} x^l \varphi(x)dx = 0$.
Nous observons que le développement (1.23) est le même trouvé dans Kessler [51] dans le cas sans sauts et il est obtenu par itération du générateur continu $\bar{A}_c$. Donc, il est complètement explicite. Nous soulignons que in Kessler [51] la partie à droite de (1.23) représente une approximation de $E[\bar{X}^\mu_{\Delta_{n,i}} \mid \bar{X}^\mu_0 = x]$ où $\bar{X}^\mu$ est la diffusion continue solution de $d\bar{X}^\mu_t = \bar{b}(\mu, \bar{X}^\mu_s)ds + \sigma(\bar{X}^\mu_s)dW_s$. Du Résultat 4, la partie à droite de (1.23) est aussi une approximation de $m(\mu, x) = \dfrac{E[X^\mu_{\Delta_{n,i}} \varphi_{\Delta^\beta_{n,i}}(X^\mu_{\Delta_{n,i}} - x) \mid X^\mu_0 = x]}{E[\varphi_{\Delta^\beta_{n,i}}(X^\mu_{\Delta_{n,i}} - x) \mid X^\mu_0 = x]}$ dans le cas d'intensité des sauts finie et pour une fonction de troncation à noyau φ qui satisfait $\forall k \in \{0,...,M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx = 0$. Nous insistons sur le fait que, dans l'expansion de m donnée dans le Résultat 4, la contribution de la partie discontinue du générateur disparaît seulement grâce à la choix d'une fonction φ oscillante.
Dans la définition de la fonction de contraste (10) nous pouvons remplacer $m(\mu, x)$ avec son approximation explicite $\widetilde{m}^k(\mu, x) := x + \sum_{h=1}^k \frac{\Delta^h_{n,i}}{h!} A_K^{(h)}(x)$, avec une erreur $R(\mu, \Delta^k_{n,i}, x)$ pour $k \leq \lfloor 2(M+1)\beta \rfloor$. Nous montrons que l'estimateur associé est efficace sous la condition $\sqrt{n}\Delta_n^{k-\frac{1}{2}} \to 0$ pour $n \to \infty$ (voir Proposition 1 dans le Chapitre 2).
Comme M et donc k peuvent être choisis arbitrairement grands, nous obtenons que le pas de discrétisation Δ_n peut converger à zéro à une vitesse polynomiale arbitrairement lente. Il se trouve que, avec un pas de discrétisation lent, il faut choisir une fonction de troncation qui annule plus de moments.

D'ailleurs nous montrons numériquement que, quand l'intensité des sauts est finie, l'estimateur que nous déduisons de l'approximation de la fonction de contraste (10) a une bonne performance et il rend le biais visiblement réduit. Quand au contraire l'intensité des sauts est infinie, nous construisons une approximation de m a partir de laquelle nous déduisons une approximation de la fonction de contraste (10) que l'on minimise dans le but d'obtenir l'estimateur $\hat{\mu}_n$. L'estimateur que nous obtenons est une version corrigée de l'estimateur qu'on aurait utilisé en considérant le schéma d'approximation d'Euler. Nous voyons numériquement que notre estimateur fonctionne bien et, en le comparant avec celui d'Euler, on remarque que le terme de correction que nous fournissons réduit drastiquement le biais, surtout lorsque α grandit.

Le deuxième chapitre porte sur l'estimation jointe des deux paramètres μ et σ apparaissants dans le modèle (2) avec une intensité des sauts qui est maintenant finie.
Nous considérons alors la fonction contraste (2.2) introduite avant comme généralisation naturelle de la fonction contraste proposée par Kessler dans [51] pour l'estimation des deux paramètres dans le cas sans sauts.
À partir de cette fonction nous définissons l'estimateur

$$\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n) := arg \min_{\theta} U_n(\mu, \sigma).$$

Le résultat principal de ce chapitre est la consistence de l'estimateur $\hat{\theta}_n$ et le fait qu'il converge vers une Gaussienne avec des variances asymptotiques explicites.

Résultat 5. *L'estimateur $\hat{\theta}_n$ est consistent en probabilité:*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \qquad n \to \infty.$$

Résultat 6. *L'estimateur $\hat{\theta}_n$ est asymptotiquement normal:*

$$(\sqrt{T_n}(\hat{\mu}_n - \mu_0), \sqrt{n}(\hat{\sigma}_n - \sigma_0)) \xrightarrow{\mathcal{L}} N(0, K) \quad \textit{pour } n \to \infty,$$

où $K = \begin{pmatrix} (\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} & 0 \\ 0 & 2(\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} \end{pmatrix}.$

En comparaison aux résultats precedents le pas de discrétisation peut être irrégulier, nous n'avons pas besoin d'introduire des conditions sur la vitesse à laquelle $\Delta_n \to 0$ et les deux paramètres sont estimés conjointement.
L'estimation jointe des deux paramètres introduit des difficultés notables : en particulier comme les deux paramètres ne s'estiment pas à la même vitesse, l'étude asymptotique de la fonction de contraste implique deux régimes asymptotiques différents.
L'étude des dérivées par rapport à tous les deux paramètres de la fonction de contraste est aussi nécessaire pour l'analyse du comportement asymptotique de l'estimateur. Les dérivées de la fonction de contraste n'étant pas explicites non plus, notre étude reposera sur des approximations explicites de m et m_2 et de leurs dérivées jusqu'à l'ordre troisième par rapport aux deux paramètres. Ces approximations des dérivées premières, deuxièmes et troisièmes sont contenues, respectivement, dans les Propositions 16, 17 et 18.
Une difficulté en plus dans ce chapitre est de montrer que, pour une fonction f à croissance polynomiale, nous avons $\frac{1}{n} \sum_{i=0}^{n-1} f(X_{t_i}, \theta) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f(x, \theta) \pi(dx)$ (voir Proposition 13). Pour cela, il nous faut une borne sur $|Cov(X_{t_i}, X_{t_j})|$. Cette convergence est indispensable pour le preuves des Résultats 5 et 6.
Il a été montré dans la Proposition 3.8 de [67] que le processus X est β - mixing avec décroissance exponentielle, c'est à dire qu'il existe $\gamma > 0$ tel que $\beta_X(k) = O(e^{-\gamma k})$; avec $\beta_X(k)$ défini comme dans Section 1.3.2 de [27]. Si le processus est β - mixing,

alors il est aussi α - mixing et donc la suivante estimation est vérifiée (voir Théorème 3 dans la Section 1.2.2 de [27])

$$|Cov(X_{t_i}, X_{t_j})| \leq c \left\|X_{t_i}\right\|_p \left\|X_{t_j}\right\|_q \alpha^{\frac{1}{r}}(X_{t_i}, X_{t_j})$$

avec p, q et r tels que $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. En utilisant que $\alpha(X_{t_i}, X_{t_j}) \leq \beta_X(|t_i - t_j|) = O(e^{-\gamma|t_i - t_j|})$, l'inégalité ci-dessus devient, pour une fonction f à croissance polynomiale, $|Cov(f(X_{t_i}, \theta), f(X_{t_j}, \theta))| \leq ce^{-\frac{1}{r}\gamma|t_i - t_j|}$.
Nous arrivons en conséquence à borner $Var(\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i}, \theta))$ avec une quantité qui tend vers 0 pour $n \to \infty$ et, donc, nous obtenons $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i}, \theta) - \int_{\mathbb{R}} f(x, \theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$.

De plus nous donnons des approximations explicites de m_2 qui, avec les approximations de m fournies dans le premier chapitre, nous permettent de contourner le fait que la fonction de contraste n'est pas explicite.
Dans ce cas aussi, comme nous avions déjà fait dans le premier chapitre pour m, nous montrons que à condition de choisir une fonction particulière φ qui soit oscillante nous sommes capables de supprimer la contribution des sauts et de fournir des développements explicites de la fonction m_2 à tout ordre. En utilisant les résultats des deux chapitres ensemble, alors, nous pouvons approximer notre fonction contraste (2.2) à un ordre arbitrairement haut avec une fonction complètement explicite exactement comme cela avait été fait par Kessler dans [51] dans le cas continu. Ceci nous donne un estimateur consistent et asymptotiquement normale sous la condition $n\Delta_n^k \to 0$ où, comme dit avant, k est lié aux propriétés d'oscillation de la fonction φ et, comme nous pouvons choisir k arbitrairement grand, notre méthode permet d'estimer conjointement les paramètres de dérive et de volatilité sous la faible condition que le pas de discrétisation aille à zéro à une vitesse polynomiale; à condition que l'intensité des sauts soit finie.

Résultat 7. *Soit φ une fonction $\mathcal{C}^\infty$ avec support compact et telle que $\varphi \equiv 1$ on $[-1, 1]$ et $\forall k \in \{0, ..., M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx = 0$ pour $M \geq 0$. Alors, pour $|x| \leq \Delta_{n,i}^{-k_0}$ avec $k_0 > 0$,*

$$m_2(\mu, \sigma, x) = \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_2}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} - (x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!})^2 + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x),$$

où $A_{K_1}^{(k)}(x) := \bar{A}_c^k(h_1)(x)$ et $A_{K_2}^{(k)}(x) := \bar{A}_c^k(h_2)(x)$, avec $\bar{A}_c(f) := \bar{b}f' + \frac{1}{2}a^2 f''$ et $\bar{b}(\mu, y) = b(\mu, y) - \int_{\mathbb{R}} \gamma(y) z F(z)dz$.
Le sigles K_1 que K_2 nous avons écrit représentent « Kessler ». Ceci est basée sur le fait que les développements que nous trouvons sont les mêmes obtenus dans le cas sans sauts par itération du générateur continu $\bar{A}_c$. Les fonctions qui apparaissent dans les définitions de $A_{K_1}^{(k)}$ et $A_{K_2}^{(k)}$ sont les suivantes: $h_1(y) := (y - x)$, $h_2(y) = y^2$.

Nous trouvons par ailleurs un développement exacte de la fonction m_2 jusqu'à l'ordre Δ_n^2, qui est valide pour n'importe quelle fonction régulière φ.

Résultat 8. *Soit $\beta \in (\frac{1}{4}, \frac{1}{2})$ et la mesure de Lévy F soit $\mathcal{C}^1$. Alors il existe $k_0 > 0$ tel que, pour $|x| \leq \Delta_{n,i}^{-k_0}$,*

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)} F(0) \int_{\mathbb{R}} v^2 \varphi(v)dv + \tag{14}$$

$$+\Delta_{n,i}^2(3\bar{b}^2(x,\mu)+h_2(x,\theta))+\Delta_{n,i}^{(1+4\beta)\wedge(2+\beta)\wedge(3-2\beta)}R(\theta,1,x);$$

où $h_2=\frac{1}{2}a^2(a')^2+\frac{1}{2}a^3a''+a^2\bar{b}'+aa'\bar{b}+\bar{b}^2$.

Nous observons que, si $\int_{\mathbb{R}}v^2\varphi(v)dv=0$, nous retombons sur le développement du Résultat 7 jusqu'à l'ordre 2. Nous voyons donc que la choix d'une fonction de troncation oscillante φ est nécessaire pour enlever la contribution des sauts.
Il est à noter que le terme le plus grand après celui principal est dû aux sauts et ne dépend pas des paramètres μ et σ. Nous verrons dans la suite qu'il nécessaire pour montrer la consistance de $\hat{\mu}_n$ que cette contribution ne dépende pas des paramètres. En considérant la différence de la fonction de contraste calculée dans deux valeurs différentes du paramètre de dérive, en effet, sa présence devient inutile.
Nous remarquons d'ailleurs que le terme de reste d'ordre $1+4\beta$ est négligeable comparé au terme d'ordre 2 parce que nous avons pris $\beta>\frac{1}{4}$.
Dans le Résultat 8 ci-dessus nous avons supposé $F\in\mathcal{C}^1$; telle condition n'est plus demandé dans le résultat plus général suivant:

Résultat 9. *Il existe* $k_0>0$ *tel que, pour* $|x|\le\Delta_{n,i}^{-k_0}$,

$$m_2(\mu,\sigma,x)=\Delta_{n,i}a^2(x,\sigma)+\frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}}u^2\varphi(u)F(u\frac{\Delta_{n,i}^{\beta}}{\gamma(x)})du+\Delta_{n,i}^2(3\bar{b}^2(x,\mu)+h_2(x,\theta))+$$

$$+\frac{\Delta_{n,i}^{2+\beta}a^2(x,\sigma)}{2\gamma(x)}\int_{\mathbb{R}}(u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du+\Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)}R(\theta,1,x),\quad(15)$$

où $h_2=\frac{1}{2}a^2(a')^2+\frac{1}{2}a^3a''+a^2\bar{b}'+aa'\bar{b}+\bar{b}^2$.

Nous voyons que la contributions des sauts dépend de la densité F dont l'argument dans l'intégrale dépend de $\Delta_{n,i}$. Si nous choisissons une fonction de densité F particulière qui est nulle dans le voisinage de 0 la contribution des sauts disparaît et, dans ce cas, nous retombons sur le développement pour m_2 trouvé par Kessler dans le cas sans sauts ([51]), jusqu'à l'ordre $\Delta_{n,i}^2$.
Le développement (15) peut sembler compliqué, cependant tous les termes sont nécessaires pour obtenir une expansion valide pour n'importe quelle densité F avec intensité finie et pour avoir un terme de reste avec un ordre qui soit explicite et strictement plus grand que 2.
Dans le cas particulier où F est $\mathcal{C}^1$, le trois premiers termes du développement nous donnent les termes principaux dans (2.14). La dernière intégrale, qui est d'ordre $\Delta_{n,i}^{2+\beta}$, est dans ce cas clairement un terme de reste.
Au contraire, dans la situation où F peut être non bornée dans le voisinage de 0, avec $\int F(z)dz<\infty$, la dernière intégrale est seulement négligeable comparé à $\Delta_{n,i}^2$. Donc, il pourrait n'être pas négligeable comparé aux termes de reste et c'est pour cela qu'il est nécessaire dans le développement.

Nous concluons cette partie avec une implémentation numérique de nos résultats principaux, en construisant deux approximations de m et m_2 à partir desquelles nous déduisons deux fonctions contrastes explicites que nous minimisons dans le but de trouver les estimateurs $\hat{\mu}_n$ et $\hat{\sigma}_n$. Nous comparons ces estimateurs avec ceux qui dérivent de l'utilisation du schéma d'Euler et nous remarquons que notre estimateur fonctionne mieux en particulier pour l'estimation paramétrique de σ.

0.3 Seconde partie: développement asymptotique de la variation quadratique tronquée et correction de biais.

La seconde partie de la thèse traite de l'estimation non - paramétrique de la volatilité intégrée à partir de l'observation dans les instants de temps $0 = t_0 \leq t_1 \leq \ldots \leq t_n =: T_n$ du processus X solution de la suivante équation différentielle stochastique avec sauts:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz) \tag{16}$$

où, comme dans la partie précédente, $W = (W_t)_{t\geq 0}$ est un mouvement Brownien en dimension 1 et $\tilde{\mu} = \mu - \bar{\mu}$ est une mesure ponctuelle de Poisson sur $(0, \infty) \times \mathbb{R}$ associée au processus de Lévy $L = (L_t)_{t\geq 0}$, avec une intensité des sauts qui peut être infinie. Il est quasiment un processus stable: son intensité est $F(dz) = \frac{g(z)}{|z|^{1+\alpha}}dz$, où $\alpha \in (0, 2)$ et $g : \mathbb{R} \to \mathbb{R}$ est une fonction continue, symétrique, non negative, bornée et telle que $g(0) = 1$.

Nous nous mettons encore dans le cadre de haute fréquence, en considérant donc un pas de discrétisation $\Delta_n := \sup_{i=0,\ldots,n-1} \Delta_{n,i}$, avec $\Delta_{n,i} = (t_{i+1} - t_i)$, tel que $\Delta_n \to 0$ pour n qui va à l'infini.

Nous étudions deux situations différentes qui nous portent à l'estimation de deux quantités différentes connectées à la volatilité de X. Dans la première nous supposons que l'horizon de temps $T_n = T \in [0, \infty[$ est fixé et indépendant de n et nous nous proposons comme objectif l'estimation de la quantité $IV_1 := \frac{1}{T_n}\int_0^{T_n} a^2(X_s)f(X_s)ds$, avec f n'importe quelle fonction à croissance polynomiale. Dans le second cas l'horizon de temps T_n est tel que $\lim_{n\to\infty} T_n = \infty$; nous voulons maintenant estimer la quantité $IV_2 := \int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$, où π est la mesure invariante dont on parle dans la Définition 3 d'ergodicité.

Si d'un coté l'estimation de IV_2, même en étant utile pour ses applications en inférence paramétrique, à notre connaissance n'a jamais été considérée avant, de l'autre IV_1 a déjà été largement étudiée puisqu'elle est très importante en finance. En effet, en prenant $f \equiv 1$, IV_1 devient ce qu'on appelle volatilité intégrée, qui est particulièrement pertinente pour mesurer et prévenir les risques des actifs : l'estimation de $\frac{1}{T_n}\int_0^{T_n} a^2(X_s)ds$ à partir des observations discrètes du processus X est en conséquence un des problèmes classiques.

La littérature précédente qui concerne la volatilité intégrée classique avec la fonction f identiquement égale à 1, que l'on notera toujours dans la suite de l'introduction comme IV_1, est en conséquence très vaste.

En absence de sauts, l'estimateur classique pour l'étude de la volatilité intégrée est la volatilité réalisée ou variation quadratique approximée au temps T_n, qui est définie de la façon suivante:

$$[X, X]_T^n := \sum_{i=0}^{n-1} (\Delta X_i)^2,$$

où $\Delta X_i = X_{t_{i+1}} - X_{t_i}$.

Sous des conditions faible sur l'integrabilité des coefficients b et a il est bien connu qu'il existe un théorème de la limite centrale avec vitesse $\sqrt{n}$ pour cet estimateur, c'est à dire que $\sqrt{n}([X, X]_T^n - IV_1)$ converge stablement en loi à une limite Z qui est définie sur une extension de l'espace initial et qui est une variable aléatoire Gaussienne centrée dont la loi conditionnelle est caractérisée par sa variance conditionnelle $V_T := 2\int_0^T a^4(X_s)ds$ (voir Section 2.4 in [52]).

Quand le processus X a des sauts, la volatilité réalisée $[X,X]_T^n$ ne converge plus vers IV_1, la volatilité intégrée que nous voulons estimer, mais vers IV_1 auquel il faut ajouter la contribution des sauts.
Il faut donc introduire des estimateurs différents, en présence de sauts. Plusieurs méthodes ont été proposées pour étudier la volatilité intégrée dans ce cas; pour en avoir un aperçu complet voir Section 3 dans [17].
Le premier type d'estimateurs qui sont robustes même en présence de sauts sont les variations Multipower (voir [9], [10], [46]). Ces estimateurs satisfont un théorème de la limite centrale avec vitesse $\sqrt{n}$ mais avec une variance plus grande que V_T, cela veut dire qu'ils sont efficaces en vitesse mais pas en variance.
Le seconde type d'estimateurs pour la volatilité dans le cadre avec sauts a été introduit par Jacod et Todorov dans [49]; ces estimateurs sont basés sur l'estimation locale de la fonction caractéristique empirique des accroissements du processus sur des blocs dont la longueur est décroissante mais qui contiennent un nombre croissant d'observations. L'estimation de la volatilité globale est donnée comme addition des estimations de la volatilité locale.
Une autre façon d'estimer la volatilité intégrale, qui est celui sur laquelle nous allons nous concentrer, a été introduit par Mancini dans [64] et consiste dans l'utilisation de la volatilité réalisée tronquée ou variation quadratique tronquée (voir aussi [46], [65]) :

$$\hat{IV}_T^n := \sum_{i=0}^{n-1}(\Delta X_i)^2 1_{\{|\Delta X_i| \le v_n\}}, \tag{17}$$

où v_n est une suite positive de niveaux de troncation, typiquement de la forme $v_n = (\Delta_n)^\beta$ pour $\beta \in (0, \frac{1}{2})$.
L'idée à la base de la definition de variation quadratique tronquée est de se ramener à la variation quadratique, qui était un bon estimateur dans le cas sans sauts, quand on pense de ne pas avoir eu des sauts dans l'interval considéré. Cela veut dire que, comme dans la première partie de la thèse, il faut introduire un critère pour juger si un saut s'est produit dans l'interval de temps $[t_i, t_{i+1}]$ ou pas. L'idée est la même qu'avant: du moment que l'accroissement d'une diffusion continue dépasse $(\Delta_n)^\beta$ avec une probabilité petite et un accroissement d'une diffusion avec même un seul saut dépasse $(\Delta_n)^\beta$ avec une grande probabilité, nous estimons qu'au moins un saut s'est produit dans $[t_i, t_{i+1}]$ si $|X_{t_{i+1}} - X_{t_i}| \ge (\Delta_n)^\beta$.

Pour ce qui concerne la littérature précédente liée a cet estimateur, il a été montré dans [45] que la variation quadratique tronquée $\hat{IV}_T^n$ définie comme dans (17) a exactement les mêmes propriétés que la variation quadratique $[X,X]_T^n$ avait dans le cas continue mais à condition que l'indice des sauts $\alpha \in (0,1)$ défini dans la Définition 2 et le seuil β soient bien choisis. En particulier, la condition qui ressort de [45] est que la variation quadratique tronquée est un estimateur à vitesse $\sqrt{n}$ pour $IV_1 := \frac{1}{T_n}\int_0^{T_n} a^2(X_s)ds$ quand

$$\beta \in [\frac{1}{2(2-\alpha)}, \frac{1}{2}). \tag{18}$$

Nous soulignons que, si $\alpha \ge 1$, il n'y a pas de $\beta \in (0, \frac{1}{2})$ pour lesquels la condition (18) ci - dessus peut être satisfaite.

Si les sauts du processus X sont ceux d'un processus stable (voir Définition 5 après) avec indice d'intensité $\alpha \geq 1$, Mancini a montré dans [65], en utilisant un pas de discrétisation uniforme $\Delta_{n,i} := \frac{1}{n}$ et une processus de Lévy additif L à la place d'une vrai EDS générale, que la différence entre la variation quadratique tronquée $\hat{IV}_T^n$ et l'objet qu'on veut estimer IV_1 est de taille $(\frac{1}{n})^{\beta(2-\alpha)}$; c'est à dire que

$$(\hat{IV}_T^n - IV_1) \overset{\mathbb{P}}{\sim} (\frac{1}{n})^{\beta(2-\alpha)}. \tag{19}$$

Cette vitesse est moins bonne que $\sqrt{n}$ et donc il n'y a pas de théorème de la limite central dans ce cas.

Quand l'horizon de temps T_n va à l'infini il n'y a pas à notre connaissance, dans la littérature, de résultat à propos de l'estimation de $IV_2 := \int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$.

Dans cette partie de la thèse nous voulons étendre les résultats montrés par Mancini des façons suivantes.

- Le résultat de Mancini avait été montré pour un processus de Lévy additif L. Nous le remplaçons avec le processus X, solution d'une vraie équation différentielle stochastique, définie dans (16).
- Nous généralisons le schéma d'échantillonnage par rapport à celui uniforme considéré par Mancini où $\Delta_{n,i} = \frac{1}{n}$ pour chaque $i \in \{0, \dots, n-1\}$. De plus, nous étudions aussi le cadre $T_n \to \infty$ pour $n \to \infty$.
- Nous voulons étendre (19) en fournissant une expansion asymptotique de $\hat{IV}_T^n$ qui nous donne, en détail, la contribution des sauts.
- En connaissant la contribution des sauts en détail, nous pouvons déduire des estimateur debiaisés pour l'estimation de $IV_1 := \frac{1}{T}\int_0^T a^2(X_s)f(X_s)ds$ et $IV_2 = \int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$.

L'estimateur que nous utilisons pour l'estimation de $IV_1 := \frac{1}{T}\int_0^T a^2(X_s)f(X_s)ds$ et $IV_2 := \int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$ est analogue à celui défini dans (17):

$$Q_n := \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(X_{t_{i+1}} - X_{t_i})^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}), \tag{20}$$

où φ est, comme dans la première partie, une version régulière de la fonction indicatrice: elle est C^∞ et s'annule quand les accroissements des données sont trop grands si comparés aux accroissements typiques d'une diffusion continue et, alors, cela peut être utilisé pour filtrer la contribution des sauts.

Le premier résultat de ce chapitre est constitué par une decomposition de la variation quadratique tronquée définie comme dans (20) en une partie continue et une partie à sauts, plus une troisième partie qui est négligeable si comparée aux deux autres.

Résultat 10. *Soient $\beta \in (0, \frac{1}{2})$ et $\alpha \in (0, 2)$. Alors, pour $\Delta_n \to 0$,*

$$Q_n = \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(X^c_{t_{i+1}} - X^c_{t_i})^2 + \Delta_n^{\beta(2-\alpha)}\tilde{Q}_n^J + \mathcal{E}_n =$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2 + \Delta_n^{\beta(2-\alpha)}\tilde{Q}_n^J + \mathcal{E}_n, \qquad où$$

$$\tilde{Q}_n^J := \frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds,dz))^2 \frac{f(X_{t_i})}{\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)$$

est la contribution des sauts, X^c est la partie continue du processus X et $\mathcal{E}_n$ est $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ et, pour chaque $\tilde{\epsilon} > 0$, $o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$.

Nous remarquons que le résultat énoncé ci -dessus est vérifié dans les deux cas que l'on étudie, c'est à dire pour $T_n = T$ fixé et T_n qui tend vers l'infini pour $n \to \infty$.

Avant donner des autres résultats obtenus dans cette partie, nous devons introduire des conditions sur le pas de discrétisation qui sont différentes dans les deux cas T fixé et $T \to \infty$.

- T fixé:
  Il nous faut supposer que, pour $\delta \in [0, 1)$, il existe une fonction mesurable $s \mapsto H(s, \delta)$ telle que, pour chaque fonction continue $h : \mathbb{R} \to \mathbb{R}$,
  $$\frac{1}{\Delta_n^\delta}\frac{1}{n}\sum_{i=0}^{n-1} h(X_{t_i})\Delta_{n,i}^\delta \to \int_0^T h(X_s)H(s,\delta)ds. \tag{21}$$
- $T \to \infty$: Nous supposons des conditions faibles et techniques sur la régularité du pas qui nous donnent le
  Lemme
  Pour chaque fonction mesurable $h : \mathbb{R} \to \mathbb{R}$ avec derivé bornée telle que $\pi(h) < \infty$ et pour $\delta \in [0, 1)$ nous avons
  $$\frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}^\delta}\sum_{i=0}^{n-1}\Delta_{n,i}^\delta h(X_{t_i}) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} h(x)\pi(dx).$$

Nous remarquons que la fonction H dépend du pas de discrétisation et, dans le cas particulier où il est uniforme, H devient tout simplement la fonction identité.

Maintenant nous nous concentrons sur l'erreur que nous commettons dans l'estimation de IV_1 et IV_2. Nous commençons en étudiant l'erreur qui dérive de l'estimation de la volatilité discrétisée. Nous décomposons cet erreur dans l'erreur statistique qui dérive de la partie continue et un bruit dû à la contribution des sauts.
Nous voulons passer de $\tilde{Q}_n^J =$

$$= \frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds,dz))^2 \frac{f(X_{t_i})}{\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) \qquad à$$

$$\hat{Q}_n^J = \frac{1}{n\Delta_n^{\beta(2-\alpha)}}c_\alpha\sum_{i=0}^{n-1} f(X_{t_i})|\gamma(X_{t_i})|^\alpha\Delta_{n,i}^{\beta(2-\alpha)}(\int_{\mathbb{R}}\varphi(u)|u|^{1-\alpha}du) + \tilde{\mathcal{E}}_n,$$

avec $\tilde{\mathcal{E}}_n = o_{\mathbb{P}}(1)$ et, si $\alpha < \frac{4}{3}$, aussi $\frac{1}{\Delta_n^{\beta(2-\alpha)}}o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$.
Le gain est que le term principal de $\hat{Q}_n^J$ est une statistique qui ne depend que des observations et α.
Pour justifier le passage de $\tilde{Q}_n^J$ à $\hat{Q}_n^J$ nous devons identifier la contribution des sauts en détail et, pour le faire, nous devons passer de notre processus, qui en temps petit a le comportement d'un processus de Lévy remis à l'échelle, à la situation d'un processus stable. Rappelons la définition de processus stable.

Definition 4. *Une variable aléatoire X est définie stable si pour chaque nombre entier $n \geq 2$ il y a deux suites à valeurs réelles (c_n) et (d_n) avec chaque $c_n > 0$ telles que la distribution de $X_1 + X_2 + \ldots + X_n$ est égale à $c_n X + d_n$, où $X_1, X_2, \ldots X_n$ sont copies indépendantes de X. En particulier, X est dit être strictement stable si chaque $d_n \equiv 0$ et X est dit être stable symétrique si X et $-X$ ont la même distribution.*

Il a été montré dans Feller [32], p.166 que la seule choix possible de c_n est de la forme $\sigma n^{\frac{1}{\alpha}}$, avec $\sigma > 0$ et $\alpha \in (0, 2]$. Le paramètre α est l'indice de stabilité du processus.

Definition 5. *Un processus stable est un processus de Lévy $(L_t)_{t \geq 0}$ tel que L_1 a distribution stable.*

Le processus stable symétrique a pour mesure de Lévy $\frac{dz}{|z|^{1+\alpha}}$.
Cela nous permet d'arriver au résultat suivant:

Résultat 11. *Pour $\Delta_n \rightarrow 0$, nous avons*

$$Q_n - \frac{1}{n} \sum_{i=0}^{n-1} f(X_{t_i}) a^2(X_{t_i}) = \frac{Z_n}{\sqrt{n}} + \Delta_n^{\beta(2-\alpha)} \hat{Q}_n^J + \mathcal{E}_n, \qquad \text{où}$$

$\mathcal{E}_n$ est négligeable.

1. *Si T est fixé $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_0^T a^4(X_s) f^2(X_s) H(s, 0) ds)$ stablement par rapport à X.*
2. *Si $\lim_{n \rightarrow \infty} T = \infty$, $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_{\mathbb{R}} a^4(x) f^2(x) \pi(dx))$.*

Pour obtenir le résultat ci - dessus la principale difficulté consiste en majorer l'erreur que nous commettons en passant du processus de Lévy remis à l'échelle au processus stable. L'outil principale qui nous permet de dire que cette erreur est négligeable si comparée à la contribution des sauts et à l'erreur statistique qui dérive de la partie continue de l'estimateur Q_n définie en (20) est le calcul de Malliavin. Nous rappelons maintenant des résultats sur le calcul de Malliavin pour processus avec sauts, en faisant référence à [77] et [14] pour une présentation complète et à [19] pour l'adaptation à notre cadre. Nous travaillons sur un espace de Poisson associé à la mesure μ_n, liée à un processus de Lévy remis à l'échelle $(\Delta_n^{-\frac{1}{\alpha}} L_{\Delta_n t})_{t \in [0,1]}$. Le compensateur de μ_n est $\bar{\mu}^n(dt, dz) =: dt F_n(z) dz$.
Il existe un opérateur linéaire L sur l'espace $D \subset \cap_{p \geq 1} L^p(F_n(z) dz)$ qui est auto-adjoint: $\forall \Phi, \Psi \in D$, $\mathbb{E} \Phi L \Psi = \mathbb{E} L \Phi \Psi$ (voir Section 8 dans [14] pour les details sur la construction de D). Nous appelons cet opérateur L opérateur de Malliavin.
Nous associons à L l'opérateur bilinéaire symétrique Γ:

$$\Gamma(\Phi, \Psi) = L(\Phi, \Psi) - \Phi L(\Psi) - \Psi L(\Phi).$$

Si f et h sont deux fonctions de test nous avons

$$\Gamma(\mu^n(f), \mu^n(h)) = \mu^n(\rho f' h'), \tag{22}$$

où une fonction de test est une fonction $f : [0, 1] \times \mathbb{R} \rightarrow \mathbb{R}$ mesurable, $\mathcal{C}^2$ par rapport à la deuxième variable et telle que $f \in \cap_{p \geq 1} L^p(\bar{\mu}^n(dt, dz))$ et ρ est une

fonctione auxiliaire (voir Section 3.7.2 pour détails).
Les opérateurs L et Γ satisfont le suivante propriété:

$$LF(\Phi) = F'(\Phi)L\Phi + \frac{1}{2}F''(\Phi)\Gamma(\Phi,\Phi), \qquad \Gamma(F(\Phi),\Psi) = F'(\Phi)\Gamma(\Phi,\Psi).$$

Ces opérateurs permettent d'établir la formule d'intégration par parties suivante (voir [14], Théorèmes 8 - 10 p.103)

Theorem 1. *Soient Φ et Ψ deux variables aléatoires dans D et soit f une fonction bornée avec dérivées bornées jusqu'à l'ordre deux. Si $\Gamma(\Phi,\Phi)$ est inversible et $\Gamma^{-1}(\Phi,\Phi) \in \cap_{p\geq 1}L^p$, alors c'est*

$$\mathbb{E}f'(\Phi)\Psi = \mathbb{E}f(\Phi)\mathcal{H}_\Phi(\Psi), \tag{23}$$

avec

$$\mathcal{H}_\Phi(\Psi) = -2\Psi\Gamma^{-1}(\Phi,\Phi)L\Phi - \Gamma(\Phi, \Psi\Gamma^{-1}(\Phi,\Phi)). \tag{24}$$

À travers l'utilisation de la formule d'intégration par parties énoncée dans le Théorème 1 nous arrivons à borner supérieurement l'erreur qui provient d'avoir remplacé le Lévy remis à l'échelle par le processus stable.
En particulier, nous étendons le Théorème 4.2 dans [19], où les auteurs utilisent le calcul de Malliavin pour montrer la borne suivante en variation totale, pour n'importe quelle fonction bornée h:

$$|\mathbb{E}[h(n^{\frac{1}{\alpha}}L_{\frac{1}{n}})] - \mathbb{E}[h(S_1^\alpha)]| \leq c\,\|h\|_\infty\,\epsilon_n, \tag{25}$$

où $(S_t^\alpha)_{t\geq 0}$ est un processus stable et, si $\alpha \leq 1$, $\forall\epsilon \in (0,1)$, $\epsilon_n = \frac{1}{n^{1-\epsilon}}$; si $\alpha > 1$, $\forall\epsilon \in (0,\frac{1}{\alpha})$, $\epsilon_n = \frac{1}{n^{\frac{1}{\alpha}-\epsilon}}$.
Dans notre cas, la fonction h est telle que $\|h\|_\infty$ est grand, la borne (25) n'est pas suffisante pour garantir que l'erreur soit négligeable dans le cas $\alpha > 1$. Nous introduisons alors une norme auxiliaire $\|h\|_{pol} := \sup_{x\in\mathbb{R}}(\frac{h(x)}{1+|x|^p})$ et démontrons un contrôle different sur l'erreur donné de la partie à droite de (25) (cf Prop 21 dans le Chapitre 3). Ce résultat est particulièrement utile dans le cas où $\|h\|_\infty$ est grande si comparée à $\|h\|_{pol}$. Cela est le cas si nous considerons, par example, une fonction h définie de la façon suivante: $h(x) := |x|^2 1_{|x|\leq M}$, pour M grand.
Nous pouvons, en consequence, identifier la contribution des sauts $\hat{Q}_n^J$ lequel terme principal est

$$\frac{1}{n\Delta_n^{\beta(2-\alpha)}}c_\alpha\sum_{i=0}^{n-1}f(X_{t_i})|\gamma|^\alpha(X_{t_i})\Delta_{n,i}^{\beta(2-\alpha)}(\int_{\mathbb{R}}\varphi(u)|u|^{1-\alpha}du),$$

où c_α est une constante qui depend de α et qui a été trouvée dans le Théorème 7.22 de [55].
En choisissant une fonction particulière φ pour laquelle l'intégrale vaut zéro nous pouvons supprimer la contribution principale du biais dû aux sauts dans l'estimateur (20). Nous pouvons par ailleurs calculer la limite de la contribution des sauts dans le deux cas (temps fixé et temps qui va à l'infini) pour l'enlever à l'estimateur initial et trouver donc une autre correction de Q_n. Nous obtenons, respectivement, les deux résultats suivants.

Résultat 12. *Soit φ une fonction telle que $\int_{\mathbb{R}} |u|^{1-\alpha}\varphi(u)du = 0$. Alors, pour $\alpha \in (0, \frac{4}{3})$ et $\beta \in (\frac{1}{4-\alpha}, (\frac{1}{2\alpha} \wedge \frac{1}{2}))$ nous avons, $\forall \tilde{\epsilon} > 0$,*

$$Q_n - \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) = \frac{Z_n}{\sqrt{n}} + o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}).$$

Résultat 13. ❖ *T fixé: pour $\Delta_n \to 0$,*

$$\hat{Q}_n^J \xrightarrow{\mathbb{P}} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_0^T |\gamma(X_s)|^\alpha f(X_s)H(s, \beta(2-\alpha))ds$$

et, en ajoutant une condition sur la regularité du pas,

$$Q_n - \frac{1}{T}\int_0^T a^2(X_s)f(X_s)ds = \frac{Z_n}{\sqrt{n}} + o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})+$$

$$+\Delta_n^{\beta(2-\alpha)}c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_0^T |\gamma(X_s)|^\alpha f(X_s)H(s, \beta(2-\alpha))ds.$$

❖ *$T_n \to \infty$, $n\Delta_n = O(T)$: pour $\Delta_n \to 0$,*

$$\hat{Q}_n^J \xrightarrow{\mathbb{P}} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_{\mathbb{R}} |\gamma(x)|^\alpha f(x)\pi(dx)$$

et, en ajoutant une condition sur la regularité du pas,

$$Q_n - \frac{1}{T}\int_0^T f(X_s)a^2(X_s)ds = \frac{Z_n}{\sqrt{n}}+$$

$$+\Delta_n^{\beta(2-\alpha)}c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_{\mathbb{R}} |\gamma(x)|^\alpha f(x)\pi(dx) + o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)}).$$

Pour conclure, le Résultat 13 réponds aux questions sur le développement asymptotique pour la volatilité intégrée tronquée en temps petit et temps long. À chaque fois le premier terme Z_n est l'erreur statistique (évalué par le Résultat 11) et le terme intégrale suivante est une estimation explicite du biais dû aux sauts.
Par rapport aux résultats precedents, qui existent seulement en temps petit, notre développement asymptotique nous donne exactement la limite à laquelle $\frac{1}{\Delta_n^{\beta(2-\alpha)}}(Q_n - IV_1)$ converge quand $\Delta_n^{\beta(2-\alpha)} > \sqrt{n}$, qui dans le cas où le pas de discrétisation est uniforme ($\Delta_n = \frac{T}{n}$) corresponde à la condition $\beta < \frac{1}{2(2-\alpha)}$.
Quand le pas de discrétisation est uniforme ce chapitre étend le résultat de Mancini [65]. En effet nous trouvons

$$Q_n - IV_1 = \frac{Z_n}{\sqrt{n}} + (\frac{1}{n})^{\beta(2-\alpha)}c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_0^T |\gamma|^\alpha(X_s)f(X_s)ds + o_{\mathbb{P}}((\frac{1}{n})^{\beta(2-\alpha)}),$$

où $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_0^T a^4(X_s)f^2(X_s)ds)$ stablement par rapport à X. Notre développement asymptotique nous permet de déduire la comportement de la variation quadratique tronquée pour n'importe quelle couple (α, β), qui est un plus par rapport à (19).
Par ailleurs, en supposant de connaître α (et si ce n'est pas le cas il suffit de l'estimer préalablement, voir par exemple [91]), nous pouvons améliorer la performance de la

variation quadratique en enlevant de l'estimateur original le bruit dû à la présence des sauts ou en choisissant des fonctions particulières φ pour lesquels le biais dérivé de la partie à saut est égal à zéro. En utilisant le développement asymptotique nous fournissons aussi la taille de l'erreur qui reste après avoir appliqué les corrections à l'estimateur initial Q_n.

De plus, quand la volatilité est constante: $a(x) = a > 0$, nous montrons numériquement que les corrections obtenues de la connaissance de l'expansion asymptotique pour la volatilité intégrée en temps petit nous permet de réduire visiblement le bruit pour n'importe quel $\beta \in (0, \frac{1}{2})$ et $\alpha \in (0, 2)$.
Pour estimer $IV_1 = a^2$ nous prenons

$$Q_n = \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 \varphi_{k\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}).$$

L'erreur normalisée commise en utilisant l'estimateur original est $E_1 := (Q_n - a^2)\sqrt{n}$. Nous voulons maintenant corriger cet estimateur. Dans ce cas la contribution de la partie à sauts est (cf Résultat 13)

$$\hat{Q}_n^J = c_\alpha \gamma^\alpha k^{2-\alpha} (\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du) + \tilde{\mathcal{E}}_n.$$

Nous pouvons réduire l'erreur par soustractions, en construisant

$$Q_n^c := Q_n - \Delta_n^{\beta(2-\alpha)} c_\alpha \gamma^\alpha k^{2-\alpha} (\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du).$$

L'erreur commise en utilisant cet estimateur corrigé est $E_2 := (Q_n^c - a^2)\sqrt{n}$.
Nous pouvons d'ailleurs considérer une fonction $\psi(\zeta)$ qui soit C^∞ et qui soit different de zéro que pour les ζ tels que $|z| \in (1, 2)$. Alors la fonction $\tilde{\varphi}(\zeta) := \varphi(\zeta) + c\psi(\zeta)$ est pour n'importe quel c toujours une version régulière de la fonction indicatrice. Nous choisissons

$$\tilde{c} := -\frac{\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du}{\int_{\mathbb{R}} \psi(u)|u|^{1-\alpha} du},$$

pour lequel on a $\int_{\mathbb{R}} (\varphi + \tilde{c}\psi(u))|u|^{1-\alpha} du = 0$.
Nous pouvons en conséquence construire l'estimateur debiaisé

$$Q_{n,c} := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 (\varphi + \tilde{c}\psi)(\frac{X_{t_{i+1}} - X_{t_i}}{k\Delta_{n,i}^{\beta}}).$$

L'erreur commise en considérant cet estimateur est $E_3 := (Q_{n,c} - a^2)\sqrt{n}$. Nous reportons dans le tableau 2.3 les résultats numériques que nous avons obtenus à travers une estimation de Monte Carlo de E_1, E_2 et E_3 avec un echantillon de 500 éléments pour $n = 700$ et $\beta = 0.49$ selon les valeurs de α et du coefficient des sauts γ.

Nous remarquons que, si le biais de l'estimateur original est presque toujours plus grand que son écart type, après avoir effectué les corrections les deux erreurs restent au dessous de l'écart type.
On peut voir que les correction effectuées sont une claire amélioration parce que, si la variation quadratique initiale fonctionnait bien seulement pour $\beta > \frac{1}{2(2-\alpha)}$ (condition jamais satisfaite pour $\alpha \geq 1$), la variation quadratique tronquée debiaisée atteint des résultats excellents pour n'importe quel couple (α, β).

α	γ	**Mean** E_1	**Rms** E_1	**Mean** E_2	**Mean** E_3
0.1	1	1.092	1.535	0.307	-0.402
	3	1.254	1.627	0.378	-0.372
0.5	1	2.503	1.690	0.754	-0.753
	3	4.680	2.146	1.651	-0.824
0.9	1	2.909	1.548	0.217	0.416
	3	8.042	1.767	0.620	-0.404
1.2	1	7.649	1.992	-0.944	-0.185
	3	64.937	9.918	-1.692	-2.275
1.5	1	25.713	3.653	-1.697	3.653
	3	218.591	21.871	-4.566	-13.027
1.9	1	238.379	14.860	-6.826	16.330
	3	2357.553	189.231	3.827	-87.353

Table 1 – Estimation de Monte Carlo de E_1, E_2 et E_3 avec un echantillon de 500 éléments. Nous avons fixé $n = 700$ et $\beta = 0.49$.

0.4 Troisième partie: estimation adaptative de la mesure stationnaire.

Cette partie de la thèse traite de l'estimation non paramétrique de la densité invariante p associée au processus $(X_t)_{t\geq 0}$, solution de l'équation différentielle stochastique multi -dimensionnelle avec sauts:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d\setminus\{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds, dz), \tag{26}$$

où W est un mouvement Brownien d - dimensionnel et $\bar{\mu}$ une mesure de Poisson ponctuelle avec activité des sauts qui peut être infinie.

0.4.1 Estimation de la densité de la loi d'un échantillon iid

Pour introduire le problème d'estimation non -paramétrique nous considérons tout d'abord un problème plus facile: soient $X_1, \dots, X_n$ des variables aléatoires à valeurs réelles identiquement distribuées dont la distribution commune est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. Le densité de cette distribution, dénotée par p, est une fonction de $\mathbb{R} \times [0, \infty)$ qui est supposée être inconnue. Le problème est l'estimation de cette densité p.
Un estimateur de p est une fonction $x \to p_n(x) = p_n(x, X_1, \dots, X_n)$ mesurable par rapport aux observations $X = (X_1, \dots, X_n)$. Si nous savons à priori que p appartient à une famille paramétrique $\{g(x, \theta) : \theta \in \Theta\}$, où $g(\cdot, \cdot)$ est une fonction donnée et Θ est un sous-ensemble de $\mathbb{R}^k$ avec dimension fixée k indépendante de n, alors l'estimation de p est équivalente à l'estimation d'un paramètre de dimension finie θ et il s'agit d'un problème d'estimation paramétrique. Si au contraire nous ne disposons pas d'une information de ce genre à priori sur p, il s'agit d'un problème d'estimation non paramétrique. Dans l'estimation non paramétrique on suppose normalement que p soit dans une classe très vaste de densités. Par exemple,

cela peut être l'ensemble des densités de probabilité continues sur $\mathbb{R}$ ou l'ensemble des densités de probabilité Lipschitz continues sur $\mathbb{R}$. Des classes de ce type sont appelées classes non paramétriques des fonctions.

Nous voulons étudier $X_1, \ldots, X_n$, variables aléatoires iid et estimer de façon non -paramétrique leur densité de probabilité p. La fonction de répartition correspondante est $F(x) = \int_{-\infty}^{x} p(t)dt$. On considère la fonction de répartition empirique:

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} 1(X_i \leq x),$$

où 1(.) denote la fonction indicatrice. De la loi forte des grands nombres nous avons

$$F_n(x) \to F(x) \qquad \forall x \in \mathbb{R}$$

presque sûrement pour $n \to \infty$. Donc, $F_n(x)$ est un estimateur consistent de $F(x)$ pour n'importe quel $x \in \mathbb{R}$.
Comment peut-on estimer la densité p? La première solution intuitive est basée sur l'argumentation suivante: pour $h > 0$ suffisamment petit nous pouvons écrire l'approximation

$$p(x) \sim \frac{F(x+h) - F(x-h)}{2h}.$$

En replaçant F avec son estimation F_n nous définissons

$$\hat{p}_n^R(x) := \frac{F_n(x+h) - F_n(x-h)}{2h}.$$

La fonction $\hat{p}_n^R$ est un estimateur de p appelé l'estimateur de Rosenblatt. On peut le réécrire dans la forme

$$\hat{p}_n^R(x) = \frac{1}{2nh}\sum_{i=1}^{n} 1(x-h \leq X_i \leq x+h) = \frac{1}{nh}\sum_{i=1}^{n} K_0(\frac{X_i - x}{h}),$$

où $K_0(u) = \frac{1}{2}1(-1 \leq u \leq 1)$. Une generalization simple de l'estimateur de Rosenblatt est donnée par

$$\hat{p}_n(x) = \frac{1}{nh}\sum_{i=1}^{n} K(\frac{X_i - x}{h}), \tag{27}$$

où $K : \mathbb{R} \to \mathbb{R}$ est une fonction integrable qui satisfait $\int_{\mathbb{R}} K(u)du = 1$. Cette fonction K est appelée fonction noyau et le paramètre h est appelé fenêtre de l'estimateur (27). La fonction $x \to \hat{p}_n(x)$ est appelée estimateur de la densité par noyau.
Dans le cadre asymptotique, pour $n \to \infty$, nous considérerons une fenêtre h qui dépend de n; on le dénotera h_n et on supposera que la suite $(h_n)_{n\geq 1}$ tende vers 0 pour $n \to \infty$. Dans la suite h sera dénotée sans l'indice n pour alléger la notation quand cela ne causera pas d'ambiguïté.
Des exemples classiques de fonctions noyaux sont les suivants:

$$K(u) := \frac{1}{2}1(|u| \leq 1) \qquad \text{(noyau rectangulaire)},$$

$$K(u) := (1-|u|)1(|u| \leq 1) \qquad \text{(noyau triangulaire)},$$

$$K(u) := \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}} \qquad \text{(noyau Gaussien)}.$$

Nous remarquons que, si la fonction noyau prend seulement des valeurs qui ne sont pas négatives et si $X_1, \ldots, X_n$ sont fixés, alors la fonction $x \to \hat{p}_n(x)$ est une densité de probabilité.

L'estimateur (27) peut être généralisé au cadre multidimensionnel. Par exemple, on peut définir un estimateur de la densité par noyau en 2 dimension de la façon suivante: on suppose avoir observé n couple de variables aléatoires $(X_1, Y_1), \ldots, (X_n, Y_n)$ tels que (X_i, Y_i) sont iid avec densité $p(x, y)$ dans $\mathbb{R}^2$. Un estimateur par noyau de $p(x, y)$ dans le cas isotropique est donné par la formule

$$\hat{p}_n(x, y) = \frac{1}{nh^2} \sum_{i=1}^{n} K(\frac{X_i - x}{h}) K(\frac{Y_i - y}{h}),$$

où $K : \mathbb{R} \to \mathbb{R}$ est une fonction noyau définie comme dessus et $h > 0$ est une fenêtre.

Une mesure de le precision de l'estimateur $\hat{p}_n$ est son risque quadratique moyen (ou erreur quadratique moyen) dans un point arbitraire $x_0 \in \mathbb{R}$:

$$MSE(x_0) = \mathbb{E}_p[(\hat{p}_n(x_0) - p(x_0))^2],$$

où MSE dénote l'erreur quadratique moyen et $\mathbb{E}_p$ dénote l'espérance par rapport à la distribution de $(X_1, \ldots, X_n)$.
Nous observons que l'on a

$$MSE(x_0) = B^2(x_0) + V^2(x_0), \qquad \text{avec}$$

$$B(x_0) := \mathbb{E}_p[(\hat{p}_n(x_0)] - p(x_0) \qquad \text{et}$$
$$V^2(x_0) := \mathbb{E}_p[(\hat{p}_n(x_0) - \mathbb{E}_p[(\hat{p}_n(x_0)])^2].$$

Definition 6. *Les quantités $B(x_0)$ et $V^2(x_0)$ sont appelées respectivement biais et variance de l'estimateur $\hat{p}_n$ dans le point x_0.*

Pour évaluer le risque quadratique moyen de $\hat{p}_n$ il faut considerer séparément sa variance et son biais. Nous commençons avec l'etude de la variance.
Il a été montré dans la Proposition 1.1 de [92] que, si la densité p satisfait $p(x) \leq p_{max} < \infty$ pour n'importe quel $x \in \mathbb{R}$ et $K : \mathbb{R} \to \mathbb{R}$ est une fonction telle que $\int_{\mathbb{R}} K^2(u)du < \infty$; alors pour n'importe quel $x_0 \in \mathbb{R}$, $h > 0$ et $n \geq 1$ nous avons

$$V^2(x_0) \leq \frac{c_1}{nh}, \tag{28}$$

où $c_1 := p_{max} \int_{\mathbb{R}} K^2(u)du$.
Nous pouvons en conclure que, si $h = h_n$ est telle que $nh \to \infty$ pour $n \to \infty$, alors la variance va à zero pour $n \to \infty$.
Pour ce qui concerne le biais de l'estimateur $\hat{p}_n$, nous savons que sa forme est

$$B(x_0) := \mathbb{E}_p[(\hat{p}_n(x_0)] - p(x_0) = \frac{1}{h} \int_{\mathbb{R}} K(\frac{z - x_0}{h}) p(z)dz - p(x_0).$$

Nous analysons maintenant le comportement de $B(x_0)$ comme fonction de h sous des conditions de régularité sur la densité p et sur la fonction noyau K. Dans la suite $\lfloor \beta \rfloor$ dénotera le nombre entier le plus grand qui est strictement inférieur au nombre réel β.
Nous avons besoin tout d'abord de quelques définitions.

Definition 7. *Soient T un interval de $\mathbb{R}$ et β et L deux nombres positifs. La classe de Hölder $\Sigma(\beta, L)$ sur T est définie comme l'ensemble des fonctions $f : T \to \mathbb{R}$ qui sont $l = \lfloor \beta \rfloor$ différentiables et dont la dérivée $f^{(l)}$ satisfait*

$$|f^{(l)}(x) - f^{(l)}(x')| \leq L|x - x'|^{\beta - l}, \qquad \forall x, x' \in T.$$

Definition 8. *Soit $l \geq 1$ un nombre entier. Nous disons que $K : \mathbb{R} \to \mathbb{R}$ est un noyau d'ordre l si les fonctions $u \mapsto u^j K(u)$, $j = 0, 1, \ldots, l$ sont intégrables et satisfont*

$$\int_{\mathbb{R}} K(u)du = 1, \quad \int_{\mathbb{R}} u^j K(u)du = 0 \quad \forall j = 1, \ldots, l.$$

Supposons que p appartient à la classe des densités $\mathcal{P} = \mathcal{P}(\beta, L)$ définie de la façon suivante:

$$\mathcal{P}(\beta, L) := \left\{ p : \, p \geq 0 \int_{\mathbb{R}} p(x)dx = 1 \, \text{et} \, p \in \Sigma(\beta, L) \, \text{sur} \, \mathbb{R} \right\}.$$

Supposons par ailleurs que K soit un noyau d'ordre $l = \lfloor \beta \rfloor$ qui satisfait

$$\int_{\mathbb{R}} |u|^{\beta} |K(u)| du < \infty;$$

il y a alors le suivant résultat sur l'estimation du biais.

$$B(x_0) \leq c_2 h^{\beta}, \tag{29}$$

où $c_2 := \frac{L}{l!} \int_{\mathbb{R}} |u|^{\beta} |K(u)| du$.

Nous voyons de (28) et (29) que les bornes supérieurs sur le biais et sur la variance se comportent de manière opposée au changer de h. Quand h augmente le borne sur le biais augmente alors que la variance diminue.
La choix d'une fenêtre h petite, correspondante à une grande variance, est appelé "under smoothing". Au contraire quand h est grande le biais ne peut pas être raisonnablement controlé, on parle dans ce cas de "oversmoothing". Une valeur de h optimale est une valeur qui balance le biais et la variance et cela est située entre ces deux extrêmes. Pour avoir un aperçu de la choix optimale de h, nous pouvons minimiser en h le borne supérieur sur le MSE obtenu comme consequence des résultats precedents. Les estimations (28) et (29) nous donnent

$$MSE \leq c_2^2 h^{2\beta} + \frac{c_1}{nh}.$$

Le minimum par rapport à h de la partie à droite de l'équation ci dessus est obtenu en choisissant

$$h_n^* := \left(\frac{c_1}{2\beta c_2^2}\right)^{\frac{1}{2\beta+1}} n^{-\frac{1}{2\beta+1}};$$

Avec cette fenêtre nous arrivons à la borne supérieure suivante (voir Théorème 1.1 dans [92])

$$\sup_{x_0 \in \mathbb{R}} \sup_{p \in \mathcal{P}(\beta, L)} \mathbb{E}_p[(\hat{p}_n(x_0) - p(x_0))^2] \leq c n^{-\frac{2\beta}{2\beta+1}},$$

où $C > 0$ est une constante qui depend de β, L et de la fonction noyau K.
La vitesse de convergence de l'estimateur $\hat{p}_n(x_0)$ est $\psi_n := n^{-\frac{\beta}{2\beta+1}}$.
Maintenant nous nous demandons si c'est il possible d'améliorer la vitesse de convergence ψ_n en utilisant des autres estimateurs de la densité et, dans ce cas, combien serait la vitesse de convergence optimale. Pour répondre à ces questions il est utile de considerer le risque minimax $\mathcal{R}_n^*$ associé à la classe $\mathcal{P}(\beta, L)$:

$$\mathcal{R}_n^*(\mathcal{P}(\beta, L)) := \inf_{T_n} \sup_{p \in \mathcal{P}(\beta, L)} \mathbb{E}_p[(T_n(x_0) - p(x_0))^2],$$

où l'inf est considéré sur la classe de tous les estimateurs possibles.
Le Théorème 5.1 dans [44] nous donne aussi un borne inférieur sur le risque minimax:

$$\mathcal{R}_n^*(\mathcal{P}(\beta, L)) \geq C'\psi_n^2$$

pour une constante $C' > 0$. Cela implique que l'estimateur à noyau atteint la vitesse de convergence optimale $n^{-\frac{\beta}{2\beta+1}}$ associée à la classe de densités $\mathcal{P}(\beta, L)$.

Nous donnons maintenant des definitions relatives à la vitesse de convergence optimale dans un cadre général, en analysant le borne inférieur du risque minimax.
On vient de voir que le problème d'estimation non - paramétrique est caractérisé par trois objets: une classe Θ non -paramétrique de fonctions qui contient la fonction θ que nous voulons estimer; une famille $\{\mathcal{P}_\theta, \theta \in \Theta\}$ de mesures de probabilité sur un espace mesurable et une distance d sur Θ qui est utilisée pour définir le risque.
La performance de un estimateur $\hat{\theta}_n$ de θ est donnée par le risque maximal de cet estimateur sur Θ:

$$r(\hat{\theta}_n) := \sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\hat{\theta}_n, \theta)],$$

où $\mathbb{E}_\theta$ denote l'espérance par rapport à $\mathcal{P}_\theta$.
Une borne supérieure sur le risque maximal est une inégalité de la forme

$$sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\hat{\theta}_n, \theta)] \leq C\psi_n^2$$

pour un estimateur $\hat{\theta}_n$, une suite positive $\psi_n \to 0$ et une constante $C < \infty$. Nous pouvons complementer ces bornes supérieurs avec les bornes inférieurs correspondents:

$$\forall \hat{\theta}_n : \quad sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\hat{\theta}_n, \theta)] \geq c\psi_n^2,$$

pour n suffisamment grand et où c est une constante positive. Dans ce contexte, il est utile de définir le risque minimax associé a d:

$$\mathcal{R}_n^* = \inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\hat{\theta}_n, \theta)],$$

où l'infimum est sur tous les estimateurs. La borne supérieure implique qu'il existe une constante $C < \infty$ telle que

$$\limsup_{n \to \infty} \psi_n^2 \mathcal{R}_n^* \leq C \tag{30}$$

pour une suite ψ_n qui tende vers zero. La borne inférieure correspondente nous dit qu'il existe une constante $c > 0$ telle que, pour la même suite ψ_n

$$\liminf_{n \to \infty} \psi_n^2 \mathcal{R}_n^* \geq c. \tag{31}$$

Definition 9. *Une suite positive* $(\psi_n)_{n=1,..,\infty}$ *est appelée vitesse optimale de convergence si elle est telle que* (30) *et* (31) *sont vraies.*
Un estimateur θ_n^* *qui satisfait*

$$\sup_{\theta \in \Theta} \mathbb{E}_\theta[d^2(\theta_n^*, \theta)] \leq C\psi_n^2,$$

où $(\psi_n)_{n=1,..,\infty}$ *est la vitesse optimale de convergence et* $C < \infty$ *est une constante, est dit estimateur avec vitesse optimale sur* (Θ, d).

0.4.2 Estimation de la mesure stationnaire pour une diffusion continue

Nous nous concentrons maintenant sur l'estimation de la mesure stationnaire d'une diffusion continue.
Pour $d = 1$ Schmisser considère, dans [83], une un processus strictement stationnaire et β- mixing $(X_t)_{t\geq 0}$ observé à des instants $t_i^n = 0, \Delta_n, ..., n\Delta_n$. L'auteur estime les dérivées $\mu^{(j)}$ de la densité stationnaire sur un ensemble compact et sur $\mathbb{R}$ en utilisant une méthode des moindres carrés pénalisés. Si la dérivée $\mu^{(j)}$ appartient à l'espace de Besov $\mathcal{B}^\alpha_{2,\infty}$, alors le risque L^2 de l'estimateur converge à la vitesse $(n\Delta_n)^{\frac{-2\alpha}{2\alpha+2j+1}}$ et la procédure ne requiert pas la connaissance de α. Quand $j = 0$, la densité invariante associée au processus $(X_t)_{t\geq 0}$ est estimée à une vitesse de convergence qui est la même que celle trouvée par Comte et Merlevède dans [21] et [22].

Dans la littérature existante dans le cadre multidimensionnelle, une référence importante a été donnée par Dalalyan et Reiss dans [25], où ils montrent une équivalence asymptotique pour l'inférence sur la dérive dans le cas de diffusion multi - dimensionnelle. Comme sous-produit de l'étude ils montrent, sous des contraintes de régularité Hölder isotropique, des vitesses de convergence de l'estimateur de la densité invariante qui sont plus rapides par rapport à ceux connus par rapport à l'estimation multivariate standard de la densité. Leur résultat est basé sur des bornes supérieurs sur la variance de fonctionnelles de diffusion, qui sont montrés à travers une application de l'inégalité de trou spectral en combinaison avec une borne sur la transition de densité du processus.

En particulier, ils supposent qu'un enregistrement continu des observations $(X_t^T)_{0\leq t\leq T}$ d'un processus d-dimensionnel jusqu'à l'instant T est disponible. Ce processus de diffusion est donné comme solution forte de l'équation différentielle stochastique

$$dX_t = b(X_t)dt + dW_t \quad X_0 = \zeta, \quad t \in [0, T], \tag{32}$$

où $b : \mathbb{R}^d \to \mathbb{R}^d$, $W = (W_t)_{t\geq 0}$ est un processus Brownien d-dimensionnel et ζ est un vecteur aléatoire indépendant de W; $b_i : \mathbb{R}^d \to \mathbb{R}$ sont les composantes de la fonction b. Ils supposent aussi que la fonction de dérive soit de la forme $b = -\nabla V$, où $V \in \mathcal{C}^2(\mathbb{R}^d)$ est un potentiel.
Cette restriction leur permet d'utiliser des résultats forts pour le semigroup de Markov de la diffusion sur l'espace L^2 engendré par la mesure invariante.
Pour des constantes positives M_1 et M_2, ils définissent $\Sigma(M_1, M_2)$ l'ensemble des fonctions $b = -\nabla V : \mathbb{R}^d \to \mathbb{R}^d$ qui satisfassent, pour chaque $x, y \in \mathbb{R}^d$,

$$|b(x)| \leq M_1(1 + |x|), \quad \text{and}$$

$$(b(x) - b(y))^T(x - y) \leq -M_2|x - y|^2. \tag{33}$$

Une fonction b qui satisfait les deux conditions ci-dessus est localement Lipschitz-continue. L'équation (32) admet alors une solution forte unique, qui est un processus de Markov homogène continu (voir Théorème 12.1 dans Roger et Williams [81]). Soient $C_b := \int_{\mathbb{R}^d} e^{-2V(u)}du$ et

$$\mu_b(x) := C_b^{-1} e^{-2V(x)}, \tag{34}$$

pour $x \in \mathbb{R}^d$. Sous la condition (33) cela donne $C_b < \infty$ et le processus X est ergodique avec une unique mesure de probabilité invariante (Bhattacharya [13], Théorème 3.5). De plus, la mesure de probabilité invariante est absolument continue par rapport à la mesure de Lebesgue et sa densité est μ_b. Dalalyan et Reiss supposent aussi que la valeur initiale ζ dans (32) suive la loi invariante, de façon que le processus X soit strictement stationnaire. Ils dénotent avec $P_{b,t}$ le semigroup de transition du processus X, i.e. $P_{b,t}f(x) = \mathbb{E}_b[f(X_t)|X_0 = x]$. La densité de transition est $p_{b,t}$: $P_{b,t}f(x) = \int_{\mathbb{R}^d} f(y)p_{b,t}(x,y)dy$.

Ils supposent que le potentiel V soit dans la classe de Hölder définie dans 7 qu'ils dénotent $H(\beta+1, L)$, pour $\beta, L > 0$. Cela implique $b_i \in H(\beta, L)$. De plus, si pour une constante $C_1 > 0$ nous avons $\max_{i=1,...,d} \max_{\alpha:|\alpha|\le\beta} |D^\alpha b_i(0)| \le C_1$, alors la fonction μ_b est Hölder continue d'ordre $\beta+1$ dans chaque ensemble borné $A \subset \mathbb{R}^d$. Un estimateur naturel à noyau pour cette densité invariante, basé sur l'observation de $(X_t^T)_{t\in[0,T]}$ est donné par

$$\hat{\mu}_{h,T}(x) := \frac{1}{T}\int_0^T K_h(x - X_t)dt, \quad x \in \mathbb{R}^d,$$

avec $K_h(x) = h^{-d}K(h^{-1}x)$ et $K : \mathbb{R}^d \to \mathbb{R}$ est une fonction à noyau d'ordre β (définie dans 8) régulière et à support compact.
Pour trouver la vitesse de convergence de cet estimateur ils utilisent la décomposition biais -variance usuelle (voir Efromovich [29], § 8.9):

$$\mathbb{E}_b[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2] \le h^{2(\beta+1)} + \frac{1}{T^2}Var_b(\int_0^T K_h((x - X_t)dt)). \tag{35}$$

Pour estimer la variance ils ont besoin d'introduire les conditions suivantes sur le semi - groupe de transition et la densité de transition:

1. Il y a $\rho > 0$ tel que, pour chaque $f : \mathbb{R}^d \to \mathbb{R}$: $\int_{\mathbb{R}^d} |f(x)|^2 \mu_b(dx) < \infty$ et pour chaque $t > 0$:
$$\|P_{b,t}f - \mu_b(f)\|_{\mu_b} \le e^{-t\rho} \|f\|_{\mu_b}.$$
2. Il y a $C_0 > 0$ tel que pour chaque $t > 0$ et pour chaque couple de points $x, y \in \mathbb{R}^d$ qui satisfont $|x-y|^2 < t$ nous avons
$$p_{b,t}(x,y) \le C_0(t^{-\frac{d}{2}} + t^{\frac{3d}{2}}). \tag{36}$$

Ils obtiennent alors

$$\frac{1}{T^2}Var_b(\int_0^T K_h((x - X_t)dt)) \le \frac{1}{T}\psi_d^2(h^d), \qquad \text{où}$$

$$\psi_d(x) = \begin{cases} \max(1, \log^2(\frac{1}{x})), & d = 2 \\ x^{\frac{1}{d}-\frac{1}{2}}, & d \ge 3 \end{cases} . \tag{37}$$

La choix optimale $h = h(T) \sim T^{-\frac{1}{2\beta+d}}$ donne les vitesses suivantes:

$$\mathbb{E}_b[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2]^{\frac{1}{2}} \lesssim \begin{cases} T^{\frac{-1}{2}}(\log T)^2, & d = 2 \\ T^{-\frac{\beta+1}{2\beta+d}}, & d \geq 3. \end{cases} \tag{38}$$

Toujours dans le cas sans sauts, dans un papier récent, Strauch [90] a étendu les travaux de Dalalyan et Reiss en construisant des estimateurs adaptatives qui atteignent des vitesses de convergence rapides sur boules anisotropiques de Hölder. L'anisotropie joue un rôle importante du moment que les propretés de régularité des éléments d'un espace des fonctions peuvent dépendre de la direction de $\mathbb{R}^d$ choisie. L'école russe a commencé à considérer des espaces anisotropiques dès le début de la théorie des espaces de fonctions dans les années 1950 - 1960. Cependant, la question de la sélection de la fenêtre optimale à partir des observations i.i.d. par rapport au risque en norme sup n'a pas été complètement résolu jusqu'aux développements plutôt récents contenus dans [60].
La méthodologie détaillée dans Goldenshluger and Lepski [39] a inspiré la procédure de sélection de la fenêtre de l'estimateur à noyau, guidée par les données, qui a été proposé par des auteurs différents comme Strauch dans [90]; Comte, Prieur and Samson dans [23] et il nous fournit aussi le point de départ pour l'étude de la procédure adaptative.

Strauch, en particulier, considère toujours un enregistrement continu des observations $(X_t^T)_{0 \leq t \leq T}$ d'un processus d-dimensionnel jusqu'à l'instant T, solution forte de l'équation différentielle stochastique (32). La fonction de dérive est toujours de la forme $b = -\nabla V$, où $V \in \mathcal{C}^2(\mathbb{R}^d)$ est le potentiel. Elle suppose aussi que (34) soit vraie.
Elle construit des estimateurs adaptatifs sur des boules anisotropiques de Hölder; ces boules sont définies de la façon suivante.

Definition 10. *Soit $\beta = (\beta_1, ..., \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$. Une fonction $g : \mathbb{R}^d \to \mathbb{R}$ appartient à la classe anisotropique d' Hölder $\mathcal{H}_d(\beta, \mathcal{L})$ de fonctions si, pour chaque $i \in \{1, ..., d\}$,*

$$\left\| D_i^k g \right\|_\infty \leq \mathcal{L}_i \qquad \forall k = 0, 1, ..., \lfloor \beta_i \rfloor,$$

$$\left\| D_i^{\lfloor \beta_i \rfloor} g(. + te_i) - D_i^{\lfloor \beta_i \rfloor} g(.) \right\|_\infty \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \qquad \forall t \in \mathbb{R},$$

pour $D_i^k g$ qui dénote la dérivée partielle d'ordre k de g par rapport à la composante i, $\lfloor \beta_i \rfloor$ denote le plus grand nombre entier qui soit strictement plus petit que β_i et $e_1, ..., e_d$ dénote la base canonique en $\mathbb{R}^d$.

Strauch note par $\tilde{\mathbb{H}}_d(\beta, \mathcal{L})$ l'ensemble de fonctions de densité $\mu = \mu_b \in \mathcal{H}_d(\beta + 1, \mathcal{L})$ qui sont associées à $b = -\nabla V \in \Sigma(M_1, M_2)$, ensemble de fonctions défini au dessus de (33).
Pour estimer une densité invariante $\mu_b \in \tilde{\mathbb{H}}_d(\beta, \mathcal{L})$ dans un point $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, elle définit un estimateur à noyau

$$\hat{\mu}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m}) du, \tag{39}$$

où $h = (h_1, \dots, h_d)$ est la fenêtre multi - dimensionnelle.
En suivant [25], elle utilise la décomposition biais -variance usuelle. Pour estimer la

variance, elle utilise les inégalités de Poincaré et de Nash lesquelles mènent l'auteur à des conditions analogues aux conditions 1 et 2 proposés par [25] et ici reportés au dessous de (35).
Strauch étend donc (38) au cadre anisotropique en trouvant

$$\mathbb{E}_b[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2] \lesssim \sum_{r=1}^{d} h_r^{2(\beta_r+1)} + \frac{1}{T}\psi_d^2(\prod_{l=1}^{d} h_l),$$

où $\psi_d(x)$ est la fonction définie dans (37).
Soit $\beta = (\beta_1, \dots, \beta_d)$, avec $\beta_r > 0$ pour chaque $r \in \{1, \dots, d\}$, nous définissons la régularité moyenne $\overline{\beta + 1}$ de $\beta + 1$ comme

$$\frac{1}{\overline{\beta+1}} := \frac{1}{d}\sum_{r=1}^{d} \frac{1}{\beta_r + 1}.$$

La choix optimal de la fenêtre $h = h(T)$ donne les vitesses suivantes:

$$\mathbb{E}_b[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2]^{\frac{1}{2}} \lesssim \begin{cases} T^{\frac{-1}{2}}(\log T)^2, & d = 2 \\ T^{-\frac{\overline{\beta+1}}{2\overline{\beta+1}+d-2}}, & d \geq 3. \end{cases} \tag{40}$$

En consequence de (40) il n'y a clairement aucun gain à implémenter une procédure de sélection de la fenêtre dans le cas mono dimensionnel et bi-dimensionnel. Strauch considère donc l'adaptivité dans le cadre $d \geq 3$. Elle définit un ensemble $\mathcal{H}_t$ de fenêtres candidates et un ensemble associé $\mathcal{F}(\mathcal{H}_t)$ des estimateurs à noyau candidats:

$$\mathcal{F}(\mathcal{H}_t) := \left\{ \hat{\mu}_{h,T}(x) = \frac{1}{T\prod_{l=1}^{d} h_l} \int_0^T \prod_{m=1}^{d} K(\frac{x_m - X_u^m}{h_m})du, \quad x \in \mathbb{R}^d, \quad h \in \mathcal{H}_t \right\}. \tag{41}$$

Strauch propose une procédure pour sélectionner un estimateur de $\mathcal{F}(\mathcal{H}_t)$ qui est basée sur l'approche de [39]. En particulier, elle montre que l'estimateur sélectionné $\hat{\mu}_{\hat{h}} \in \mathcal{F}(\mathcal{H}_t)$ satisfait, pour chaque $T > 0$,

$$\mathbb{E}_b[\|\hat{\mu}_{\hat{h}} - \mu_b(x)\|_\infty^q]^{\frac{1}{q}} \leq c_1 \inf_{h \in \bar{\mathcal{H}}_T} \left\{ B_b(h) + \prod_{j=1}^{d} h_j^{\frac{1}{d}-\frac{1}{2}} \sqrt{\frac{\log T}{T}} \right\} + \frac{c_2}{\sqrt{T}}, \tag{42}$$

où c_1 et c_2 sont deux constants, $\bar{\mathcal{H}}_T \subset \mathcal{H}_t$ dénote une grille dyadique et $B_b(\cdot)$ peut être vu comme une approximation de l'erreur de μ_b mesuré en norme sup. L'inégalité (42) est l'outil principal pour l'étude de la vitesse de convergence de l'estimateur adaptive $\hat{\mu}_{\hat{h}}$. En effet, Strauch montre aussi que, pour chaque $q \geq 1$,

$$\lim_{T\to\infty} \sup_{\mu \in \mathbb{H}_d(\beta, \mathcal{L}} \sup (\mathbb{E}_b[\varphi_T^{-q}(\beta + 1)\, \|\hat{\mu}_{\hat{h}} - \mu_b(x)\|_\infty^q])^{\frac{1}{q}} < \infty,$$

où $\varphi_T(\beta + 1) := (\frac{\log T}{T})^{\frac{\overline{\beta+1}}{2\overline{\beta+1}+d-2}}$.

0.4.3 Estimation de la mesure stationnaire pour une diffusion avec sauts

Dans le quatrième chapitre de la thèse nous considérons le problème de l'estimation de la mesure stationnaire pour le diffusion à sauts

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d\setminus\{0\}} \gamma(X_{s^-})z(\mu - \bar{\mu})(ds, dz), \quad t \in [0, T], \tag{43}$$

où $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$; $W = (W_t, t \geq 0)$ est un mouvement Brownien d- dimensionnel et μ est une mesure ponctuelle de Poisson sur $(0, \infty) \times \mathbb{R}^d$ associée au processus de Lévy $L = (L_t)_{t \geq 0}$, avec $L_t := \int_0^t \int_{\mathbb{R}^d} z(\mu(ds, dz) - \bar{\mu}(ds, dz))$. Nous supposons toujours qu'un enregistrement continu des observations $X^T = (X_t)_{0 \leq t \leq T}$ soit disponible.
En présence des sauts, nous n'avons connaissance que de quelques travaux dans le cadre non paramétrique. Par exemple, Schmisser étudie dans [84] l'estimation adaptative non paramétrique des coefficients d'un processus de diffusion avec sauts et avec Funke dans [34], l'estimation adaptative non paramétrique de la dérive d'un processus de diffusion à sauts intégré.

Nous fournissons dans le Chapitre 4 un estimateur non - paramétrique de la densité invariante μ sur des boules anisotropiques avec une procédure de sélection de la fenêtre complètement guidée par les données.
Nous proposons d'estimer μ avec un estimateur à noyau, nous devons donc introduire une fonction noyau $K : \mathbb{R} \to \mathbb{R}$. Un estimateur naturel de μ en $x \in \mathbb{R}^d$ dans le cadre anisotropique est, comme on l'a déjà vu,

$$\hat{\mu}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m}) du,$$

où $h = (h_1, ..., h_d)$ est une fenêtre multi -dimensionnelle qui peut être choisie à travers une procédure de sélection que l'on décrira.
Nous commençons en donnant deux définitions:

Definition 11. *1.Un processus X est dit exponentiellement ergodique s'il est ergodique, au sens de la Définition 3, et s'il y a des constantes positives c et ρ tels que*

$$\|P_t f\|_{L^1(\mu)} \leq c e^{-\rho t} \|f\|_\infty .$$

2. Un processus X est dit exponentiellement β - mixing s'il existe une constante $\gamma > 0$ telle que $\beta_X(t) = O(e^{-\gamma t})$ pour $t \to \infty$, où β_X est le coefficient de β - mixing du processus X défini dans la Section 1.3.2 de [27].

Nous trouvons des bornes sur le semi-groupe de transition et sur la densité de transition qui seront utiles pour trouver des bornes supérieurs sur la variance des fonctionnels intégrales de la diffusion X.

Résultat 14. *1. $\exists!$ densité de transition $p_t(x, y)$ pour laquelle, pour chaque $T \geq 0$, il y a $c_0 > 0$ et $\lambda_0 > 0$ tels que, pour chaque $t \in [0, T]$ et pour chaque couple de points $x, y \in \mathbb{R}^d$, nous avons*

$$|p_t(x, y)| \leq c_0 (t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{t}} + \frac{t}{(t^{\frac{1}{2}} + |y - x|)^{d+\alpha}}).$$

2. Le processus X est exponentiellement ergodique et exponentiellement β - mixing.

Nous soulignons que, si dans [25] et [90] les auteurs avaient besoin de supposer l'existence de la densité de transition et de la borner, nous dérivons ces propriétés à travers le Résultat 14 ci-dessus; les hypothèses qu'il nous faut sont exprimées directement sur le model (43).

De plus, nous ne supposons plus que la fonction de dérive soit de la forme $b = -\nabla V$, comme il l'était dans [25] et [90].

En utilisant les bornes contenus dans le Résultat 14, nous pouvons borner la variance de la façon suivante:

Résultat 15. *Soit $f : \mathbb{R}^d \to \mathbb{R}$ une fonction bornée et mesurable avec support $\mathcal{S}$ qui satisfait $|\mathcal{S}| < 1$. Alors il existe une constante C qui ne dépend pas de f telle que*

- $Var(\int_0^T f(X_t)dt) \le CT \left\| f \right\|_\infty^2 |\mathcal{S}|^2 (1 + (\log(\frac{1}{|\mathcal{S}|}))^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{|\mathcal{S}|}))$ *pour $d = 1$,*
- $Var(\int_0^T f(X_t)dt) \le CT \left\| f \right\|_\infty^2 |\mathcal{S}|^2 (1 + \log(\frac{1}{|\mathcal{S}|}))$ *pour $d = 2$,*
- $Var(\int_0^T f(X_t)dt) \le CT \left\| f \right\|_\infty^2 |\mathcal{S}|^{1+\frac{2}{d}}$ *pour $d \ge 3$.*

En utilisant la décomposition de biais -variance dans le cadre anisotropique (voir Proposition 1 dans [20]) nous trouvons la borne suivante:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^d h_l^{2\beta_l} + T^{-2} Var(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_t^m}{h_m})dt).$$

Nous voulons borner la variance ci-dessus en utilisant le Résultat 15 sur la fonction $f(y) := \frac{1}{\prod_{l=1}^d h_l} \prod_{m=1}^d K(\frac{x_m - y_m}{h_m})$, qui nous amène aux vitesses de convergence suivantes pour l'estimation ponctuelle de la densité invariante de notre diffusion avec sauts:

Résultat 16.

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \begin{cases} \frac{(\log T)^{(2-\frac{(1+\alpha)}{2}) \vee 1}}{T} & pour\ d = 1, \\ \frac{\log T}{T} & pour\ d = 2, \\ T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} & pour\ d \ge 3, \end{cases}$$

où $\alpha \in (0, 2)$ est l'indice d'activité des sauts du processus de Lévy et $\bar{\beta}$ est la moyenne harmonique de la régularité sur les d dimensions différentes définie, en analogie avec $\overline{\beta + 1}$, de la façon suivante:

$$\frac{1}{\bar{\beta}} = \frac{1}{d} \sum_{r=1}^d \frac{1}{\beta_r}.$$

Nous remarquons que la vitesse que nous trouvons pour $d \ge 3$ est la même trouvée par Strauch dans [90] en absence de sauts, qui est aussi la vitesse trouvée dans [25] à condition de replacer la régularité moyenne avec β, la régularité commune dans le d dimensions.
Le cas $d = 1$ met en évidence la différence principale entre les deux cadres avec et sans sauts. En effet, si dans le cas continu la vitesse de convergence optimale était $\frac{1}{T}$, maintenant la vitesse que nous trouvons dépend de α et est entre $\frac{\log T}{T}$ et $\frac{(\log T)^{\frac{3}{2}}}{T}$. Nous soulignons ici que cette vitesse de convergence n'est pas nécessairement la vitesse optimale. En réalité dans le cas continu des approches différentes, comme le temps local de la diffusion, avaient été utilisées pour obtenir la vitesse optimale $\frac{1}{T}$; nous n'excluons pas la possibilité que aussi en présence de sauts l'implémentation

d'autres méthodes puisse conduire à une vitesse de convergence plus rapide de celle présentée ci-dessus pour le cadre mono -dimensionnel.
Pour compléter le comparaison avec le cadre continu, nous rappelons que dans [25] et [90] la vitesse de convergence trouvée dans le cas $d = 2$ était $\frac{(\log T)^4}{T}$ et, donc, la convergence de l'estimateur semble être plus rapide en présence de sauts que en leur absence. Le motif pour lequel cela se passe est que, pour trouver la vitesse de convergence, la densité de transition $(p_t)_{t\in\mathbb{R}^+}$ doit être bornée supérieurement. Dans [25] les auteurs supposent avoir $p_t(x, y) \le c(t^{-\frac{d}{2}} + t^{\frac{3d}{2}})$ et dans [90] les inégalités de Nash et Poincaré mènent l'auteur à une borne analogue à celle présentée dans [25]. Cependant le Résultat 14 nous donne une borne differente qui nous conduit à une vitesse différente. Toutefois, en absence du terme $t^{\frac{3d}{2}}$ dans la condition (36), qui se passe par exemple en considérant une fonction de dérive bornée, aussi dans le cadre continu la vitesse de convergence se trouve être, comme dans le cas de diffusion avec sauts, égal à $\frac{\log T}{T}$.

Après avoir donné les vitesses de convergence des estimateurs nous proposons, dans le cas $d \ge 3$, une procédure de sélection de la fenêtre de l'estimateur à noyau totalement guidée par les données, inspirée par la méthodologie détaillée dans Goldenshluger et Lespki [39]. Le méthode a l'avantage d'être anisotropique: les fenêtres sélectionnées dans le diverses directions sont en général différentes, qui est cohérent avec la possibilité d'avoir des régularités différentes dans les diverses directions.

Nous introduisons un ensemble de fenêtres potentielles $\mathcal{H}_T$ et un ensemble associé d'estimateurs potentiels $\mathcal{F}(\mathcal{H}_T)$, défini comme dans (41).
Nous dénotons, pour $y \in \mathbb{R}^d$,

$$\mathbb{K}_h(y) := \prod_{l=1}^{d} \frac{1}{h_l} \prod_{m=1}^{d} K(\frac{y_m}{h_m}). \tag{44}$$

Nous écrirons, donc, $\hat{\mu}_h(x) := \frac{1}{T}\int_0^T \mathbb{K}_h(X_u - x)du$.
Notre objectif est de sélectionner un estimateur dans la famille $\mathcal{F}(\mathcal{H}_T)$ avec une procédure guidée par les données, basée sur l'observation de la trajectoire continue du processus X jusqu'à l'instant T.
Nous décrivons maintenant cette procédure, qui est basée sur les estimateurs auxiliaires qui reposent sur l'opérateur de convolution. À notre connaissance, ce méthode a été introduit dans [61] comme système pour contourner l'absence d'ordre parmi les estimateurs dans le cas anisotropique, où l'augmentation de la variance d'un estimateur n'implique pas la diminution de son biais.
Pour n'importe quelles fenêtres $h = (h_1, ..., h_d)^T$, $\eta = (\eta_1, ..., \eta_d)^T \in \mathcal{H}_T$ et $x \in \mathbb{R}^d$, nous définissons

$$\mathbb{K}_h * \mathbb{K}_\eta(x) := \prod_{j=1}^{d} (K_{h_j} * K_{\eta_j})(x_j) = \prod_{j=1}^{d} \int_{\mathbb{R}} K_{h_j}(u - x_j) K_{\eta_j}(u) du.$$

Nous définissons aussi les estimateurs à noyau

$$\hat{\mu}_{h,\eta}(x) := \frac{1}{T}\int_0^T (\mathbb{K}_h * \mathbb{K}_\eta)(X_u - x)du, \quad x \in \mathbb{R}^d.$$

Nous remarquons que, de par la définition des estimateurs à noyau et puisque la convolution est commutative, nous avons $\hat{\mu}_{h,\eta} = \hat{\mu}_{\eta,h}$.
La procédure de sélection proposée se base sur la comparaison des différences $\hat{\mu}_{h,\eta} -$

$\hat{\mu}_\eta$.
Nous définissons

$$A(h) := \sup_{\eta \in \mathcal{H}_T} (\|\hat{\mu}_{h,\eta} - \hat{\mu}_\eta\|_A^2 - V(\eta))_+, \tag{45}$$

où avec $\|.\|_A^2$ nous dénotons la norme L^2 sur A, un sous ensemble compact de $\mathbb{R}^d$, et avec

$$V(h) := \frac{k}{T} (\prod_{l=1}^d h_l)^{\frac{2}{d}-1}.$$

La constante k est une constante numérique qui est grande. En particulier, il est suffisant de la choisir plus grande que les constantes $2k_0^*$ et $2k_0$ qui apparaissent dans le Lemme 38 du Chapitre 4. Même si k n'est pas explicite, il peut être calibré à travers des simulations comme cela a été fait par exemple dans la Section 5 de [23] en utilisant l'implémentation d'une méthode inspiré par Lacour, Massart et Rivoirard dans [59].
Heuristiquement, $A(h)$ est une estimation du carré du biais et $V(h)$ une borne de la variance. Nous soulignons que le terme de pénalité $V(h)$ qui est utilisé ici vient du Résultat 15, pour la fonction f qui est maintenant la fonction à noyau $\mathbb{K}_h$.
Donc, la sélection est faite en définissant

$$\tilde{h} := \arg \min_{h \in \mathcal{H}_T} (A(h) + V(h)). \tag{46}$$

Nous introduisons la notation suivante: $\mu_h := \mathbb{K}_h * \mu$, qui est la fonction estimée sans biais par $\hat{\mu}_h$, c'est à dire $\mathbb{E}[\hat{\mu}_h(x)] = \mu_h(x)$.
De plus, nous définissons un terme de biais $B(h) := \|\mu_h - \mu\|_{\tilde{A}}^2$, où nous avons dénoté comme $\|.\|_{\tilde{A}}$ la norme L^2 sur $\tilde{A}$, un ensemble compact de $\mathbb{R}^d$ qui est tel que $\tilde{A} := \left\{\zeta \in \mathbb{R}^d : d(\zeta, A) \leq 2\sqrt{d}\right\}$.
Nous avons alors le résultat suivant

Résultat 17.

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 \inf_{h \in \mathcal{H}_T} (B(h) + V(h)) + c_1 e^{-c_2 (\log T)^2},$$

pour c_1 et c_2 des constantes positives.

La borne énoncée dans le Résultat 17 montre que l'estimateur conduit à un trade - off automatique entre le biais $\|\mu_h - \mu\|_{\tilde{A}}^2$ et la variance $V(h)$, à une constante multiplicative c_1 près. Le dernier terme est en effet négligeable.

Nous remarquons que le Résultat 16 nous fournit la choix optimale pour la fenêtre $h(T)$ pour $d \geq 3$, qui est $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}}$.
Pour ce $h(T)$ appartenant à la classe de fenêtres potentielles $\mathcal{H}_T$, nous avons le résultat suivant:

Résultat 18.

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 (\frac{1}{T})^{\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2 (\log T)^2},$$

pour c_1 et c_2 des constantes positives.

En soulignant encore une fois que le dernier terme est négligeable comparé au premier, nous avons que le risque obtenu en utilisant la fenêtre fournie par notre procédure de sélection converge à zéro rapidement. En particulier, sa vitesse de convergence coïncide avec celle optimale fourni de [25] et [90] dans le cas sans sauts.

Part I

Contrast function estimation for the drift and volatility parameters of ergodic jump diffusion processes

Introduction

In the first part of this thesis we suppose to observe the process $(X_{t_i})_{i=0,...n}$, where the discretization step $\Delta_n := \max_i(t_{i+1} - t_i)$ goes to 0. X is the solution of the stochastic differential equation (1) for $d = 1$, with the drift and the volatility coefficients which depend on two unknown parameters that we will denote μ and σ, respectively. The main goal is to improve the existing literature by proposing a contrast function which can remove the restrictive conditions on the discretization step, so that it can go to 0 arbitrarily slowly.

Shimizu uses in [87] a contrast function to estimate the drift, volatility and jumps parameters. The asymptotic normality of the estimators is obtained under some conditions on the discretization step which are more and more restrictive when the jump intensity is high in the neighborhood of zero. In the most favourable case, which corresponds to a finite jump intensity, the condition is $n\Delta_n^2 \to 0$ and, for α next to 1, it becomes $n\Delta_n \to 0$, in contradiction with $n\Delta_n \to \infty$. In [38] the condition on the step is weaken, for the drift estimation only, and it becomes $n\Delta_n^3 \to 0$ in the case of finite intensity. These two works assume that the jumps of the process are summable ($\alpha \leq 1$). Considering a model without jumps, in order to obtain some conditions less restrictive than $n\Delta_n^3 \to 0$, it is necessary to introduce some corrections of the contrast function as in Kessler [51].

In the first chapter we propose a correction of the Shimizu's contrast [87] which allows us to estimate the drift parameter without requiring any condition on the rate at which Δ_n converges to 0. We extend the results of [88] and [38] as well by suppressing the assumption that the jumps of the process are summable.
Moreover, in the case where the intensity is finite, we can propose an explicit correction of the Shimizu's contrast function and we connect it to Kessler's correction. The chapter is based on the paper "Contrast function estimation for the drift parameter of ergodic jump diffusion process" [3], published in Scandinavian Journal of Statistics.

The second chapter regards the joint estimation of the volatility and drift parameters in a similar framework, under the condition that the jump intensity is finite. The joint estimation of the two parameters introduces some significant difficulties: since the drift and the volatility parameters aren't estimated at the same rate, we have to deal with asymptotic properties in two different regimes. Comparing to earlier results ([88], [87], [38], [3]), the sampling step $(t_i^n)_{i=0,...,n}$ can be irregular, no condition is needed on the rate at which $\Delta_n \to 0$ and the parameters of drift and diffusion are jointly estimated. The chapter relies on the submitted paper "Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function" [4], under revision for Statistical Inference for Stochastic Processes.

Chapter 1

Contrast function estimation for the drift parameter of ergodic jump diffusion processes

Abstract :

In this chapter we consider an ergodic diffusion process with jumps whose drift coefficient depends on an unknown parameter θ*. We suppose that the process is discretely observed at the instants* $(t_i^n)_{i=0,\dots,n}$ *with* $\Delta_n = \sup_{i=0,\dots,n-1}(t_{i+1}^n - t_i^n) \to 0$*. We introduce an estimator of* θ*, based on a contrast function, which is efficient without requiring any conditions on the rate at which* $\Delta_n \to 0$*, and where we allow the observed process to have non summable jumps. This extends earlier results where the condition* $n\Delta_n^3 \to 0$ *was needed (see [38],[88]) and where the process was supposed to have summable jumps. In general situations, our contrast function is not explicit and one has to resort to some approximation. In the case of a finite jump activity, we propose explicit approximations of the contrast function, such that the efficient estimation of* θ *is feasible under the condition that* $n\Delta_n^k \to 0$ *where* $k > 0$ *can be arbitrarily large. This extends the results obtained by Kessler [51] in the case of continuous processes.*

Keys words : Efficient drift estimation, ergodic properties, high frequency data, Lévy-driven SDE, thresholding methods.

1.1 Introduction

Diffusion processes with jumps have been widely used to describe the evolution of phenomenon arising in various fields. In finance, jump-processes were introduced to model the dynamic of asset prices ([70],[56]), exchange rates ([11]), or volatility processes ([8],[31]). Utilization of jump-processes in neuroscience can be found for instance in [26].

Practical applications of these models has lead to the recent development of many statistical methods. In this chapter, our aim is to estimate the drift parameter θ from a discrete sampling of the process X^θ solution to

$$X_t^\theta = X_0^\theta + \int_0^t b(\theta, X_s^\theta)ds + \int_0^t \sigma(X_s^\theta)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s-}^\theta) z \tilde{\mu}(ds, dz),$$

where W is a one dimensional Brownian motion and $\tilde{\mu}$ a compensated Poisson random measure, with a possible infinite jump activity. We assume that the process is sampled at the times $(t_i^n)_{i=0,\dots,n}$ where the sampling step $\Delta_n := \sup_{i=0,\dots,n-1} t_{i+1}^n - t_i^n$ goes to zero. Due to the presence of a Gaussian component, we know that it is impossible to estimate the drift parameter on a finite horizon of time. Thus, we assume that $t_n \to \infty$ and the ergodicity of the process X^θ.

Generally, the main difficulty while considering statistical inference of discretely observed stochastic processes comes from the lack of explicit expression for the likelihood. Indeed, the transition density of a jump-diffusion process is usually unknown explicitly. Several methods have been developed to circumvent this difficulty. For instance, closed form expansions of the transition density of jump-diffusions is studied in [2], [62]. In the context of high frequency observation, the asymptotic behaviour of estimating functions are studied in [50], and conditions are given to ensure rate optimality and efficiency. Another approach, fruitful in the case of high frequency observation, is to consider pseudo-likelihood method, for instance based on the high frequency approximation of the dynamic of the process by the one of the Euler scheme. This leads to explicit contrast functions with Gaussian structures (see e.g. [88],[87],[68]).

The validity of the approximation by the Euler pseudo-likelihood is justified by the high frequency assumption of the observations, and actually proving that the estimators are asymptotic normal usually necessitates some conditions on the rate at which Δ_n should tend to zero. For applications, it is important that the condition on $\Delta_n \to 0$ is less stringent as possible.

In the case of continuous processes, Florens-Zmirou [33] proposes estimation of drift and diffusion parameters under the fast sampling assumption $n\Delta_n^2 \to 0$. Yoshida [95] suggests a correction of the contrast function of [33] that yields to the condition $n\Delta_n^3 \to 0$. In Kessler [51], the author introduces an explicit modification of the Euler scheme contrast such that the associated estimators are asymptotically normal, under the condition $n\Delta_n^k \to 0$ where $k \geq 2$ is arbitrarily large. Hence, the result by Kessler allows for any arbitrarily slow polynomial decay to zero of the sampling step.

In the case of jump-diffusions, Shimizu [87] proposes parametric estimation of drift, diffusion and jump coefficients. The asymptotic normality of the estimators are obtained under some explicit conditions relating the sampling step and jump intensity of the process. These conditions on Δ_n are more restrictive as the intensity of jumps near zero is high. In the situation where this jump intensity is finite, the conditions of [87] reduces to $n\Delta_n^2 \to 0$. In [38], the condition on the sampling step is relaxed to $n\Delta_n^3 \to 0$, when one estimates the drift parameter only.

In this chapter, we focus on the estimation of the drift parameter, and our aim is to weaken the conditions on the decay of the sampling step in way comparable to Kessler's work [51], but in the framework of jump-diffusion processes.

One of the idea in Kessler's paper is to replace, in the Euler scheme contrast function, the contribution of the drift by the exact value of the first conditional moment $m^{(1)}_{\theta,t_i,t_{i+1}}(x) = E[X^\theta_{t_{i+1}} \mid X^\theta_{t_i} = x]$ or some explicit approximation with arbitrarily high order when $\Delta_n \to 0$. In presence of jumps, the contrasts functions in [88] (see also [87], [38]) resort to a filtering procedure in order to suppress the contribution of jumps and recover the continuous part of the process. Based on those ideas, we introduce a contrast function (see Definition 13), whose expression relies on the quantity $m_{\theta,t_i,t_{i+1}}(x) = \frac{E[X^\theta_{t_{i+1}}\varphi((X^\theta_{t_{i+1}}-X^\theta_{t_i})/(t_{i+1}-t_i)^\beta)|X^\theta_{t_i}=x]}{E[\varphi((X^\theta_{t_{i+1}}-X^\theta_{t_i})/(t_{i+1}-t_i)^\beta)|X^\theta_{t_i}=x]}$, where φ is some compactly supported function and $\beta < 1/2$. The function φ is such that $\varphi((X^\theta_{t_{i+1}} - X^\theta_{t_i})/(t_{i+1} - t_i)^\beta)$ vanishes when the increments of the data are too large compared to the typical increments of a continuous diffusion process, and thus can be used to filter the contribution of the jumps.

The main result of this chapter is that the associated estimator converges at rate $\sqrt{t_n}$, with some explicit asymptotic variance and is efficient. Comparing to earlier results ([88], [87], [38]), the sampling step $(t_i^n)_{i=0,\dots,n}$ can be irregular, no condition is needed on the rate at which $\Delta_n \to 0$ and we have suppressed the assumption that the jumps of the process are summable. Let us stress that when the jumps activity is so high that the jumps are not summable, we have to choose $\beta < 1/3$ (see Assumption A_β).

Moreover, in the case where the intensity is finite and with the specific choice of φ being an oscillating function, we prove that we can approximate our contrast function by a completely explicit one, exactly as in the paper by Kessler [51]. This yields to an efficient estimator under the condition $n\Delta_n^k \to 0$, where k is related to the oscillating properties of the function φ. As k can be chosen arbitrarily high, up to a proper choice of φ, our method allows to estimate efficiently the drift parameter, under the assumption that the sampling step tends to zero at some polynomial rate. We also show numerically that, when the jump activity is finite, the estimator we deduce from the explicit approximation of the contrast function performs well, making the bias visibly reduced.
On the other side, considering the case of infinite jumps activity (taking in particular a tempered α-stable jump process with $\alpha < 1$), we implement our main results building an approximation of m (see Theorem 3 below) from which we deduce an approximation of the contrast that we minimize in order to get the estimator of the drift coefficient. The estimator we found is a corrected version of the estimator that would result from the choice of an Euler scheme approximation. We see numerically that our estimator is well-performed and that the correction term we give drastically reduces the bias, especially as α gets bigger.

The outline of the chapter is the following. In Section 3.2 we present the assumptions on the process X. The Section 3.3 contains the main results of the paper: in Section 1.3.1 we define the contrast function while the consistency and asymptotic normality of the estimator are stated in Section 1.3.2. In Section 1.4 we explain how to use in practice the contrast function and so we deal with its approximations in Section 1.4.1 while its explicit modification is presented in the case of finite jump activity in Section 1.4.2. The Section 1.5 is devoted to numerical results and perspectives for practical applications. In Section 1.6 we state limit theorems useful to study the asymptotic behavior of the contrast function. The proofs of the main statistical results are given in Section 1.7, while the proofs of the limit theorems and some technical results are presented in the Appendix.

1.2 Model, assumptions

Let Θ be a compact subset of $\mathbb{R}$ and X^θ a solution to

$$X_t^\theta = X_0^\theta + \int_0^t b(\theta, X_s^\theta)ds + \int_0^t a(X_s^\theta)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}^\theta) z\tilde{\mu}(ds,dz), \quad t \in \mathbb{R}_+, \tag{1.1}$$

where $W = (W_t)_{t\geq 0}$ is a one dimensional Brownian motion, μ is a Poisson random measure associated to the Lévy process $L = (L_t)_{t\geq 0}$, with $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds,dz)$ and $\tilde{\mu} = \mu - \bar{\mu}$ is the compensated one, on $[0,\infty) \times \mathbb{R}$. We denote $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which W and μ are defined.
We suppose that the compensator has the following form: $\bar{\mu}(dt,dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.
The initial condition X_0^θ, W and L are independent.

1.2.1 Assumptions

We suppose that the functions $b : \Theta \times \mathbb{R} \to \mathbb{R}$, $a : \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ satisfy the following assumptions:

ASSUMPTION 1: *The functions $a(x)$, $\gamma(x)$ and, for all $\theta \in \Theta$, $b(x,\theta)$ are globally Lipschitz. Moreover, the Lipschitz constant of b is uniformly bounded on Θ.*

Under Assumption 1 the equation (1.1) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [7] (Theorems 6.2.9. and 6.4.6.).

ASSUMPTION 2: *For all $\theta \in \Theta$ there exists a constant $t > 0$ such that X_t^θ admits a density $p_t^\theta(x,y)$ with respect to the Lebesgue measure on $\mathbb{R}$; bounded in $y \in \mathbb{R}$ and in $x \in K$ for every compact $K \subset \mathbb{R}$. Moreover, for every $x \in \mathbb{R}$ and every open ball $U \in \mathbb{R}$, there exists a point $z = z(x,U) \in supp(F)$ such that $\gamma(x)z \in U$.*

The last assumption was used in [67] to prove the irreducibility of the process X^θ. Other sets of conditions, sufficient for irreducibility, are in [67].

ASSUMPTION 3 (Ergodicity):

1. *For all $q > 0$, $\int_{|z|>1} |z|^q F(z)dz < \infty$.*
2. *For all $\theta \in \Theta$ there exists $C > 0$ such that $xb(x,\theta) \leq -C|x|^2$, if $|x| \to \infty$.*
3. *$|\gamma(x)|/|x| \to 0$ as $|x| \to \infty$.*
4. *$|a(x)|/|x| \to 0$ as $|x| \to \infty$.*
5. *$\forall \theta \in \Theta$, $\forall q > 0$ we have $\mathbb{E}|X_0^\theta|^q < \infty$.*

Assumption 2 ensures, together with the Assumption 3, the existence of unique invariant distribution π^θ, as well as the ergodicity of the process X^θ, as stated in the Lemma 26 below.

ASSUMPTION 4 (Jumps):

1. *The jump coefficient* γ *is bounded from below, that is*

$$\inf_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$$

2. *The Lévy measure* F *is absolutely continuous with respect to the Lebesgue measure and we denote* $F(z) = \frac{F(dz)}{dz}$.
3. *We suppose that* $\exists\, c > 0$ *s.t., for all* $z \in \mathbb{R}$, $F(z) \le \frac{c}{|z|^{1+\alpha}}$, *with* $\alpha \in (0, 2)$.

Assumptions 4.1 is useful to compare size of jumps of X and L.

ASSUMPTION 5 (Non-degeneracy): *There exists some* $\alpha > 0$, *such that* $a^2(x) \ge \alpha$ *for all* $x \in \mathbb{R}$

The Assumption 5 ensures the existence of the contrast function defined in Section 3.1.

ASSUMPTION 6 (Identifiability): *For all* $\theta \neq \theta'$,$(\theta, \theta') \in \Theta^2$,

$$\int_{\mathbb{R}} \frac{(b(\theta, x) - b(\theta', x))^2}{a^2(x)} d\pi^\theta(x) > 0$$

We can see that this last assumption is equivalent to

$$\forall \theta \neq \theta', \qquad (\theta, \theta') \in \Theta^2, \qquad b(\theta, .) \neq b(\theta', .). \tag{1.2}$$

We also need the following technical assumption:

ASSUMPTION 7:

1. The derivatives $\frac{\partial^{k_1+k_2} b}{\partial x^{k_1} \partial \theta^{k_2}}$, with $k_1 + k_2 \le 4$ and $k_2 \le 3$, exist and they are bounded if $k_1 \ge 1$. If $k_1 = 0$, for each $k_2 \le 3$ they have polynomial growth.
2. The derivatives $a^{(k)}(x)$ exist and they are bounded for each $1 \le k \le 4$.
3. The derivatives $\gamma^{(k)}(x)$ exist and they are bounded for each $1 \le k \le 4$.

Define the asymptotic Fisher information by

$$I(\theta) = \int_{\mathbb{R}} \frac{(\dot{b}(\theta, x))^2}{a^2(x)} \pi^\theta(dx). \tag{1.3}$$

ASSUMPTION 8: *For all* $\theta \in \Theta$, $I(\theta) > 0$.

Remark 1. *If $\alpha < 1$, using Assumption 4.3 the stochastic differential equation (1.1) can be rewritten as follows:*

$$X_t^\theta = X_0^\theta + \int_0^t \bar{b}(\theta, X_s^\theta)ds + \int_0^t a(X_s^\theta)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}^\theta)z\mu(ds, dz), \quad t \in \mathbb{R}_+, \tag{1.4}$$

where $\bar{b}(\theta, X_s^\theta) = b(\theta, X_s^\theta) - \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}^\theta)zF(z)dz$.
This expression implies that X follows diffusion equation $X_t^\theta = X_0^\theta + \int_0^t \bar{b}(\theta, X_s^\theta)ds + \int_0^t a(X_s^\theta)dW_s$ in the interval in which no jump occurred.

From now on we denote the true parameter value by θ_0, an interior point of the parameter space Θ that we want to estimate. We shorten X for X^{θ_0}.
We will use some moment inequalities for jump diffusions, gathered in the following lemma:

Lemma 1. *Let X satisfies Assumptions 1-4. Let $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$ and let $\mathcal{F}_s := \sigma\left\{(W_u)_{0<u\leq s}, (L_u)_{0<u\leq s}, X_0\right\}$.*
Then, for all $t > s$,
1) for all $p \geq 2$, $\mathbb{E}[|X_t - X_s|^p]^{\frac{1}{p}} \leq c|t-s|^{\frac{1}{p}}$,
2) for all $p \geq 2$, $p \in \mathbb{N}$, $\mathbb{E}[|X_t - X_s|^p|\mathcal{F}_s] \leq c|t-s|(1 + |X_s|^p)$.
3) for all $p \geq 2$, $p \in \mathbb{N}$, $\sup_{h\in[0,1]} \mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] \leq c(1 + |X_s|^p)$.

The first two points follow from Theorem 66 of [80] and Proposition 3.1 in [88]. The last point is a consequence of the second one: $\forall h \in [0, 1]$,

$$\mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] = \mathbb{E}[|X_{s+h} - X_s + X_s|^p|\mathcal{F}_s] \leq c(\mathbb{E}[|X_{s+h} - X_s|^p|\mathcal{F}_s] + \mathbb{E}[|X_s|^p|\mathcal{F}_s]),$$

where c may change value line to line. Using the second point of Lemma 25 and the measurability of X_s with respect to $\mathcal{F}_s$, it is upper bounded by $c|h|(1+|X_s|^p)+c|X_s|^p$. Therefore

$$\sup_{h\in[0,1]} \mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] \leq \sup_{h\in[0,1]} c|h|(1 + |X_s|^p) + c|X_s|^p \leq c(1 + |X_s|^p).$$

1.2.2 Ergodic properties of solutions

An important role is playing by ergodic properties of solution of equation (1.1)
The following Lemma states that Assumptions $1-4$ are sufficient for the existence of an invariant measure π^θ such that an ergodic theorem holds and moments of all order exist.

Lemma 2. *Under assumptions 1 to 4, for all $\theta \in \Theta$, X^θ admits a unique invariant distribution π^θ and the ergodic theorem holds:*

1. *For every measurable function $g : \mathbb{R} \to \mathbb{R}$ satisfying $\pi^\theta(g) < \infty$, we have a.s.*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t g(X_s^\theta)ds = \pi^\theta(g).$$

2. *For all $q > 0$, $\pi^\theta(|x|^q) < \infty$.*

3. *For all $q > 0$, $\sup_{t\geq 0} \mathbb{E}[|X_t^\theta|^q] < \infty$.*

A proof is in [38] (Section 8 of Supplement) in the case $\alpha \in (0,1)$, the proof relies on [67]. In order to use it also in the case $\alpha \geq 1$ we have to show that, taken $q > 2$ q even and $f^\star(x) = |x|^q$, $f^\star$ satisfies the drift condition $Af^\star = A_d f^\star + A_c f^\star \leq -c_1 f^\star + c_2$, where $c_1 > 0$ and $c_2 > 0$.
Using Taylor's formula up to second order we have

$$|A_d f^\star(x)| \leq c \int_{\mathbb{R}} \int_0^1 |z|^2 \left\|\gamma\right\|_\infty |f''^\star(x + sz\gamma(y))| F(z) ds dz =$$

$$= c \int_{\mathbb{R}} \int_0^1 |z|^2 \left\|\gamma\right\|_\infty q(q-1)|x + sz\gamma(y)|^{q-2} F(z) ds dz = o(|x|^q). \tag{1.5}$$

Concerning the generator's continuous part, we use the second point of Assumption 3 to get

$$A_c f^\star(x) = \frac{1}{2}\sigma^2(x) q(q-1) x^{q-2} + b(\theta, x) q\, x\, x^{q-2} \leq o(|x|^q) - cq|x|^2 x^{q-2} \leq o(|x|^q) - cf^\star(x). \tag{1.6}$$

By (1.5) and (1.6), the drift condition holds.

1.3 Construction of the estimator and main results

We exhibit a contrast function for the estimation of a parameter in the drift coefficient. We prove that the derived estimator is consistent and asymptotically normal.

1.3.1 Construction of the estimator

Let X^θ be the solution to (1.1). Suppose that we observe a finite sample

$$X_{t_0}, ..., X_{t_n}; \qquad 0 = t_0 \leq t_1 \leq ... \leq t_n,$$

where X is the solution to (1.1) with $\theta = \theta_0$. Every observation time point depends also on n, but to simplify the notation we suppress this index. We will be working in a high-frequency setting, i.e.

$$\Delta_n := \sup_{i=0,...,n-1} \Delta_{n,i} \longrightarrow 0, \quad n \to \infty,$$

with $\Delta_{n,i} := (t_{i+1} - t_i)$.
We assume $\lim_{n\to\infty} t_n = \infty$ and $n\Delta_n = O(t_n)$ as $n \to \infty$.
We introduce a jump filtered version of the gaussian quasi-likelihood. This leads to the following contrast function:

Definition 12. *For $\beta \in (0, \frac{1}{2})$ and $k > 0$, we define the contrast function $U_n(\theta)$ as follows:*

$$U_n(\theta) := \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m_{\theta,t_i,t_{i+1}}(X_{t_i}))^2}{a^2(X_{t_i})(t_{i+1} - t_i)} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} \tag{1.7}$$

where

$$m_{\theta,t_i,t_{i+1}}(x) := \frac{\mathbb{E}[X_{t_{i+1}}^\theta \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]} \tag{1.8}$$

and

$$\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) = \varphi(\frac{X_{t_{i+1}} - X_{t_i}}{\Delta_{n,i}^{\beta}}),$$

with φ a smooth version of the indicator function, such that $\varphi(\zeta) = 0$ for each ζ, with $|\zeta| \geq 2$ and $\varphi(\zeta) = 1$ for each ζ, with $|\zeta| \leq 1$.
The last indicator aims to avoid the possibility that $|X_{t_i}|$ is big. The constant k is positive and it will be choosen later, related to the development of $m_{\theta,t_i,t_{i+1}}(x)$ (cf. Remark 2 below).
Moreover we define

$$m_{\theta,h}(x) := \frac{\mathbb{E}[X_h^{\theta}\varphi_{h^{\beta}}(X_h^{\theta} - X_0^{\theta})|X_0^{\theta} = x]}{\mathbb{E}[\varphi_{h^{\beta}}(X_h^{\theta} - X_0^{\theta})|X_0^{\theta} = x]}.$$

By the homogeneity of the equation we get that $m_{\theta,t_i,t_{i+1}}(x)$ depends only on the difference $t_{i+1} - t_i$ and so $m_{\theta,t_i,t_{i+1}}(x) = m_{\theta,t_{i+1}-t_i}(x)$ that we may denote simply as $m_{\theta}(x)$, in order to make the notation easier.

We define an estimator $\hat{\theta}_n$ of θ_0 as

$$\hat{\theta}_n \in \arg\min_{\theta \in \Theta} U_n(\theta). \tag{1.9}$$

The idea, with a finite intensity, is to use the size of $X_{t_{i+1}} - X_{t_i}$ in order to judge the existence of a jump in an interval $[t_i, t_{i+1})$. The increment of X with continuous transition could hardly exceed the threshold $\Delta_{n,i}^{\beta}$ with $\beta \in (0, \frac{1}{2})$. Therefore we can judge a jump occurred if $|X_{t_{i+1}} - X_{t_i}| > \Delta_{n,i}^{\beta}$. We keep the idea even when the intensity is no longer finite.
With a such defined $m_{\theta}(X_{t_i})$, using the true parameter value θ_0, we have that

$$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0,t_i,t_{i+1}}(X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|X_{t_i} = x] =$$

$$= \mathbb{E}[X_{t_{i+1}}\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - x)|X_{t_i} = x] +$$

$$-\frac{\mathbb{E}[X_{t_{i+1}}\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|X_{t_i} = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|X_{t_i} = x]}\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|X_{t_i} = x] = 0,$$

where we have just used the definition and the measurability of $m_{\theta_0,t_i,t_{i+1}}(X_{t_i})$.
But, as the transition density is unknown, in general there is no closed expression for $m_{\theta,h}(x)$, hence the contrast is not explicit. However, in the proof of our results we will need an explicit development of (2.4).

In the sequel, for $\delta \geq 0$, we will denote $R(\theta, \Delta_{n,i}^{\delta}, x)$ for any function $R(\theta, \Delta_{n,i}^{\delta}, x) = R_{i,n}(\theta, x)$, where $R_{i,n} : \Theta \times \mathbb{R} \longrightarrow \mathbb{R}$, $(\theta, x) \mapsto R_{i,n}(\theta, x)$ is such that

$$\exists c > 0 \qquad |R_{i,n}(\theta, x)| \leq c(1 + |x|^c)\Delta_{n,i}^{\delta} \tag{1.10}$$

uniformly in θ and with c independent of i, n.
The functions R represent the term of rest and have the following useful property, consequence of the just given definition:

$$R(\theta, \Delta_{n,i}^{\delta}, x) = \Delta_{n,i}^{\delta} R(\theta, \Delta_{n,i}^{0}, x). \tag{1.11}$$

We point out that it does not involve the linearity of R, since the functions R on the left and on the right side are not necessarily the same but only two functions on which the control (3.23) holds with $\Delta_{n,i}^{\delta}$ and $\Delta_{n,i}^{0}$, respectively.

We state asymptotic expansions for $m_{\theta,\Delta_{n,i}}$. The cases $\alpha < 1$ and $\alpha \geq 1$ yield to different magnitude for the rest term.

Case $\alpha \in (0,1)$**:**

Theorem 2. *Suppose that Assumptions 1 to 4 hold and that* $\beta \in (0, \frac{1}{2})$ *and* $\alpha \in (0,1)$ *are given in definition* 1 *and the third point of Assumption* 4*, respectively. Then*

$$\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] = 1 + R(\theta, \Delta_{n,i}^{(1-\alpha\,\beta)\wedge(2-3\beta)}, x). \tag{1.12}$$

Theorem 3. *Suppose that Assumptions 1 to 4 hold and that* $\beta \in (0, \frac{1}{2})$ *and* $\alpha \in (0,1)$ *are given in definition* 1 *and the third point of Assumption* 4*, respectively. Then*

$$\mathbb{E}[(X_{t_{i+1}}^{\theta} - x)\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] = \Delta_{n,i}\, b(x,\theta) + \tag{1.13}$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x)\, [1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\, F(z)dz \,+\, R(\theta, \Delta_{n,i}^{2-2\beta}, x).$$

There exists $k_0 > 0$ *such that, for* $|x| \leq \Delta_{n,i}^{-k_0}$*,*

$$m_{\theta,\Delta_{n,i}}(x) = x + \Delta_{n,i}\, b(x,\theta) + \tag{1.14}$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x)\, [1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\, F(z)dz \,+\, R(\theta, \Delta_{n,i}^{2-2\beta}, x).$$

.

Case $\alpha \in [1,2)$**:**

Theorem 4. *Suppose that Assumptions 1 to 4 hold and that* $\beta \in (0, \frac{1}{2})$ *and* $\alpha \in [1,2)$ *are given in definition* 1 *and the third point of Assumption* 4*, respectively. Then*

$$\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] = 1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-4\beta)}, x). \tag{1.15}$$

Theorem 5. *Suppose that Assumptions 1 to 4 hold and that* $\beta \in (0, \frac{1}{3})$ *and* $\alpha \in [1,2)$ *are given in definition* 1 *and the third point of Assumption* 4*, respectively. Then*

$$\mathbb{E}[(X_{t_{i+1}}^{\theta} - x)\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] = \Delta_{n,i}\, b(x,\theta) + \tag{1.16}$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x)\, [1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\, F(z)dz \,+\, R(\theta, \Delta_{n,i}^{2-3\beta}, x).$$

There exists $k_0 > 0$ *such that, for* $|x| \leq \Delta_{n,i}^{-k_0}$*,*

$$m_{\theta,\Delta_{n,i}}(x) = x + \Delta_{n,i}\, b(x,\theta) + \tag{1.17}$$

$$-\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(x)\, [1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\, F(z)dz \,+\, R(\theta, \Delta_{n,i}^{2-3\beta}, x).$$

Remark 2. *The constant k in the definition* (2.4) *of contrast function can be taken in the interval* $(0, k_0]$. *In this way* $\Delta_{n,i}^{-k} \le \Delta_{n,i}^{-k_0}$ *and so* (2.31) *or* (1.17) *holds for* $|x| = |X_{t_i}|$ *smaller than* $\Delta_{n,i}^{-k}$.
If it is not the case the contribution of the observation X_{t_i} *in the contrast function is just* 0. *However we will see that suppressing the contribution of too big* $|X_{t_i}|$ *does not effect the efficiency property of our estimator.*

Remark 3. *In the development* (1.13) *or* (1.16) *the term* $\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(x)\,[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\,F(z)dz$ *is independent of* θ, *hence it will disappear in the difference* $m_\theta(x) - m_{\theta_0}(x)$, *but it is not negligible compared to* $\Delta_{n,i}\,b(x,\theta)$ *since its order is* $\Delta_{n,i}$ *if* $\alpha \in (0,1)$ *and at most* $\Delta_{n,i}^{\frac{1}{2}}$ *if* $\alpha \in [1,2)$. *Indeed, by the definition of the function* φ, *we know that we can consider as support of* $\varphi_{\Delta_{n,i}^{\beta}}(0) - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)$ *the interval* $c \times [-\frac{\Delta_{n,i}^{\beta}}{\|\gamma\|_\infty}, \frac{\Delta_{n,i}^{\beta}}{\|\gamma\|_\infty}]^c$. *If* $\alpha < 1$, *using moreover the third point of Assumption 4 we get the following estimation:*

$$|\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(x)\,[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\,F(z)dz| \le R(\theta_0, \Delta_{n,i}^{1}, X_{t_i}). \tag{1.18}$$

Otherwise, if $\alpha \ge 1$, *we have*

$$|\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(x)\,[1 - \varphi_{\Delta_{n,i}^{\beta}}(\gamma(x)z)]\,F(z)dz| \le$$

$$\le c|\Delta_{n,i}| \int_{c\times[-\frac{\Delta_{n,i}^{\beta}}{\|\gamma\|_\infty}, \frac{\Delta_{n,i}^{\beta}}{\|\gamma\|_\infty}]^c} |z|^{-\alpha} = R(\theta, \Delta_{n,i}^{1+\beta(1-\alpha)}, x),$$

with $\beta \in (0, \frac{1}{2})$ *and* $\alpha \in [1,2)$, *hence the exponent on* $\Delta_{n,i}$ *is always more than* $\frac{1}{2}$.
We can therefore write in the first case

$$m_{\theta,\Delta_{n,i}}(x) = x + R(\theta, \Delta_{n,i}, x) = R(\theta, \Delta_{n,i}^{0}, x) \tag{1.19}$$

and in the second

$$m_{\theta,\Delta_{n,i}}(x) = x + R(\theta, \Delta_{n,i}^{1+\beta(1-\alpha)}, x) = R(\theta, \Delta_{n,i}^{0}, x). \tag{1.20}$$

Remark 4. *In Theorems 2 - 4 we do not need conditions on* β *because, for each* $\beta \in (0, \frac{1}{2})$ *and for each* $\alpha \in (0,2)$ *the exponent on* $\Delta_{n,i}$ *is positive and therefore the last term of* (1.15) *is negligible compared to* 1. *In Theorem 5, instead,* R *is a negligible function if and only if* $2 - 3\beta \ge 1$, *it means that it must be* $\beta \le \frac{1}{3}$. *We have taken* $\beta \in (0, \frac{1}{3})$ *and so such a condition is always respected.*

1.3.2 Main results

Let us introduce the Assumption A_β that turns out starting from Theorems 2, 3, 4 and 5:

ASSUMPTION A_β: We choose $\beta \in (0, \frac{1}{2})$ if $\alpha \in (0,1)$. If on the contrary $\alpha \in [1,2)$, then we take β in $(0, \frac{1}{3})$.

The following theorems give a general consistency result and the asymptotic normality of the estimator $\hat{\theta}_n$, that hold without further assumptions on n and Δ_n.

Theorem 6. *(Consistency)*
Suppose that Assumptions 1 to 7 and A_β hold and let k of the definition of the contrast function (2.4) be in $(0, k_0)$. Then the estimator $\hat{\theta}_n$ is consistent in probability:

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \qquad n \to \infty.$$

Recalling that the Fisher information I is given by (1.3), we give the following theorem.

Theorem 7. *(Asymptotic normality)*
Suppose that Assumptions 1 to 8 and A_β hold, and $0 < k < k_0$.
Then the estimator $\hat{\theta}_n$ is asymptotically normal:

$$\sqrt{t_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta_0)), \qquad n \to \infty.$$

Remark 5. *Furthermore, the estimator $\hat{\theta}_n$ is asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem.*
The Hájek−LeCam convolution theorem states that any regular estimator in a parametric model which satisfies LAN property is asymptotically equivalent to a sum of two independent random variables, one of which is normal with asymptotic variance equal to the inverse of Fisher information, and the other having arbitrary distribution. The efficient estimators are those with the second component identically equal to zero.
The model (1.1) is LAN with Fisher information $I(\theta) = \int_{\mathbb{R}} \frac{(\dot{b}(\theta,x))^2}{a^2(x)} \pi^\theta(dx)$ (see [38]) and thus $\hat{\theta}_n$ is efficient.

Remark 6. *We point out that, contrary to the papers [38] and [88], in this case there is not any condition on the sampling, that can be irregular and with Δ_n that goes slowly to zero. On the other hand, our contrast function relies on the quantity $m_{\theta,h}(x)$ which is not explicit in general.*

1.4 Practical implementation of the contrast method

In order to use in practice the contrast function (2.4), one need to know the values of the quantities $m_{\theta,t_i,t_{i+1}}(X_{t_i})$. In most cases, it seems impossible to find an explicit expression for the function $m_{\theta,h}$ appearing in Definition 13. However, explicit or numerical approximations of this function seem available in many situations.

1.4.1 Approximate contrast function

Let us assume that one has at disposal an approximation of the function $m_{\theta,h}(x)$, denoted by $\widetilde{m}_{\theta,h}(x)$ which satisfies, for $|x| \le h^{-k_0}$,

$$|\widetilde{m}_{\theta,h}(x) - m_{\theta,h}(x)| \le R(\theta, h^\rho, x)$$

where the constant $\rho > 1$ assesses the quality of the approximation. We assume that the first three derivatives of $\tilde{m}_{h,\theta}$ with respect to the parameter provide approximation of the derivatives of $m_{h,\theta}$, in the following way

$$|\frac{\partial^i \widetilde{m}_{\theta,h}(x)}{\partial \theta^i} - \frac{\partial^i m_{\theta,h}(x)}{\partial \theta^i}| \le R(\theta, h^{1+\epsilon}, x), \quad \text{for } i = 1, 2, \tag{1.21}$$

$$|\frac{\partial^3 \widetilde{m}_{\theta,h}(x)}{\partial \theta^3} - \frac{\partial^3 m_{\theta,h}(x)}{\partial \theta^3}| \le R(\theta, h, x), \tag{1.22}$$

for all $|x| \leq h^{-k_0}$ and where $\epsilon > 0$. Let us stress that from Proposition 8 below, we know the derivatives with respect to θ of the quantity $m_{h,\theta}$.

Now, we consider $\widetilde{\theta}_n$ the estimator obtained from minimization of the contrast function (2.4) where one has replaced $m_{\theta,t_i,t_{i+1}}(X_{t_i})$ by its approximation $\widetilde{m}_{\theta,\Delta_{n,i}}(X_{t_i})$. Then, the result of Theorem 7 can be extended as follows.

Proposition 1. *Suppose that Assumptions 1 to 8 and A_β hold, with $0 < k < k_0$, and that $\sqrt{n}\Delta_n^{\rho-1/2} \to 0$ as $n \to \infty$.*
Then, the estimator $\widetilde{\theta}_n$ is asymptotically normal:

$$\sqrt{t_n}(\widetilde{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, I^{-1}(\theta_0)), \qquad n \to \infty.$$

We give below several examples of approximations of $m_{\theta,h}$. Let us stress that, in general, Theorem 3 (resp. Theorem 5) provides an explicit approximation of $m_{\theta,\Delta_{n,i}}(x)$ with an error of order $\Delta_{n,i}^{2-2\beta}$ (resp. of order $\Delta_{n,i}^{2-3\beta}$). They can be used to construct an explicit contrast function. In the next section we show that when the intensity is finite, it is possible to construct an explicit approximation of $m_{\theta,h}$ with arbitrarily high order.

1.4.2 Explicit contrast in the finite intensity case.

In the case with finite intensity it is possible to make the contrast explicit, using the development of $m_{\theta,\Delta_{n,i}}$ proved in the next proposition. We need the following assumption:

ASSUMPTION A_f:

1. We have $F(z) = \lambda F_0(z)$, $\int_{\mathbb{R}} F_0(z)dz = 1$ and F is a $\mathcal{C}^\infty$ function.
2. We assume that $x \mapsto a(x)$, $x \mapsto b(x,\theta)$ and $x \mapsto \gamma(x)$ are $\mathcal{C}^\infty$ functions, they have at most uniform in θ polynomial growth as well as their derivatives.

Let us define $A_K^{(k)}(x) = \bar{A}_c^k(g)(x)$, with $g(y) = (y - x)$ and $\bar{A}_c(f) = \bar{b}f' + \frac{1}{2}a^2 f''$; $\bar{b}(\theta, y) = b(\theta, y) - \int_{\mathbb{R}} \gamma(y) z F(z)dz$ as in the Remark 1.

Proposition 2. *Assume that A_f holds and let φ be a $\mathcal{C}^\infty$ function that has compact support and such that $\varphi \equiv 1$ on $[-1,1]$ and $\forall k \in \{0,...,M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx = 0$ for $M \geq 0$. Then, for $|x| \leq \Delta_{n,i}^{-k_0}$ with some $k_0 > 0$,*

$$m_{\theta,\Delta_{n,i}}(x) = x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_K^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x). \qquad (1.23)$$

In order to say that (1.23) holds, we have to prove the existence of a function φ with a compact support such that $\varphi \equiv 1$ on $[-1,1]$ and, $\forall k \in \{0,...,M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx$. We build it through ψ, a function with compact support, $\mathcal{C}^\infty$, such that $\psi|_{[-1,1]}(x) = \frac{x^M}{M!}$. We then define $\varphi(x) := \frac{\partial^M}{\partial x^M}\psi(x)$.
In this way we have $\varphi \equiv 1$ on $[-1,1]$, φ is $\mathcal{C}^\infty$, with compact support and such that for each $l \in \{0,...M\}$, using the integration by parts, $\int_{\mathbb{R}} x^l \varphi(x)dx = 0$, as we wanted.

Remark 7. *The development* (1.23) *is the same found in Kessler [51] in the case without jumps and it is obtained by the iteration of the continuous generator* $\bar{A}_c$*. Hence, it is completely explicit. Let us stress that in Kessler [51] the right hand side of* (1.23) *stands for an approximation of* $E[\bar{X}^\theta_{\Delta_{n,i}} \mid \bar{X}^\theta_0 = x]$ *where* $\bar{X}^\theta$ *is the continuous diffusion solution of* $d\bar{X}^\theta_t = \bar{b}(\theta, \bar{X}^\theta_s)ds + \sigma(\bar{X}^\theta_s)dW_s$*. From Proposition 2, the right hand side of* (1.23) *is also an approximation of* $m_{\theta,\Delta_{n,i}}(x) = \dfrac{E[X^\theta_{\Delta_{n,i}}\varphi_{\Delta^\beta_{n,i}}(X^\theta_{\Delta_{n,i}} - x) \mid X^\theta_0 = x]}{E[\varphi_{\Delta^\beta_{n,i}}(X^\theta_{\Delta_{n,i}} - x) \mid X^\theta_0 = x]}$ *in the case of finite activity jumps, and for a truncation kernel* φ *satisfying* $\forall k \in \{0, ..., M\}$, $\int_{\mathbb{R}} x^k\varphi(x)dx = 0$*. We emphasize that in the expansion of* $m_{\theta,\Delta_{i,n}}$ *given in Proposition 2, the contribution of the discontinuous part of the generator disappears only thanks to the choice of an oscillating function* φ*.*

Remark 8. *In the definition of the contrast function* (2.4) *we can replace* $m_{\theta,\Delta_{i,n}}(x)$ *with the explicit approximation* $\widetilde{m}^k_{\theta,\Delta_{n,i}}(x) := x + \sum_{h=1}^k \frac{\Delta^h_{n,i}}{h!}A^{(h)}_K(x)$*, with an error* $R(\theta, \Delta^k_{n,i}, x)$*, for* $k \leq \lfloor \beta(M + 2)\rfloor$*. Using* $A^{(1)}_K(x) = \Delta_{n,i}[b(\theta, x) - \int_{\mathbb{R}} \gamma(y)zF(z)dz]$ *and the expansions* (1.120)*–*(1.122) *we deduce that the conditions* (1.21) *–* (1.22) *are valid. Then, by application of Proposition 1, we can see that the associated estimator is efficient under the assumption* $\sqrt{n}\Delta_n^{k-\frac{1}{2}} \to 0$ *for* $n \to \infty$*. As* M*, and thus* k*, can be chosen arbitrarily large, we see that the sampling step* Δ_n *is allowed to converge to zero in a arbitrarily slow polynomial rate as a function of* n*. It turns out that a slow sampling step necessitates to choose a truncation function with more vanishing moments.*

1.5 Numerical experiments

1.5.1 Finite jump activity

Let us consider the model

$$X_t = X_0 + \int_0^t (\theta_1 X_s + \theta_2)ds + \sigma W_t + \gamma \int_0^t \int_{\mathbb{R}\setminus\{0\}} z\tilde{\mu}(ds, dz), \tag{1.24}$$

where the compensator of the jump measure is $\overline{\mu}(ds, dz) = \lambda F_0(z)dsdz$ for F_0 the probability density of the law $\mathcal{N}(\mu_J, \sigma^2_J)$ with $\mu_J \in \mathbb{R}$, $\sigma_J > 0$, $\sigma > 0$, $\theta_1 < 0$, $\theta_2 \in \mathbb{R}$, $\gamma \geq 0$, $\lambda \geq 0$. Since the jump activity is finite, we know from Section 1.4.2 that the function $m_{(\theta_1,\theta_2),\Delta_{n,i}}(x)$ can be approximated at any order using (1.23). As the latter is also the asymptotic expansion of the first conditional moment for the continuous S.D.E. $\bar{X}_t = \bar{X}_0 + \int_0^t(\theta_1\bar{X}_s + \theta_2 - \gamma\lambda\mu_J)ds + \sigma W_t$, which is explicit due to the linearity of the model, we decide to directly use the expression of the conditional moment and set

$$\widetilde{m}_{(\theta_1,\theta_2),\Delta_{n,i}}(x) = (x + \frac{\theta_2}{\theta_1} - \frac{\gamma\lambda\mu_J}{\theta_1})e^{\theta_1\Delta_{n,i}} + \frac{\gamma\lambda\mu_J - \theta_2}{\theta_1}. \tag{1.25}$$

Following Nikolskii [76], we construct oscillating truncation functions in the following way. First, we choose $\varphi^{(0)} : \mathbb{R} \to [0, 1]$ a $\mathcal{C}^\infty$ symmetric function with support on $[-2, 2]$ such that $\varphi^{(0)}(x) = 1$ for $|x| \leq 1$. We let, for $d > 1$, $\varphi^{(1)}_d(x) =$

$(d\varphi^{(0)}(x) - \varphi^{(0)}(x/d))/(d-1)$, which is a function equal to 1 on $[-1,1]$, vanishing on $[-d,d]^c$ and such that $\int_{\mathbb{R}} \varphi_d^{(1)}(x)dx = 0$. For $l \in \mathbb{N}$, $l \geq 1$, and $d > 1$, we set $\varphi_d^{(l)}(x) = c_d^{-1} \sum_{k=1}^{l} C_l^k (-1)^{k+1} \frac{1}{k} \varphi_d^{(1)}(x/k)$, where $c_d = \sum_{k=1}^{l} C_l^k (-1)^{k+1} \frac{1}{k}$. One can check that $\varphi_d^{(l)}$ is compactly supported, equal to 1 on $[-1,1]$, and that for all $k \in \{0,\dots,l\}$, $\int_{\mathbb{R}} x^k \varphi_d^{(l)}(x)dx = 0$, for $l \geq 1$. With these notations, we estimate the parameter $\theta = (\theta_1, \theta_2)$ by minimization of the contrast function

$$U_n(\theta) = \sum_{i=0}^{n-1} (X_{t_{i+1}} - \widetilde{m}_{(\theta_1,\theta_2),\Delta_{n,i}}(X_{t_i}))^2 \varphi^{(l)}_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}), \tag{1.26}$$

where $l \in \mathbb{N}$ and $c > 0$ will be specified latter.

For numerical simulations, we choose $T = 2000$, $n = 10^4$, $\Delta_{i,n} = \Delta_n = 1/5$, $\theta_1 = -0.5$, $\theta_2 = 2$ and $X_0 = x_0 = 4$. We estimate the bias and standard deviation of our estimators using a Monte Carlo method based on 5000 replications. As a start, we consider a situation without jumps $\lambda = 0$, in which we remove the truncation function φ in the contrast, as it is useless in absence of jumps. In Table 1.1, we compare the estimator $\widetilde{\theta}_n$ which uses the Kessler exact bias correction given by (1.25), with an estimator based on the Euler scheme approximation where one uses the approximation $\widetilde{m}^{\text{Euler}}_{(\theta_1,\theta_2),\Delta_{n,i}}(x) = x + \Delta_{n,i}(\theta_1 x + \theta_2)$. From Table 1.1 we see that the estimator $\tilde{\theta}_n^{\text{Euler}}$ based on Euler contrast exhibits some bias which is completely removed using Kessler's correction. Next, we set a jump intensity $\lambda = 0.1$, with jumps size whose common law is $\mathcal{N}(0,2)$ and set $\gamma = 1$. We use the contrast function relying on (1.25). Results are given for three choices of truncation function, $\varphi^{(0)}$, $\varphi_d^{(2)}$ and $\varphi_d^{(3)}$ where $d = 3$. Plots of these functions are given in Figure 1.1. We choose $\beta = 0.49$ and $c = 1$. As the true value of the volatility is $\sigma = 0.3$, this choice enables most of the increments without jumps of X on $[t_i, t_{i+1}]$ to be such that $\varphi^{(l)}_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) = 1$. Let us stress that, if σ is unknown, it is possible to estimate, even roughly, the local value of the volatility in order to choose c accordingly (see [38] for analogous discussion). Results in Table 1.2 show that the estimator works well, with a reduced bias for all choices of truncation function. Especially the bias is much smaller than the one of the Euler scheme contrast in absence of jumps. It shows the benefit of using (1.25) in the contrast function, even if the truncation function is not oscillating as is it when we consider $\varphi^{(0)}$. We remark that by the choice of a symmetric truncation function one has $\int_{\mathbb{R}} u\varphi^{(0)}(u)du = 0$ and inspecting the proof of Proposition 8 it can be seen that this conditions is sufficient, in the expansion of $m_{\theta,\Delta_{n,i}}$, to suppress the largest contribution of the discrete part of the generator.

If the number of jumps is greater, e.g. for $\lambda = 1$, we see in Table 1.3 that using the oscillating kernels $\varphi_d^{(2)}$, $\varphi_d^{(3)}$ yields to a smaller bias than using $\varphi^{(0)}$, whereas it tends to increase the standard deviation of the estimator. The estimator we get using $\varphi_d^{(3)}$ performs well in this situation, it has a negligible bias and a standard deviation comparable to the one in the case where the process has no jump.

1.5.2 Infinite jumps activity

Let us consider X solution to the stochastic differential equation (1.24), where the compensator of the jump measure is $\overline{\mu}(ds,dz) = \frac{e^{-z}}{z^{1+\alpha}} 1_{(0,\infty)}(z)dsdz$ with $\alpha \in (0,1)$. This situation corresponds to the choice of the Levy process $(\int_0^t \int_{\mathbb{R}\setminus\{0\}} z\tilde{\mu}(ds,dz))_t$

	Mean (std) for $\theta_1 = -0.5$	Mean (std) for $\theta_2 = 2$
$\widetilde{\theta}_n^{\text{Euler}}$	-0.4783 (0.0213)	1.9133 (0.0856)
$\widetilde{\theta}_n$	-0.5021 (0.0236)	2.0084 (0.0947)

Table 1.1 – Process without jump

	Mean (std) for $\theta_1 = -0.5$	Mean (std) for $\theta_2 = 2$
$\widetilde{\theta_n}$ using $\varphi^{(0)}$	-0.4967 (0.0106)	1.9869 (0.0430)
$\widetilde{\theta_n}$ using $\varphi_d^{(2)}$	-0.4990 (0.0153)	1.9959 (0.0622)
$\widetilde{\theta_n}$ using $\varphi_d^{(3)}$	-0.5006 (0.0196)	2.0023 (0.0798)

Table 1.2 – Gaussian jumps with $\lambda = 0.1$

being a tempered α-stable jump process. In the case of infinite jump activity, we have no result providing approximations at any arbitrary order of $m_{\theta,\Delta_{n,i}}$. However, we can use Theorem 3 to find some useful explicit approximation.

According to (2.31) and taking into account that the threshold level is $c\Delta_{n,i}^{\beta}$ for some $c > 0$, we have $m_{\theta,\Delta_{n,i}}(x) =$

$$
\begin{aligned}
&= x + \Delta_{n,i}(\theta_1 x + \theta_2) - \Delta_{n,i}\gamma \int_0^\infty \frac{e^{-z}}{z^\alpha} dz + \Delta_{n,i}\gamma \int_0^\infty \varphi_{c\Delta_{n,i}^\beta}(\gamma z) \frac{e^{-z}}{z^\alpha} dz + \\
&+ R(\theta, \Delta_{n,i}^{2-2\beta}, x) \\
&= x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \Delta_{n,i}^{1+\beta(1-\alpha)} c^{1-\alpha}\gamma^\alpha \int_0^\infty \varphi(v) \frac{e^{-\frac{cv\Delta_{n,i}^\beta}{\gamma}}}{v^\alpha} dv + R(\theta, \Delta_{n,i}^{2-2\beta}, x),
\end{aligned}
$$

where in the last line, following the notation of Remark 1, we have set $\bar{b}(x, \theta_1, \theta_2) = (\theta_1 x + \theta_2) - \gamma \int_0^\infty \frac{e^{-z}}{z^\alpha} dz$, and we make the change of variable $v = \frac{\gamma z}{c\Delta_{n,i}^\beta}$. This leads us to consider the approximation

$$
\widetilde{m}_{\theta,\Delta_{n,i}}(x) = x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \Delta_{n,i}^{1+\beta(1-\alpha)} c^{1-\alpha}\gamma^\alpha \int_0^\infty \varphi(v) \frac{1}{v^\alpha} dv, \tag{1.27}
$$

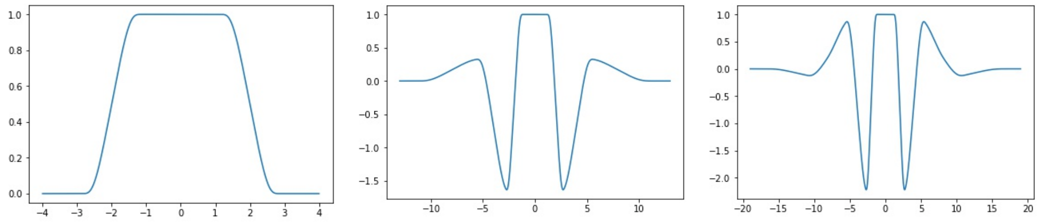
which is such that $|\widetilde{m}_{\theta,\Delta_{n,i}}(x) - m_{\theta,\Delta_{n,i}}(x)| \le R(\theta, \Delta_{n,i}^{(2-2\beta)\wedge(1+\beta(2-\alpha))}, x)$.

For numerical simulations, we choose $T = 100$, $n = 10^4$, $\Delta_{i,n} = \Delta_n = 1/100$, $\theta_1 = -0.5$, $\theta_2 = 2$, $X_0 = x_0 = 4$, $\gamma = 1$, $\sigma = 0.3$ and $\alpha \in \{0.1,\ 0.3,\ 0.5\}$. To illustrate the estimation method, we focus on the estimation of the parameter θ_2 only, as the minimisation of the contrast defined by (1.26)–(1.27) yields to the explicit estimator,

$$
\begin{aligned}
\widetilde{\theta}_{2,n} &= \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_n\theta_1 X_{t_i})\varphi_{c\Delta_n^\beta}(X_{t_{i+1}} - X_{t_i})}{\Delta_n \sum_{i=0}^{n-1} \varphi_{c\Delta_n^\beta}(X_{t_{i+1}} - X_{t_i})} - \gamma \int_0^\infty \frac{e^{-z}}{z^\alpha} dz \\
&\qquad - \Delta_n^{\beta(1-\alpha)} c^{1-\alpha}\gamma^\alpha \int_0^\infty \varphi(v) \frac{1}{v^\alpha} dv \\
&=: \widetilde{\theta}_{2,n}^{\text{Euler}} - \Delta_n^{\beta(1-\alpha)} c^{1-\alpha}\gamma^\alpha \int_0^\infty \varphi(v) \frac{1}{v^\alpha} dv.
\end{aligned} \tag{1.28}
$$

	Mean (std) for $\theta_1 = -0.5$	Mean (std) for $\theta_2 = 2$
$\widetilde{\theta_n}$ using $\varphi^{(0)}$	-0.4623 (0.0059)	1.8495 (0.0256)
$\widetilde{\theta_n}$ using $\varphi_d^{(2)}$	-0.4886 (0.0161)	1.9549 (0.0710)
$\widetilde{\theta_n}$ using $\varphi_d^{(3)}$	-0.5033 (0.0243)	2.0136 (0.1059)

Table 1.3 – Gaussian jumps with $\lambda = 1$



(a) $\varphi^{(0)}$ (b) $\varphi_d^{(2)}$ with $d = 3$ (c) $\varphi_d^{(3)}$ with $d = 3$

Figure 1.1 – Plot of the truncation functions

We can see that the estimator $\widetilde{\theta}_{2,n}$ is a corrected version of the estimator $\widetilde{\theta}_{2,n}^{\text{Euler}}$, that would result from the choice of the approximation $m_{\theta,\Delta_n}(x) \approx x + \Delta_n \bar{b}(x, \theta_1, \theta_2)$ in the definition of the contrast function. Comparing with estimators of earlier works (e.g. [38], [87]), the presence of this correction term appears new. In lines 2–3 of Table 1.4, we compare the mean and standard deviation of $\widetilde{\theta}_{2,n}$ and $\widetilde{\theta}_{2,n}^{\text{Euler}}$ for $\alpha \in \{0.1,\ 0.3,\ 0.5\}$ and with the choice $c = 1$, $\beta = 0.49$ and $\varphi = \varphi^{(0)}$ (see Figure 1.1). We see that the estimator $\widetilde{\theta}_{2,n}$ performs well and the correction term in (1.28) drastically reduces the bias present in $\widetilde{\theta}_{2,n}^{\text{Euler}}$, especially when the jump activity near 0 is high, corresponding to larger values of α. If we take a threshold level $c = 1.5$ higher, we see in line 5 of Table 1.4 that the bias of the estimator $\widetilde{\theta}_{2,n}^{\text{Euler}}$ increases, since the estimator $\widetilde{\theta}_{2,n}^{\text{Euler}}$ keeps more jumps that induce a stronger bias. On the other hand, the bias of the estimator $\widetilde{\theta}_{2,n}$ remains small (see line 4 of Table 1.4), as the correction term in (1.28) increases with c.

		$\alpha = 0.1$	$\alpha = 0.3$	$\alpha = 0.5$
c=1	$\widetilde{\theta}_{2,n}$	1.99 (0.0315)	1.98 (0.0340)	1.97 (0.0367)
	$\widetilde{\theta}_{2,n}^{\text{Euler}}$	2.20 (0.0315)	2.37 (0.0340)	2.76 (0.0367)
c=1.5	$\widetilde{\theta}_{2,n}$	1.97 (0.0340)	1.96 (0.0363)	1.94 (0.0397)
	$\widetilde{\theta}_{2,n}^{\text{Euler}}$	2.28 (0.0340)	2.48 (0.0363)	2.90 (0.0397)

Table 1.4 – Mean (std) for the estimation of $\theta_2 = 2$

1.5.3 Conclusion and perspectives for practical applications

In this chapter, we have shown that it is theoretically possible to estimate the drift parameter efficiently under the sole condition of a sampling step converging to zero. However, the contrast function relies on the quantity $m_{h,\theta}(x)$ which is usually not explicit. For practical implementation, the question of approximation of $m_{h,\theta}(x)$ is crucial, and one also has face the question of choosing the threshold level, characterized here by c, β and φ. On contrary to more conventional threshold methods, it appears here that the estimation quality seems less sensitive to choice of the threshold level, as the quantity $m_{h,\theta}(x)$ depends by construction on this threshold level and may compensate for too large threshold. On the other hand, the quantity $m_{h,\theta}(x)$ can be numerically very far from the approximation derived form the Euler scheme approximation. This can be seen in the example of Section 1.5.2, where the correction term of the estimator is, on this finite sample example, essentially of the same magnitude as the estimated quantity. A perspective, in the situation of infinite jump activity, would be to numerically approximate the function $x \mapsto m_{h,\theta}(x)$, using for instance a Monte Carlo approach, and provide more precise corrections than the explicit correction used in Section 1.5.2.

In the specific situation of finite activity, we proposed an explicit approximation of $m_{h,\theta}(x)$ with arbitrary order. This approximation is the same one as Kessler's approximation in absence of jumps, and it relies on the choice of oscillating truncation functions. A crucial point in the proof of the expansion of $m_{h,\theta}(x)$ given in Proposition 2 is that the support of the truncation function $\varphi_{c\Delta_{n,i}^{\beta}}$ is small compared to the typical scale where the density of the jumps law varies. However, our construction of oscillating function is such that the support of $\varphi = \varphi_d^{(l)}$ tends to be larger as the number of oscillations l is larger, which yields to restrictions for the choice of l on finite sample. Moreover, the truncation function takes large negative values as well, which makes the minimization of the contrast function unstable if the parameter set is too large. Perspective for further works would be to extend Proposition 2 for a non oscillating function φ. We expect that the resulting asymptotic expansion would involve additional terms related to the quantities $\int u^k \varphi(u) du$.

1.6 Limit theorems

The asymptotic properties of estimators are deduced from the asymptotic behavior of the contrast function. We therefore prepare some limit theorems for triangular arrays of the data, that we will prove in the Appendix.

Proposition 3. *Suppose that Assumptions 1 to 4 hold, $\Delta_n \to 0$ and $t_n \to \infty$. Moreover suppose that f is a differentiable function $\mathbb{R} \times \Theta \to \mathbb{R}$ such that $|f(x,\theta)| \leq c(1+|x|)^c$, $|\partial_x f(x,\theta)| \leq c(1+|x|)^c$ and $|\partial_\theta f(x,\theta)| \leq c(1+|x|)^c$.*
Then, $x \mapsto f(x,\theta)$ is a π-integrable function for any $\theta \in \Theta$ and the following convergence result holds as $n \to \infty$:
(i) $\sup_{\theta\in\Theta} |\frac{1}{t_n} \sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i},\theta) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$,
(ii) $\sup_{\theta\in\Theta} |\frac{1}{t_n} \sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i},\theta) \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0$.

The next proposition will be used in order to prove the consistency.
First, we prepare some notations. We define

$$\zeta_i := \int_{t_i}^{t_{i+1}} a(X_s) dW_s + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}) z \tilde{\mu}(ds,dz) + \tag{1.29}$$

$$+\Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{t_i})\,[1-\varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i})z)]\,F(z)dz.$$

We now observe that using the dynamic of the process X and the development (2.31) of m we get

$$X_{t_{i+1}} - m_\theta(X_{t_i}) + R(\theta, \Delta_{n,i}^{2-2\beta}, X_{t_i}) = (\int_{t_i}^{t_{i+1}} b(X_s, \theta_0)ds - \Delta_{n,i} b(X_{t_i}, \theta)) + \zeta_i, \quad (1.30)$$

if $\alpha < 1$ and the same but with the different rest term $R(\theta, \Delta_{n,i}^{2-3\beta}, X_{t_i})$ if $\alpha \geq 1$. From the choice that we have made on α and β in Theorems 3 and 5, the exponent on $\Delta_{n,i}$ in the rest function is always more than 1. Hence, from now on, we will call it simply $R(\theta, \Delta_{n,i}^{1+\delta}, X_{t_i})$, with $\delta > 0$. That is the reason why we choose such a definition for ζ_i.

Proposition 4. *Suppose that Assumptions 1 to 4 and A_β hold, $\Delta_n \to 0$ and $t_n \to \infty$ and, $\forall i \in \{0, ..., n-1\}$, $f_{i,n}$: $\mathbb{R}\times\Theta \to \mathbb{R}$. Moreover we suppose that $\exists c$: $|f_{i,n}(x, \theta)| \leq c(1 + |x|^c)$ $\forall i, n$.*
Then, $\forall \theta \in \Theta$,

$$\frac{1}{t_n} \sum_{i=0}^{n-1} f_{i,n}(X_{t_i}, \theta)\, \zeta_i\, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} \xrightarrow{\mathbb{P}} 0.$$

The proof relies on the following lemma:

Lemma 3. *Suppose that Assumptions 1 to 4 and A_β hold. Then*

1. $$\mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} | \mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}^{(1+\delta)\wedge\frac{3}{2}}, X_{t_i}), \quad (1.31)$$

2. $$\mathbb{E}[\zeta_i^2\, \varphi_{\Delta_{n,i}^\beta}^2(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} | \mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}, X_{t_i}), \quad (1.32)$$

and

3. $$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2\, \varphi_{\Delta_{n,i}^\beta}^2(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} | \mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}, X_{t_i}), \quad (1.33)$$

where $(\mathcal{F}_s)_s$ is the filtration defined in Lemma 25 and δ is positive as defined above.

We now give an asymptotic normality result:

Proposition 5. *Suppose that Assumptions 1 to 4 and A_β hold, $\Delta_n \to 0$, $t_n \to \infty$. Moreover suppose that f is a continuous function $\Theta\times\mathbb{R} \to \mathbb{R}$ that satisfies conditions in Proposition 3. Then for all θ*

$$\frac{1}{\sqrt{t_n}} \sum_{i=0}^{n-1} (X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) f(X_{t_i}, \theta)\, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} \xrightarrow{\mathcal{L}}$$

$$N(0, \int_{\mathbb{R}} f^2(x, \theta)\, a^2(x)\, \pi(dx)).$$

1.7 Proof of main results

We state a proposition that will be used repeatedly in the proof of Theorems 2,3,4 and 5. This proposition is an estimation of some expectations related to the event that increments of the process X lies where $\varphi_{\Delta_{n,i}}$, that is the smooth version of the indicator function, becomes singular for $\Delta_n \to 0$. The proof is postponed to Section 1.8.3.

Proposition 6. *Suppose that Assumptions 1 to 4 and A_β hold. Moreover suppose that $h : \mathbb{R} \times \Theta \longrightarrow \mathbb{R}$ is a function for which $\exists c > 0 : \sup_{\theta\in\Theta} |h(x,\theta)| \le c(1+|x|)^c$. Then $\forall k \ge 1$ $\forall \epsilon > 0$, we have*

$$\sup_{u\in[t_i,t_{i+1}]} \mathbb{E}[|h(X_u^\theta,\theta)|\,|\varphi^{(k)}_{\Delta_{n,i}^\beta}(X_u^\theta - X_{t_i}^\theta)|\,\Big|X_{t_i}^\theta = x] = R(\theta, \Delta_{n,i}^{1-\alpha\beta-\epsilon}, x).$$

with α and β given in the third point of Assumption 4 and Definition 1. We have used $\varphi^{(k)}_{\Delta_{n,i}^\beta}(y)$ in order to denote $\varphi^{(k)}(\frac{y}{\Delta_{n,i}^\beta})\Delta_{n,i}^{-\beta}$.

Proposition 19 is a consequence of the following more general proposition:

Proposition 7. *Suppose that Assumption 1 to 4 and A_β hold. For $c > 0$, we define*

$$\mathcal{Z}_{h,c,p} := \Big\{Z = (Z_\theta)_{\theta\in\Theta} \text{family of r. v. } \mathcal{F}_h \text{ measurable s t} \\ \sup_{\theta\in\Theta} \mathbb{E}[|Z_\theta|^p|X_0^\theta = x] \le c(1+|x|^c)\Big\}.$$

Then $\forall k \ge 1$ we have, $\forall \epsilon \ge \frac{1}{p}$,

$$\sup_{Z\in\mathcal{Z}_{h,c,p}} \mathbb{E}[\,|Z_\theta|\,|\varphi^{(k)}_{h^\beta}(X_h^\theta - X_0^\theta)|\,\Big|X_0^\theta = x] \le R(\theta, h^{(1-\alpha\beta)(1-\epsilon)}, x),$$

where $R(\theta, h^\delta, x)$ denotes any function such that $\exists c > 0$: $|R(\theta, h^\delta, x)| \le c(1+|x|^c)h^\delta$ uniformly in θ, with c independent of h.

1.7.1 Development of $m_{\theta,\Delta_{n,i}}(x)$

In order to study the asymptotic behavior of the contrast function we need some explicit approximation of $m_{\theta,\Delta_{n,i}}$. We study the asymptotic expansion of $m_{\theta,\Delta_{n,i}}(x)$ as $\Delta_{n,i} \to 0$. The main tools is the iteration of the Dynkin's formula that provides us the following expansion for every function f: $\mathbb{R} \to \mathbb{R}$ such that f is in $C^{2(k+1)}$:

$$\mathbb{E}[f(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] = \sum_{j=0}^{k} \frac{\Delta_{n,i}^j}{j!} A^j f(x) + \tag{1.34}$$

$$+ \int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} ... \int_{t_i}^{u_k} \mathbb{E}[A^{k+1} f(X_{u_{k+1}}^\theta)|X_{t_i}^\theta = x]\, du_{k+1} ... du_2\, du_1,$$

where A denotes the generator of the diffusion. A is the sum of the continuous and discrete part: $A := A_c + A_d$, with

$$A_c f(x) = \frac{1}{2} a^2(x) f''(x) + b(x,\theta) f'(x)$$

and

$$A_d f(x) = \int_{\mathbb{R}} (f(x+\gamma(x)z) - f(x) - z\gamma(x) f'(x))\, F(z) dz.$$

We set $A^0 = Id$.

1.7.1.1 Proof of Theorem 2:

Proof. We have to show (1.12). Using the formula (2.123) in the case $k = 1$, we get

$$\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x] =$$

$$= A^0\varphi_{\Delta_{n,i}^\beta}(0) + (t_{i+1} - t_i)A\varphi_{\Delta_{n,i}^\beta}(0) + \int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1} \mathbb{E}[A^2\varphi_{\Delta_{n,i}^\beta}(X_{u_2}^\theta)|X_{t_i}^\theta = x]du_2du_1. \tag{1.35}$$

We have defined φ as a smooth version of the indicator function, it means that in a neighborhood of 0 its value is 1 and so that $\varphi^{(k)}(0) = 0$ for each $k \geq 1$.
We denote $f_{i,n}(y) := \varphi_{\Delta_{n,i}^\beta}(y - x) = \varphi(\frac{y-x}{\Delta_{n,i}^\beta})$, with $\beta \in (0, \frac{1}{2})$. By the building, $f_{i,n}(x) = 1$ and $f_{i,n}^{(k)}(x) = 0$ for each $k \geq 1$, so we get $A_cf_{i,n}(x) = 0$ and $A_df_{i,n}(x) = \int_{\mathbb{R}\setminus\{0\}}[f_{i,n}(x + \gamma(x)z) - 1]\,F(z)dz$.
In the sequel the constant $c > 0$ may change from line to line.
From the definition of $f_{i,n}$ and the fact that $\varphi = 1$ on $[-1, 1]$ we have that $f_{i,n}(y) = 1$ for $|y - x| \leq \Delta_{n,i}^\beta$. Thus

$$|A_df_{i,n}(x)| \leq\; 2\left\|\varphi_{\Delta_{n,i}^\beta}\right\|_\infty \int_{\left\{z:|z\gamma(x)|\geq\Delta_{n,i}^\beta\right\}} F(z)dz \leq$$

$$\leq 2\left\|\varphi_{\Delta_{n,i}^\beta}\right\|_\infty \int_{\left\{z:|z|\geq\frac{\Delta_{n,i}^\beta}{|\gamma(x)|}\right\}} |z|^{-1-\alpha}dz \leq c\left\|\varphi_{\Delta_{n,i}^\beta}\right\|_\infty |\gamma(x)|^\alpha\Delta_{n,i}^{-\beta\alpha} = R(\theta, \Delta_{n,i}^{-\alpha\beta}, x),$$

where the second inequality follows from point 3 of Assumption 4. Substituting in (1.35) we get

$$\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x] = \tag{1.36}$$

$$= 1 + \Delta_{n,i}R(\theta, \Delta_{n,i}^{-\alpha\beta}, x) + \int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1} \mathbb{E}[A^2\varphi_{\Delta_{n,i}^\beta}(X_{u_2}^\theta)|X_{t_i}^\theta = x]du_2du_1.$$

In order to prove (1.12), we want to show that the last term is negligible.
We consider the generator's decomposition in discrete and continuous part $A = A_c + A_d$ that yields: $A^2f_{i,n}(y) = (A_c^2f_{i,n})(y) + A_c(A_df_{i,n})(y) + A_d(A_cf_{i,n})(y) + (A_d^2f_{i,n})(y)$.
We observe that we can write $(A_c^2f_{i,n})(y)$ as

$$\sum_{j=1}^4 \Delta_{n,i}^{-\beta j}h_j(y, \theta)\varphi_{\Delta_{n,i}^\beta}^{(j)}(y - x),$$

where $\varphi_{\Delta_{n,i}^\beta}^{(j)}(y - x) = \varphi^{(j)}(\frac{(y-x)}{\Delta_{n,i}^\beta})$. For each $j \in \{1, 2, 3, 4\}$, h_j is a function of a, b and their derivatives up to second order: $h_1 = \frac{1}{2}a^2b'' + bb'$, $h_2 = \frac{1}{2}a^2(a')^2 + \frac{1}{2}a^3a'' + a^2b' + aa'b + b^2$, $h_3 = a^3a' + a^2b$ and $h_4 = \frac{1}{4}a^4$.
Using the Proposition 19 we get that $\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[(A_c^2f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]|$ is upper bounded by

$$\sup_{u_2\in[t_i,t_{i+1}]} |\sum_{j=1}^4 \Delta_{n,i}^{-\beta j}\,\mathbb{E}[h_j(X_{u_2}^\theta, \theta)\varphi_{\Delta_{n,i}^\beta}^{(j)}(X_{u_2}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]| =$$

$$= |\sum_{j=1}^4 \Delta_{n,i}^{-\beta j}R(\theta, \Delta_{n,i}^{1-\alpha\beta-\epsilon}, x)| = R(\theta, \Delta_{n,i}^{1-\alpha\beta-\epsilon-4\beta}, x).$$

Let us now consider $A_c(A_d f_{i,n})(y)$. Substituting the definition of $A_d f_{i,n}$ we get

$$A_c(A_d f_{i,n})(y) = A_c(\int_{\mathbb{R}} g_n(\cdot, z) F(z) dz)(y), \tag{1.37}$$

where

$$g_n(y,z) := \varphi_{\Delta_{n,i}^{\beta}}(y - x + z\gamma(y)) - \varphi_{\Delta_{n,i}^{\beta}}(y - x) - \Delta_{n,i}^{-\beta} \varphi'_{\Delta_{n,i}^{\beta}}(y - x)\gamma(y) z \tag{1.38}$$

and where the notation used means that we are applying the differential operator A_c with respect to the variable represented with a dot. In order to estimate it we observe that

$$|g_n(y,z)| \le \Delta_{n,i}^{-\beta} \left\| \varphi' \right\|_\infty |z||\gamma(y)|, \tag{1.39}$$

$$|\frac{\partial}{\partial y} g_n(y,z)| \le \Delta_{n,i}^{-2\beta} P(y)|z| \qquad \text{and} \tag{1.40}$$

$$|\frac{\partial^2}{\partial y^2} g_n(y,z)| \le \Delta_{n,i}^{-3\beta} P(y)(|z| + |z|^2); \tag{1.41}$$

where $P(y)$ is a polynomial function in y, that may change from line to line.
Since the functions a^2 and b have polynomial growth, we obtain

$$|A_c g_n(\cdot, z)(y)| \le \Delta_{n,i}^{-3\beta} P(y)(|z| + |z|^2). \tag{1.42}$$

Using the dominated convergence theorem we get

$$A_c(\int_{\mathbb{R}} g_n(\cdot, z) F(z) dz)_{(y)} = \int_{\mathbb{R}} (A_c g_n)(\cdot, z)_{(y)} F(z) dz,$$

Therefore, using (1.42),

$$|A_c(\int_{\mathbb{R}} g_n(\cdot, z) F(z) dz)_{(y)}| \le \Delta_{n,i}^{-3\beta} P(y) \int_{\mathbb{R}} (|z| + |z|^2) F(z) dz,$$

that is upper bounded by $c\Delta_{n,i}^{-3\beta} P(y)$ since α is less than 1. It turns

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_c(A_d f_{i,n}))(X_{u_2}^{\theta})|X_{t_i}^{\theta} = x]| \le$$

$$\le \sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[c\Delta_{n,i}^{-3\beta} P(X_{u_2}^{\theta})|X_{t_i}^{\theta} = x]| = R(\theta, \Delta_{n,i}^{-3\beta}, x)$$

where, in the last equality, we have used the third point of Lemma 25.
We reason in the same way on $A_d(A_c f_{i,n})(y)$, which is equal to

$$\int_{\mathbb{R}} [A_c f_{i,n}(y + z\gamma(y)) - A_c f_{i,n}(y) - z\gamma(y)(A_c f_{i,n})'(y)] F(z) dz. \tag{1.43}$$

It is, in module, upper bounded by

$$c \int_0^1 \int_{\mathbb{R}} [|(A_c f_{i,n})'(y + z\gamma(y)s)| + |(A_c f_{i,n})'(y)|]|z||\gamma(y)| F(z) ds\, dz. \tag{1.44}$$

We observe that, $\forall y'$, $(A_c f_{i,n})'(y') = (b' f'_{i,n} + b f''_{i,n} + a a' f''_{i,n} + \frac{1}{2} a^2 f'''_{i,n})(y')$.
By the fact that $|\frac{\partial^j}{\partial y^j} \varphi_{\Delta_{n,i}^{\beta}}(y)| \le c\Delta_{n,i}^{-\beta j}$ for $j = 1, 2, 3$ and recalling $f_{i,n}(y) = \varphi_{\Delta_{n,i}^{\beta}}(y - x)$, we get that

$$|(A_c f_{i,n})'(y')| \le c\, P(y') \Delta_{n,i}^{-3\beta}, \tag{1.45}$$

where we have used that b and a^2 have polynomial growth. We obtain that (1.44) is upper bounded by

$$\Delta_{n,i}^{-3\beta}\int_0^1\int_{\mathbb{R}}(P(y+z\gamma(y)s)+P(y))|z||\gamma(y)|F(z)ds\,dz\leq$$

$$\leq\Delta_{n,i}^{-3\beta}\int_{\mathbb{R}}P(y)P(z)|z|F(z)dz\leq c\Delta_{n,i}^{-3\beta}P(y),$$

where we have used the first point of Assumptions 3 and the third of Assumption 4, with $\alpha\in(0,1)$, in order to get $\int_{\mathbb{R}}P(z)|z|F(z)dz\leq\infty$.
Considering the controls (1.44) and (1.45) on (1.43) it yields, using again the third point of Lemma 25,

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[(A_d(A_cf_{i,n}))(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=R(\theta,\Delta_{n,i}^{-3\beta},x).$$

To conclude, we consider $A_d(A_df_{i,n})(y)$:

$$\int_{\mathbb{R}}[A_df_{i,n}(y+z\gamma(y))-A_df_{i,n}(y)-z\gamma(y)(A_df_{i,n})'(y)]F(z)dz.\tag{1.46}$$

Again, (1.46) is, in module, upper bounded by

$$c\int_0^1\int_{\mathbb{R}}[|(A_df_{i,n})'(y-x+z\gamma(y)s)|+|(A_df_{i,n})'(y)|]|z||\gamma(y)|F(z)ds\,dz\tag{1.47}$$

But

$$A_df_{i,n}(y')=\int_{\mathbb{R}}g_n(y',z)F(z)dz,\tag{1.48}$$

with $g_n(y',z)$ given in (1.38) Using control equation (1.40) and dominated convergence theorem, we get that its derivative is upper bounded by $c\Delta_{n,i}^{-2\beta}P(y')$.
Using also (1.46) and (1.47),

$$|A_d^2f_{i,n}(y)|\leq\Delta_{n,i}^{-2\beta}P(y)\int_{\mathbb{R}}|z|F(z)dz$$

and it turns, using third point of Lemma 25,

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[(A_d^2f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=R(\theta,\Delta_n^{-2\beta},x).$$

By the decomposition of the generator in A_c and A_d we get

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[A^2f_{i,n}(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=$$

$$=R(\theta,\Delta_{n,i}^{1-\alpha\beta-4\beta-\epsilon},x)+R(\theta,\Delta_{n,i}^{-3\beta},x)+R(\theta,\Delta_{n,i}^{-2\beta},x),$$

with $\alpha\in(0,1)$ and $\beta\in(0,\frac{1}{2})$, so it is $R(\theta,\Delta_{n,i}^{-3\beta},x)$, since the other R functions are always negligible compared to it.
Using (1.36) we get

$$\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta-X_{t_i}^\theta)|X_{t_i}^\theta=x]=1+\Delta_{n,i}R(\theta,\Delta_{n,i}^{-\alpha\beta},x)+\frac{\Delta_{n,i}^2}{2}R(\theta,\Delta_{n,i}^{-3\beta},x).$$

We deduce, using the definition of $\Delta_{n,i}$ and (3.24), that it is

$$1+R(\theta,\Delta_{n,i}^{1-\alpha\beta},x)+R(\theta,\Delta_{n,i}^{2-3\beta},x)=1+R(\theta,\Delta_{n,i}^{(1-\alpha\beta)\wedge(2-3\beta)},x),$$

as we wanted. □

1.7.2 Proof of Theorem 4

Proof. Let α now be in $[1, 2)$. In the sequel we skip the study of the case $\alpha = 1$ for simplicity, in order to avoid the appearance of logarithmic functions. However, such a specific case is embedded in the case $\alpha > 1$ by taking $\alpha = 1 + \epsilon$ with a choice of $\epsilon > 0$ arbitrarily small.
Using again Dynkin formula, we have that (1.36) is still true. Considering the generator's decomposition, we act like in the case where α is less than 1 to get that

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_c^2 f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{1-\alpha\beta-\epsilon-4\beta}, x). \tag{1.49}$$

Concerning $A_c(A_d f_{i,n})(y)$, we use (1.37) with g_n defined in (1.38). Using Taylor development to the second order we get

$$|g_n(y, z)| \leq \left\| \varphi''_{\Delta_{n,i}^\beta} \right\|_\infty |\Delta_{n,i}|^{-2\beta} \frac{|z|^2 \gamma(y)^2}{2}. \tag{1.50}$$

In the same way we get the following two estimations:

$$|\frac{\partial}{\partial y} g_n(y, z)| \leq$$

$$\leq |\Delta_{n,i}|^{-2\beta} \left\| \varphi''_{\Delta_{n,i}^\beta} \right\|_\infty |\gamma(y)||\gamma'(y)||z|^2 + |\frac{\Delta_{n,i}|^{-3\beta}}{2} \left\| \varphi'''_{\Delta_{n,i}^\beta} \right\|_\infty |z|^2 \gamma^2(y)|1 + \gamma'(y)z|,$$

$$|\frac{\partial^2}{\partial y^2} g_n(y, z)| \leq \tag{1.51}$$

$$\leq |\Delta_{n,i}|^{-2\beta} |z|^2 P(y) + |\Delta_{n,i}|^{-3\beta} |P(y)(|z|^2 + |z|^3) + |\Delta_{n,i}|^{-4\beta} P(y)(|z|^2 + |z|^3).$$

Since a^2 and b have polynomial growth, (1.51) provides us an estimation on $|A_c g_n(\cdot, z)(y)|$. Using dominated convergence theorem, (1.37), the estimation of $|A_c g_n(\cdot, z)(y)|$ obtained from (1.51) and the fact that $\int_{\mathbb{R}} (|z|^2 + |z|^3) F(z) dz < \infty$, we get

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_c A_d f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = \tag{1.52}$$

$$= R(\theta, \Delta_{n,i}^{-2\beta}, x) + R(\theta, \Delta_{n,i}^{-3\beta}, x) + R(\theta, \Delta_{n,i}^{-4\beta}, x) = R(\theta, \Delta_{n,i}^{-4\beta}, x).$$

We now consider $A_d(A_c f_{i,n})(y)$. Using (1.43) and the development to the second order of the function $A_c f_{i,n}(y + z\gamma(y))$ we obtain

$$|A_d(A_c f_{i,n})(y)| \leq c \int_{\mathbb{R}} \int_0^1 |(A_c f_{i,n})''(y + s\, z\gamma(y))||z|^2 |\gamma^2(y)| F(z) ds dz. \tag{1.53}$$

We observe that $(A_c f_{i,n})''(y') = [b'' f'_{i,n} + 2b' f''_{i,n} + b f'''_{i,n} + (a')^2 f''_{i,n} + a(a'' f''_{i,n} + a' f'''_{i,n}) + 2aa' f'''_{i,n} + \frac{1}{2} a^2 f^{(4)}_{i,n}](y')$. By the fact that $|\frac{\partial^j}{\partial y^j} \varphi_{\Delta_{n,i}^\beta}(y)| \leq c \Delta_{n,i}^{-\beta j}$ for $j = 1, 2, 3$ and recalling $f_{i,n}(y) = \varphi_{\Delta_{n,i}^\beta}(y - x)$, we get that

$$|(A_c f_{i,n})''(y')| \leq c\, P(y') \Delta_{n,i}^{-4\beta}. \tag{1.54}$$

Using (1.53) and (1.54) it yields

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_d A_c f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{-4\beta}, x). \tag{1.55}$$

To conclude, we consider $A_d A_d f_{i,n}$. Using (1.46) and the development up to the second order we get

$$|A_d(A_d f_{i,n})(y)| \le c \int_{\mathbb{R}} \int_0^1 |(A_d f)''(y + s\, z\gamma(y))||z|^2|\gamma^2(y)|F(z)dsdz.$$

We recall that (1.48) still holds, with g_n defined in (1.38). In order to estimate $(A_d f)''(y)$ in the case where $\alpha \in [1,2)$ we use therefore (1.51) joint with dominated convergence theorem. It provides us

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_d A_d f_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{-4\beta}, x). \tag{1.56}$$

Using (1.49), (1.52), (1.55) and (1.56) we put the pieces together and so we obtain

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[A^2 f_{i,n}(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{1-\alpha\beta-4\beta-\epsilon}, x) + R(\theta, \Delta_{n,i}^{-4\beta}, x).$$

We replace it in the Dynkin formula (1.36) getting

$$\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x] = 1 + \Delta_{n,i} R(\theta, \Delta_{n,i}^{-\alpha\beta}, x) + \frac{\Delta_{n,i}^2}{2} R(\theta, \Delta_{n,i}^{(1-\alpha\beta-4\beta-\epsilon)\wedge(-4\beta)}, x).$$

Using the definition of $\Delta_{n,i}$ and (3.24) it is

$$1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(3-\alpha\beta-4\beta-\epsilon)\wedge(2-4\beta)}, x). \tag{1.57}$$

Since ϵ is arbitrarily small, for each choice of α and β there exists ϵ such that $3 - \alpha\beta - 4\beta - \epsilon$ is greater than $2 - 4\beta$ and (1.15) follows. □

1.7.3 Proof of Theorem 3

Proof. We observe that

$$m_{\theta,\Delta_{n,i}}(x) := \frac{\mathbb{E}[X_{t_{i+1}}^\theta \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} = x + \frac{\mathbb{E}[g_{i,n}(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}, \tag{1.58}$$

with $g_{i,n}(y) = (y - x)\varphi_{\Delta_{n,i}^\beta}(y - x)$.

We have already found a development for the denominator of (1.58) given by (1.12), we use again the Dynkin's formula (2.123) for $k = 1$ in order to find a development for the numerator. By the building, $g_{i,n}(x) = 0$, $g'_{i,n}(x) = 1$ and $g''_{i,n}(x) = 0$, so we get

$$A_c g_{i,n}(x) = b(x, \theta)$$

and

$$A_d g_{i,n}(x) = \int_{\mathbb{R}\setminus\{0\}} [g_{i,n}(x + z\gamma(x)) - z\gamma(x)]F(z)dz =$$

$$= \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^\beta}(z\gamma(x)) - 1]F(z)dz$$

where we have used, in the last equality, simply the definition of $g_{i,n}$.
Substituting in the Dynkin's formula we get

$$\mathbb{E}[g_{i,n}(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] = \Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^\beta}(z\gamma(x)) - 1]F(z)dz +$$

$$+\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\mathbb{E}[A^2 g_{i,n}(X_{u_2})|X_{t_i}=x]du_2du_1. \qquad (1.59)$$

In order to show that the last term is negligible, we have to estimate $(A^2 g_{i,n})(y)$ using the decomposition in continuous and discrete part of the generator, as we have already done.
Since $g_{i,n}(y)=(y-x)\varphi_{\Delta_{n,i}^{\beta}}(y-x)$, we have

$$g_{i,n}^{(h)}(y)=\sum_{k=0}^{h}\binom{h}{k}\frac{\partial^k}{\partial y^k}(y-x)\frac{\partial^{h-k}}{\partial y^{h-k}}(\varphi_{\Delta_{n,i}^{\beta}}(y-x)),$$

with $\binom{h}{k}$ binomial coefficients. So we get, observing that the derivatives of $(y-x)$ after the second order are zero, the following useful control for $h\ge 1$:

$$|g_{i,n}^{(h)}(y)|\le|\varphi_{\Delta_{n,i}^{\beta}}^{(h)}(y-x)|\Delta_{n,i}^{-\beta h}|y-x|+|\varphi_{\Delta_{n,i}^{\beta}}^{(h-1)}(y-x)|\Delta_{n,i}^{-\beta(h-1)}|h|. \qquad (1.60)$$

By the definition of φ as a smooth version of the indicator function, we know that it exists $c>0$ such that if $\frac{|y-x|}{\Delta_{n,i}^{\beta}}>c$, then φ and its derivatives are zero when evaluated at the point $\frac{(y-x)}{\Delta_{n,i}^{\beta}}$.
So we can say that $|\varphi_{\Delta_{n,i}^{\beta}}^{(h)}(y-x)||y-x|\le c|\varphi_{\Delta_{n,i}^{\beta}}^{(h)}(y-x)|\Delta_{n,i}^{\beta}$ and consequently

$$|g_{i,n}^{(h)}(y)|\le c|\varphi_{\Delta_{n,i}^{\beta}}^{(h)}(y-x)|\Delta_{n,i}^{-\beta(h-1)}+c|\varphi_{\Delta_{n,i}^{\beta}}^{(h-1)}(y-x)|\Delta_{n,i}^{-\beta(h-1)}. \qquad (1.61)$$

Reasoning as in the proof of Theorem 2, we start with $(A_c^2 g_{i,n})(y)$ and we get that it is $\sum_{j=1}^{4}h_j(y,\theta)g_{i,n}^{(j)}(y)$ where again, for each $j\in\{1,2,3,4\}$, h_j is a function of a, b and their derivatives up to second order.
We substitute in $\mathbb{E}[(A_c^2 g_{i,n})(X_{u_2}^{\theta})|X_{t_i}^{\theta}=x]$, getting $\sum_{j=1}^{4}\mathbb{E}[h_j(X_{u_2}^{\theta},\theta)g_{i,n}^{(j)}(X_{u_2}^{\theta})|X_{t_i}^{\theta}=x]$. Using the estimation (1.60) we obtain

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[|A_c^2 g_{i,n}(X_{u_2}^{\theta})||X_{t_i}^{\theta}=x]|\le$$

$$\le\sup_{u_2\in[t_i,t_{i+1}]}|\sum_{j=1}^{4}c\Delta_{n,i}^{-\beta(j-1)}\mathbb{E}[|h_j(X_{u_2}^{\theta},\theta)|(|\varphi_{\Delta_{n,i}^{\beta}}^{(j)}(X_{u_2}^{\theta}-X_{t_i}^{\theta})|+|\varphi_{\Delta_{n,i}^{\beta}}^{(j-1)}(X_{u_2}^{\theta}-X_{t_i}^{\theta})|)|X_{t_i}^{\theta}=x]|.$$

We observe that we can see $\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[|h_1(X_{u_2}^{\theta},\theta)||\varphi_{\Delta_{n,i}^{\beta}}(X_{u_2}^{\theta}-X_{t_i}^{\theta})||X_{t_i}^{\theta}=x]|$ as $R(\theta,\Delta_{n,i}^{0},x)=R(\theta,1,x)$ and we use the Proposition 19 on the other terms, getting

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[|A_c^2 g_{i,n}(X_{u_2}^{\theta})||X_{t_i}^{\theta}=x]|\le$$

$$\le R(\theta,\Delta_{n,i}^{1-\alpha\beta-\epsilon},x)+R(\theta,1,x)+\sum_{j=2}^{4}c\Delta_{n,i}^{-\beta(j-1)}R(\theta,\Delta_{n,i}^{1-\alpha\beta-\epsilon},x)= \qquad (1.62)$$

$$=R(\theta,\Delta_{n,i}^{1-\alpha\beta-\epsilon-3\beta},x)+R(\theta,1,x).$$

Let us now consider $A_c(A_d g_{i,n})(y)$

$$A_c(A_d g_{i,n})(y)=A_c(\int_{\mathbb{R}}[g_{i,n}(\cdot+z\gamma(\cdot))-g_{i,n}(\cdot)-z\gamma(\cdot)g_{i,n}'(\cdot)]F(z)dz)_{(y)}. \qquad (1.63)$$

Let us denote

$$h_{i,n}(y,z) := g_{i,n}(y + z\gamma(y)) - g_{i,n}(y) - z\gamma(y)g'_{i,n}(y). \tag{1.64}$$

We observe that

$$\frac{\partial h_{i,n}}{\partial y}(y,z) = g'_{i,n}(y+z\gamma(y)) - g'_{i,n}(y) + z\gamma'(y)(g'_{i,n}(y+z\gamma(y)) - g'_{i,n}(y)) - z\gamma(y)g''_{i,n}(y), \tag{1.65}$$

$$\frac{\partial^2 h_{i,n}}{\partial y^2}(y,z) = g''_{i,n}(y + z\gamma(y))(1 + z\gamma'(y))^2 + g'_{i,n}(y + z\gamma(y))z\gamma''(y)+$$
$$-g''_{i,n}(y) - g'''_{i,n}(y)\gamma(y)z - 2g''_{i,n}(y)z\gamma'(y) - g'_{i,n}(y)z\gamma''(y). \tag{1.66}$$

Using the estimation (1.61), we have

$$|g'_{i,n}(y)| \le c \qquad |g''_{i,n}(y)| \le c\Delta_{n,i}^{-\beta} \qquad |g'''_{i,n}(y)| \le c\Delta_{n,i}^{-2\beta} \tag{1.67}$$

Hence

$$|\frac{\partial h_{i,n}}{\partial y}(y,z)| \le \left\| g''_{i,n} \right\|_\infty P(y)(|z| + |z|^2) \le (|z| + |z|^2)P(y)\Delta_{n,i}^{-\beta}, \tag{1.68}$$

and similarly

$$|\frac{\partial^2 h_{i,n}}{\partial y^2}(y,z)| \le \Delta_{n,i}^{-2\beta} P(y)(|z| + |z|^2 + |z|^3).$$

Since functions a^2 and b have polynomial growth, we obtain

$$|A_c h_{i,n}(y,z)| \le \Delta_{n,i}^{-2\beta} P(y)(|z| + |z|^2 + |z|^3).$$

Using dominated convergence theorem we get

$$|A_c(\int_{\mathbb{R}} h_{i,n}(\cdot,z)F(z)dz)_{(y)}| \le \Delta_{n,i}^{-2\beta} P(y) \int_{\mathbb{R}} (|z| + |z|^2 + |z|^3)F(z)dz$$

and so, using also the third point of Lemma 25 and (1.63), we get

$$\sup_{u_2 \in [t_i,t_{i+1}]} |\mathbb{E}[A_c(A_d g_{n,i})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{-2\beta}, x). \tag{1.69}$$

We reason on the same way on $A_d(A_c g_{n,i})(y)$:

$$A_d(A_c g_{n,i})(y) = \int_{\mathbb{R}} [A_c g_{n,i}(y + z\gamma(y)) - A_c g_{n,i}(y) - z\gamma(y)(A_c g_{n,i})'(y)]F(z)dz. \tag{1.70}$$

It is, in module, upper bounded by

$$c\int_0^1 \int_{\mathbb{R}} [|(A_c g_{n,i})'(y + z\gamma(y)s)| + |(A_c g_{n,i})'(y)|]|z||\gamma(y)|F(z)dsdz.$$

In order to estimate it we observe that, $\forall y'$,

$$(A_c g_{n,i})'(y') = (aa'g''_{n,i} + \frac{1}{2}a^2 g'''_{n,i} + b'g'_{n,i} + bg''_{n,i})(y').$$

Using (1.67) and the polynomial growth of a, b and their derivatives, we get

$$|(A_c g_{n,i})'(y')| \le c + P(y')\Delta_{n,i}^{-\beta} + P(y')\Delta_{n,i}^{-2\beta} \le P(y')\Delta_{n,i}^{-2\beta}.$$

It yields

$$c\int_0^1\int_{\mathbb{R}}[|(A_c g_{n,i})'(y+z\gamma(y)s)|+|(A_c g_{n,i})'(y)|]|z||\gamma(y)|F(z)dsdz\le$$

$$\le \Delta_{n,i}^{-2\beta}\int_0^1\int_{\mathbb{R}}(P(y+z\gamma(y)s)+P(y)]|z||\gamma(y)|F(z)dsdz\le$$

$$\le \Delta_{n,i}^{-2\beta}\int_{\mathbb{R}}P(y)P(z)|z|F(z)dz\le c\Delta_{n,i}^{-2\beta}P(y),$$

where we have used the first point of Assumptions 3 and the second of Assumption 4. Hence $|A_d(A_c g)(y)|\le \Delta_{n,i}^{-2\beta}P(y)$.
Taking the expected value and using the third point of Lemma 25, we obtain

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[A_d(A_c g_{n,i})(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=R(\theta,\Delta_{n,i}^{-2\beta},x).$$

In conclusion, we consider $A_d^2(g_{n,i})(y)$

$$A_d^2(g_{n,i})(y)=\int_{\mathbb{R}}[A_d g_{n,i}(y+z\gamma(y))-A_d g_{n,i}(y)-z\gamma(y)(A_d g_{n,i})'(y)]F(z)dz. \quad (1.71)$$

Again it is, in module, upper bounded by

$$c\int_0^1\int_{\mathbb{R}}[|(A_d g_{n,i})'(y+z\gamma(y)s)|+|(A_d g_{n,i})'(y)|]|z||\gamma(y)|F(z)ds\,dz \quad (1.72)$$

But

$$A_d g_{n,i}(y')=\int_{\mathbb{R}}[g_{n,i}(y'+z\gamma(y'))-g_{n,i}(y')-z\gamma(y')g_{n,i}'(y')]F(z)dz=\int_{\mathbb{R}}h_{i,n}(y',z)F(z)dz, \quad (1.73)$$

with $h_{i,n}$ defined in (1.64). Using control equation (1.68) and dominated convergence theorem, we get that (1.73) is upper bounded by $P(y')\Delta_{n,i}^{-\beta}$.
It follows from (1.71) and (1.72) that

$$|A_d^2 g_{n,i}(y)|\le c\Delta_{n,i}^{-\beta}P(y)\int_{\mathbb{R}}P(z)F(z)dz$$

and it turns, using again the third point of Lemma 25,

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[(A_d^2 g_{n,i})(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=R(\theta,\Delta_{n,i}^{-\beta},x).$$

Pieces things together we get

$$\sup_{u_2\in[t_i,t_{i+1}]}|\mathbb{E}[A^2 g_{n,i}(X_{u_2}^\theta)|X_{t_i}^\theta=x]|=$$

$$=R(\theta,\Delta_{n,i}^{1-\alpha\beta-\epsilon-3\beta},x)+R(\theta,\Delta_{n,i}^{-2\beta},x)+R(\theta,\Delta_{n,i}^{-\beta},x)=R(\theta,\Delta_{n,i}^{-2\beta},x),$$

where $R(\theta,\Delta_{n,i}^{1-\alpha\beta-\epsilon-3\beta},x)$ is negligible compared to $R(\theta,\Delta_{n,i}^{-2\beta},x)$ because, for each choice of α and β, we can find an ϵ arbitrarily small such that $1-\alpha\beta-\epsilon-\beta$ is more than 0. We substitute it in Dynkin's formula and we obtain

$$\mathbb{E}[g_{n,i}(X_{t_{i+1}}^\theta)|X_{t_i}^\theta=x]=$$

$$= \Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) - 1]F(z)dz) + \frac{\Delta_{n,i}^{2}}{2}R(\theta, \Delta_{n,i}^{-2\beta}, x). \quad (1.74)$$

We use the definition of $\Delta_{n,i}$ and the property (3.24) on R, then we substitute in (1.74) getting (1.13).
We now want to prove (2.31). From the expansion (1.13) and the property (3.23) of R, there exists $k_0 > 0$ such that for $|x| \leq \Delta_{n,i}^{k_0}$, $\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] \geq \frac{1}{2}$ $\forall n, i \leq n$: we are avoiding the possibility that the denominator is in the neighborhood of 0. Using (1.58), (1.74) and (1.12) we have that

$$m_\theta(x) = x + \frac{\Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) - 1]F(z)dz) + R(\theta, \Delta_{n,i}^{2-2\beta}, x)}{1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-3\beta)}, x)}. \quad (1.75)$$

Now we can use that R in the denominator is a rest function and so we obtain

$$\frac{1}{1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-3\beta)}, x)} \sim 1 - R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-3\beta)}, x). \quad (1.76)$$

Replacing (1.76) in (1.75) we get

$$m_\theta(x) = x + [\Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) - 1]F(z)dz) +$$

$$+ R(\theta, \Delta_{n,i}^{2-2\beta}, x)](1 - R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-3\beta)}, x)).$$

The expansion (2.31) follows. □

1.7.4 Proof of Theorem 5

Proof. Let us now consider an expansion of (1.58) in the case where α is in $[1, 2)$. Again, we skip the study of $\alpha = 1$ to avoid the emergence of logarithmic functions; as it is embedded in the study of $\alpha > 1$ with the choice of α arbitrarily close to 1.
We start observing that (1.59) and (1.61) still hold; we want to show that even in this case the last term of (1.59) is negligible compared to the others. Again, we consider its decomposition in continuous and discrete part.
Concerning $A_c^2 g_{i,n}$, (1.62) is still true. Let us now consider $A_c(A_d g_{i,n})(y)$ as written in (1.63). We act as in the proof of Theorem 4, using Taylor development up to second order, on the function $h_{i,n}$ defined in (1.64). Hence we obtain the following estimation:

$$|h_{i,n}(y,z)| \leq \left\| g_{i,n}'' \right\|_\infty \frac{|z|^2 \gamma(y)^2}{2}$$

and in the same way, using also (1.67),

$$|\frac{\partial h_{i,n}}{\partial y}(y,z)| \leq \left\| g_{i,n}'' \right\|_\infty |z|^2 |\gamma(y)\gamma'(y)| + \left\| g_{i,n}''' \right\|_\infty |z|^2 \frac{\gamma^2(y)}{2} |1 + \gamma'(y) z| \leq$$

$$\leq |z|^2 P(y) |\Delta_{n,i}|^{-\beta} + |\Delta_{n,i}|^{-2\beta} P(y)(|z|^2 + |z|^3), \quad (1.77)$$

$$|\frac{\partial^2 h_{i,n}}{\partial y^2}(y,z)| \leq |\Delta_{n,i}^{-\beta}| |z|^2 P(y) + |\Delta_{n,i}|^{-2\beta} P(y)(|z|^2 + |z|^3) + |\Delta_{n,i}|^{-3\beta} P(y)(|z|^2 + |z|^3). \quad (1.78)$$

Since a^2 and b have polynomial growth, (1.78) provides us an estimation on $|A_c h_{i,n}(\cdot, z)(y)|$. Using dominated convergence theorem, (1.63), the estimation of $|A_c h_{i,n}(\cdot, z)(y)|$ obtained from (1.78) and the fact that both $\int_{\mathbb{R}}(|z|^2 + |z|^3)F(z)dz$ and $\int_{\mathbb{R}}(|z|^2 + |z|^3)F(z)dz$ are finite, we get

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[A_c(A_d g_{n,i})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = \tag{1.79}$$

$$= R(\theta, \Delta_{n,i}^{-\beta}, x) + R(\theta, \Delta_{n,i}^{-2\beta}, x) + R(\theta, \Delta_{n,i}^{-3\beta}, x) = R(\theta, \Delta_{n,i}^{-3\beta}, x).$$

We now consider $A_d(A_c g_{i,n})(y)$. Using (1.70) and the development to the second order of the function $A_c g_{i,n}(y + z\gamma(y))$ we obtain

$$|A_d(A_c g_{i,n})(y)| \le c \int_{\mathbb{R}} \int_0^1 |(A_c g_{i,n})''(y + s\, z\gamma(y))||z|^2 |\gamma^2(y)| F(z) ds dz. \tag{1.80}$$

We observe that $(A_c g_{i,n})''(y) = [b'' g'_{i,n} + 2b' g''_{i,n} + b g'''_{i,n} + (a')^2 g''_{i,n} + a(a'' g''_{i,n} + a' g'''_{i,n}) + 2aa' g'''_{i,n} + \frac{1}{2} a^2 g^{(4)}_{i,n}](y)$. Using (1.67), to which we add $|g^{(4)}_{i,n}(y)| \le c\Delta_{n,i}^{-3\beta}$, we get

$$|(A_c g_{i,n})''(y)| \le c\, P(y) \Delta_{n,i}^{-3\beta}. \tag{1.81}$$

Using (1.80) and (1.81) it yields

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_d A_c g_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{-3\beta}, x). \tag{1.82}$$

To conclude, we consider $A_d A_d g_{i,n}$. Using (1.71) and the development up to the second order we get

$$|A_d(A_d g_{i,n})(y)| \le c \int_{\mathbb{R}} \int_0^1 |(A_d g_{i,n})''(y + s\, z\gamma(y))||z|^2 |\gamma^2(y)| F(z) ds dz.$$

We recall that (1.73) still holds, with $h_{i,n}$ defined in (1.64). In order to estimate $(A_d g_{i,n})''(y)$ in the case where $\alpha \in [1, 2)$ we use therefore (1.78) joint with dominated convergence theorem. It provides us

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[(A_d A_d g_{i,n})(X_{u_2}^\theta)|X_{t_i}^\theta = x]| = R(\theta, \Delta_{n,i}^{-3\beta}, x). \tag{1.83}$$

Using (1.62), (1.79), (1.82) and (1.83) we put the pieces together and so we obtain

$$\sup_{u_2 \in [t_i, t_{i+1}]} |\mathbb{E}[A^2 f_{i,n}(X_{u_2}^\theta)|X_{t_i}^\theta = x]| =$$

$$= R(\theta, \Delta_{n,i}^{1-\alpha\beta-3\beta-\epsilon}, x) + R(\theta, 1, x) + R(\theta, \Delta_{n,i}^{-3\beta}, x) = R(\theta, \Delta_{n,i}^{-3\beta}, x).$$

Indeed, since ϵ is arbitrarily small, for each choice of α and β we can find ϵ such that $1 - \alpha\beta - 3\beta - \epsilon > -3\beta$. We substitute in the Dynkin formula (1.59) and so we get

$$\mathbb{E}[g_{n,i}(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] =$$

$$= \Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^\beta}(z\gamma(x)) - 1]F(z)dz) + \frac{\Delta_{n,i}^2}{2} R(\theta, \Delta_{n,i}^{-3\beta}, x). \tag{1.84}$$

We use the definition of $\Delta_{n,i}$ and the property (3.24) on R, then we substitute in (1.84) getting (1.16).

In order to prove (1.17), we observe again that from the expansion (1.16) and the property (3.23) of R, there exists $k_0 > 0$ such that for $|x| \le \Delta_{n,i}^{k_0}$, $\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|X_{t_i}^{\theta} = x] \ge \frac{1}{2}$ $\forall n, i \le n$. Using (1.58), (1.84) and (1.15) we have that

$$m_\theta(x) = x + \frac{\Delta_{n,i}(b(x,\theta) + \int_{\mathbb{R}\setminus\{0\}} z\gamma(x)[\varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) - 1]F(z)dz) + R(\theta, \Delta_{n,i}^{2-3\beta}, x)}{1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-4\beta)}, x)}. \tag{1.85}$$

Now R in the denominator is a rest function and so

$$\frac{1}{1 + R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-4\beta)}, x)} \sim 1 - R(\theta, \Delta_{n,i}^{(1-\alpha\beta)\wedge(2-4\beta)}, x). \tag{1.86}$$

We now replace (1.86) in (1.85) and we observe that multiplying by R we obtain negligible functions, hence we get (1.17). □

Let us now prove the development of $m_{\theta,\Delta_{n,i}}$ in the particular case with finite intensity that makes possible to approximate explicitly the contrast function.

1.7.5 Proof of Proposition 2

Proof. We want to use again Dynkin's formula (2.123). We consider the decomposition of the generator: $A = A_c + A_d$ and, by the Remark 1 and the fact that we are in the finite intensity case, we can take $A_d f(x) = \int_{\mathbb{R}} \lambda[f(x+\gamma(x)z) - f(x)]F_0(z)dz$, where $F(z) = \lambda F_0(z)$ and $\int_{\mathbb{R}} F_0(z)dz = 1$.
Concerning the denominator, we denote again $f_{i,n}(y) := \varphi_{\Delta_{n,i}^{\beta}}(y - x)$ and, in order to calculate $A^k f_{i,n}(y)$ we introduce the following set of functions:

$$\mathcal{F}^p := \left\{ g(y)\, s.\,t.\, g(y) = \sum_{k=0}^{p} \varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})\Delta_{n,i}^{-k\beta}(\sum_{j=0}^{k} h_{k,j}(y)\Delta_{n,i}^{\beta j}) \right\}$$

where, $\forall k, j$, $\forall l \ge 0$ $\exists c$ such that $|\frac{\partial^l}{\partial y^l} h_{k,j}(y)| \le c(1 + |y|^c)$ and $\forall k, j$ $h_{k,j}$ is $\mathcal{C}^\infty$.
We observe that, if $g \in \mathcal{F}^p$, then $g' \in \mathcal{F}^{p+1}$, bg and $a^2 g$ are in $\mathcal{F}^p$ and therefore if $g \in \mathcal{F}^p$, then $Ag \in \mathcal{F}^{p+2}$.
We now want to show that, for $g \in \mathcal{F}^p$, A_d acts like $-\lambda I_d$ up to an error term. Indeed,

$$A_d g(y) = \int_{\mathbb{R}} \lambda[g(y+\gamma(y)z) - g(y)]F_0(z)dz = \lambda \int_{\mathbb{R}} g(y+\gamma(y)z)F_0(z)dz - \lambda g(y). \tag{1.87}$$

Let us start considering $g(y) = \varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})h(y)$, where $k \le p$ and $h \in \mathcal{C}^\infty$ is such that $\forall l \ge 0$ $\exists c$: $|\frac{\partial^l}{\partial y^l} h(y)| \le c(1+|y|^c)$. Then,

$$\int_{\mathbb{R}} g(y+\gamma(y)z)F_0(z)dz = \int_{\mathbb{R}} \varphi^{(k)}((y+\gamma(y)z - x)\Delta_{n,i}^{-\beta})\, h(y+\gamma(y)z)F_0(z)dz.$$

With the change of variable $u := (y + \gamma(y)z - x)\Delta_{n,i}^{-\beta}$ it becomes equal to

$$\frac{\Delta_{n,i}^{\beta}}{\gamma(y)} \int_{\mathbb{R}} \varphi^{(k)}(u)h(x + u\Delta_{n,i}^{\beta})F_0(\frac{x-y}{\gamma(y)} + \frac{\Delta_{n,i}^{\beta} u}{\gamma(y)})du. \tag{1.88}$$

We define $\tilde{F}(x,y,s) := \frac{h(x+s)}{\gamma(y)} F_0(\frac{x-y}{\gamma(y)} + \frac{s}{\gamma(y)})$ and we develop it up to the M-order, getting

$$\tilde{F}(x,y,\Delta_{n,i}^{\beta}u) =$$

$$= \sum_{j=0}^{M} \frac{\partial^j \tilde{F}}{\partial s^j}(x,y,0)(\Delta_{n,i}^{\beta}u)^j + \int_0^1 \frac{\partial^{M+1}}{\partial s^{M+1}} \tilde{F}(x,y,t\Delta_{n,i}^{\beta}u) \frac{(1-t)^M}{M!}(\Delta_{n,i}^{\beta}u)^{M+1} dt.$$

Replacing the development in (1.88) and recalling that by the definition of φ we have $\int_{\mathbb{R}} u^j \varphi^{(k)}(u) du = 0$, we get

$$\int_{\mathbb{R}} \varphi^{(k)}(u) \tilde{F}(x,y,\Delta_{n,i}^{\beta}u) du = \tag{1.89}$$

$$= \sum_{j=0}^{M} 0 + \int_{\mathbb{R}} \int_0^1 \varphi^{(k)}(u) \frac{\partial^{M+1}}{\partial s^{M+1}} \tilde{F}(x,y,t\Delta_{n,i}^{\beta}u) \frac{(1-t)^M}{M!}(\Delta_{n,i}^{\beta}u)^{M+1} dt\, du.$$

We observe that it is $|\frac{\partial^{l_1+l_2+l_3}}{\partial s^{l_1} \partial x^{l_2} \partial y^{l_3}} \tilde{F}(x,y,s)| \le c(1+|x|^c+|y|^c+|s|^c)$. Therefore, since the support of $\varphi^{(k)}$ is compact, we get

$$\int_{\mathbb{R}} \int_0^1 \varphi^{(k)}(u) \frac{\partial^{M+1}}{\partial s^{M+1}} \tilde{F}(x,y,t\Delta_{n,i}^{\beta}u) \frac{(1-t)^M}{M!}(\Delta_{n,i}^{\beta}u)^{M+1} dt\, du \le \tag{1.90}$$

$$\le c(\Delta_{n,i}^{\beta})^{M+1}(1+|x|^c+|y|^c).$$

Hence using (1.88) and (1.90) on $| \int_{\mathbb{R}} g(y+\gamma(y)z) F_0(z) dz|$ and the differentiation of (1.89) on $|\frac{\partial^l}{\partial y^l} \int_{\mathbb{R}} g(y+\gamma(y)z) F_0(z) dz|$ we get that both of them are upper bounded by $c(1+|x|^c+|y|^c)\Delta_{n,i}^{\beta(M+2)}$, where in the second case the constant c depends on l. Turning to a general function $g \in \mathcal{F}^p$, the estimations above become

$$| \int_{\mathbb{R}} g(y+\gamma(y)z) F_0(z) dz| \le c(1+|x|^c+|y|^c)\Delta_{n,i}^{\beta(M+2)} \Delta_{n,i}^{-\beta p} \tag{1.91}$$

and, $\forall l \ge 1$,

$$|\frac{\partial^l}{\partial y^l} \int_{\mathbb{R}} g(y+\gamma(y)z) F_0(z) dz| \le c_l(1+|x|^{c_l}+|y|^{c_l})\Delta_{n,i}^{\beta(M+2)} \Delta_{n,i}^{-\beta p}. \tag{1.92}$$

We introduce the set of functions

$$\mathcal{R}^p := \left\{ r(x,y,\Delta_{n,i}^{p}) \text{ such that } \forall l \ge 0 \ \exists c_l \ |\frac{\partial^l}{\partial y^l} r(x,y,\Delta_{n,i}^{p})| \le c_l(1+|x|^{c_l}+|y|^{c_l})\Delta_{n,i}^{p} \right\}.$$

Hence, using (1.87), (2.88) and (1.92) we have proved that, $\forall g \in \mathcal{F}^p$,

$$A_d g(y) = -\lambda g(y) + r(x,y,\Delta_{n,i}^{\beta(M+2-p)}). \tag{1.93}$$

We observe that if a function r is in $\mathcal{R}^p$, then both $A_d r$ and $A_c r$ are in $\mathcal{R}^p$. We can therefore now calculate for $f_{i,n}(y) = \varphi((y-x)\Delta_{n,i}^{-\beta})$, $f_{i,n} \in \mathcal{F}^0$,

$$A_{i_1} f_{i,n}(y) = \begin{cases} A_c f_{i,n}(y) & \text{if } i_1 = c \\ A_d f_{i,n}(y) = -\lambda f_{i,n}(y) + r(x,y,\Delta_{n,i}^{\beta(M+2)}) & \text{if } i_1 = d, \end{cases} \tag{1.94}$$

We want to show, by recurrence, that

$$A_{i_N} \circ ... \circ A_{i_1}(f_{i,n})(y) = A_c^{l(i_1,...,i_N)} f_{i,n}(y)(-\lambda)^{N-l(i_1,...,i_N)} + r(x,y,\Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N)}), \tag{1.95}$$

with $l(i_1, ..., i_N)$ the number of c in $\{i_1, ..., i_N\}$. Let us consider the base case

$$A_{i_2} \circ A_{i_1} f_{i,n}(y) = \begin{cases} A_c^2 f_{i,n}(y) & \text{if } i_2 = i_1 = c \\ -\lambda A_c f_{i,n}(y) + r(x, y, \Delta_{n,i}^{\beta(M+2)}) & \text{if } i_2 = c,\ i_1 = d \\ -\lambda A_c f_{i,n}(y) + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta}) & \text{if } i_2 = d,\ i_1 = c \\ \lambda^2 f_{i,n}(y) + r(x, y, \Delta_{n,i}^{\beta(M+2)}) & \text{if } i_2 = i_1 = d, \end{cases} \tag{1.96}$$

where in the third case we have used $A_c f_{i,n} \in \mathcal{F}^2$. So we have

$$A_{i_2} \circ A_{i_1} f_{i,n}(y) = A_c^{l(i_1,i_2)} f_{i,n}(y)(-\lambda)^{2-l(i_1,i_2)} + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,i_2)}),$$

as we wanted. For the inductive step, we assume that (1.95) holds, now

$$A_{i_{N+1}} \circ A_{i_N} \circ ... \circ A_{i_1}(f_{i,n})(y) =$$

$$= \begin{cases} A_c \circ A_c^{l(i_1,...,i_N)} f_{i,n}(y)(-\lambda)^{N-l(i_1,...,i_N)} + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N)}) & \text{if } i_{N+1} = c, \\ (-\lambda) A_c^{l(i_1,...,i_N)} f_{i,n}(y)(-\lambda)^{N-l(i_1,...,i_N)} + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N)}) & \text{if } i_{N+1} = d, \end{cases} \tag{1.97}$$

where in the first case we have used that $A_c r(x, y, \Delta_{n,i}^h) \in \mathcal{R}^h$, $\forall h$, and in the second case that $A_d r(x, y, \Delta_{n,i}^h) \in \mathcal{R}^h$ and that $A_c^{l(i_1,...,i_N)} f_{i,n} \in \mathcal{F}^{2l(i_1,...,i_N)}$ while using (1.93). It is equal to

$$A_c^{l(i_1,...,i_N,i_{N+1})} f_{i,n}(y)(-\lambda)^{N+1-l(i_1,...,i_N,i_{N+1})} + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N,i_{N+1})})$$

and therefore the recurrence is proved. We can now calculate $A^k f_{i,n}(x)$ in the Dynkin's formula (2.123) using (1.95):

$$A^k f_{i,n}(x) = \sum_{(i_1,...,i_k) \in \{c,d\}^k} (A_{i_k} \circ ... \circ A_{i_1}) f_{i,n}(x) =$$

$$= \sum_{(i_1,...,i_k) \in \{c,d\}^k} A_c^{l(i_1,...,i_k)} f_{i,n}(x)(-\lambda)^{k-l(i_1,...,i_k)} + r(x, x, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_k)}). \tag{1.98}$$

Recalling that $A_c^l f_{i,n}(x) = 0$ $\forall l \geq 1$, (1.98) becomes
$(-\lambda)^k f_{i,n}(x) + r(x, x, \Delta_{n,i}^{\beta(M+2)-2\beta k})$.
Therefore, the principal term in the development of the denominator of $m_{\theta,\Delta_{n,i}}(x)$ from Dynkin's formula up to order N is

$$\sum_{k=0}^{N} \frac{\Delta_{n,i}^k}{k!} A^k f_{i,n}(x) = \sum_{k=0}^{N} \frac{\Delta_{n,i}^k}{k!} (-\lambda)^k f_{i,n}(x) + r(x, x, \Delta_{n,i}^{\beta(M+2)-2\beta k+k}).$$

Let us now consider the term of rest in the Dynkin's formula (2.123). Observing that

$$|A_c^{N+1} f_{i,n}(y)| \leq \Delta_{n,i}^{-2\beta(N+1)}(1 + |y|^c)$$

using (1.95) and the definition of the function r, we get that

$$|A^{N+1} f_{i,n}(y)| \leq c(\Delta_{n,i}^{-2\beta(N+1)} + \Delta_{n,i}^{\beta(M+2)-2\beta(N+1)})(1 + |y|^c). \tag{1.99}$$

Therefore

$$\mathbb{E}[|A^{N+1} f_{i,n}(X_{u_{n+1}})||X_{t_i} = x] \leq c(\Delta_{n,i}^{-2\beta(N+1)} + \Delta_{n,i}^{\beta(M+2)-2\beta(N+1)})(1 + |x|^c). \tag{1.100}$$

Replacing in (2.123) it yields

$$|\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}...\int_{t_i}^{u_N}\mathbb{E}[A^{N+1}f_{i,n}(X_{u_{n+1}})|X_{t_i}=x]du_{N+1}...du_2du_1|\leq$$

$$\leq c\Delta_{n,i}^{N+1}(\Delta_{n,i}^{-2\beta(N+1)}+\Delta_{n,i}^{\beta(M+2)-2\beta(N+1)})(1+|x|^c).$$

Since $\Delta_{n,i}^{\beta(M+2)-2\beta(N+1)}$ is negligible compared to $\Delta_{n,i}^{-2\beta(N+1)}$, it is enough to have $(N+1)(1-2\beta)\geq\lfloor\beta(M+2)\rfloor$ in order to get the following development of the denominator $d_{\Delta_{n,i}}(x)$ of $m_{\theta,\Delta_{n,i}}(x)$:

$$d_{\Delta_{n,i}}(x)=\sum_{k=0}^{N}\frac{\Delta_{n,i}^k}{k!}(-\lambda)^k f_{i,n}(x)+r(x,x,\Delta_{n,i}^{\beta(M+2)+(1-2\beta)k})+r(x,x,\Delta_{n,i}^{(1-2\beta)(N+1)})=$$

$$=\sum_{k=0}^{\lfloor\beta(M+2)\rfloor}\frac{\Delta_{n,i}^k}{k!}(-\lambda)^k+r(x,x,\Delta_{n,i}^{\beta(M+2)}),$$

where we have also used that, by the definition of $f_{i,n}$, $f_{i,n}(x)=1$ and in the sum we have considered only the terms up to $k=\lfloor\beta(M+2)\rfloor$ because the others are rest terms.
Let us now study the numerator $n_{\Delta_{n,i}}(x)$ of $m_{\theta,\Delta_{n,i}}(x)$: acting like in the proof of Theorem 3 we consider $\bar{g}(y):=(y-x)\varphi((y-x)\Delta_{n,i}^{-\beta})$. Let us introduce, in place of $\mathcal{F}^p$, the set $\tilde{\mathcal{F}}^p$.

$$\tilde{\mathcal{F}}^p:=\left\{\tilde{g}(y)\ s.t.\ \tilde{g}(y)=\sum_{k=0}^{p}\varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})\Delta_{n,i}^{-k\beta}(\sum_{j=0}^{k}h_{k,j}(x,y)\Delta_{n,i}^{\beta j})\right\}$$

where, $\forall k,j$, $\forall l\geq 0$, $\exists c_l$ such that $|\frac{\partial^l}{\partial y^l}h_{k,j}(x,y)|\leq c_l(1+|x|^{c_l}+|y|^{c_l})$. We observe that, as it was for $\mathcal{F}^p$, if $\tilde{g}\in\tilde{\mathcal{F}}^p$ then $A\tilde{g}\in\tilde{\mathcal{F}}^{p+2}$ and, for all $\tilde{g}\in\tilde{\mathcal{F}}^p$,

$$A_d\tilde{g}(y)=-\lambda\tilde{g}(y)+r(x,y,\Delta_{n,i}^{\beta(M+2-p)}).\tag{1.101}$$

It turns that the same relation as (1.95) holds with $\bar{g}$ in place of $f_{i,n}$. Hence we get

$$A^k\bar{g}(y)=(A_c+A_d)^k\bar{g}(y)=$$

$$=\sum_{(i_1,...,i_k)\in\{c,d\}^k}A_c^{l(i_1,...,l_k)}\bar{g}(y)(-\lambda)^{k-l(i_1,...,i_k)}+r(x,y,\Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_k)})=\tag{1.102}$$

$$=\sum_{l=0}^{k}\binom{k}{l}(-\lambda)^{k-l}A_c^l\bar{g}(y)+r(x,y,\Delta_{n,i}^{\beta(M+2)-2\beta k}),$$

where $l(i_1,...,i_k)$ is the number of c in $\{i_1,...,i_k\}$ and $\binom{k}{l}$ are the binomial coefficients.
Now, concerning the continuous part of the generator, since it is local and $\bar{g}(y)=(y-x)$ in the neighborhood of x, we find $A_c^l\bar{g}(x)=A_K^{(l)}(x)$, which are exactly the coefficients found in the case without jump studied by Kessler in [51].
By (1.102), the principal term in the development of the numerator is therefore

$$\sum_{k=0}^{N}\frac{\Delta_{n,i}^k}{k!}A^kg(x)=\sum_{k=0}^{N}\frac{\Delta_{n,i}^k}{k!}(\sum_{l=0}^{k}\binom{k}{l}(-\lambda)^{k-l}A_K^{(l)}(x)+r(x,x,\Delta_{n,i}^{\beta(M+2)-2\beta k}))=$$

$$= \sum_{k=0}^{N} \frac{\Delta_{n,i}^k}{k!} (\sum_{l=0}^{k} \binom{k}{l} (-\lambda)^{k-l} A_K^{(l)}(x)) + r(x, x, \Delta_{n,i}^{\beta(M+2)}). \tag{1.103}$$

Changing the order of summation and introducing $k' := k - l$ we get that the first term of the previous equation is equal to

$$\sum_{l=0}^{N} \sum_{k=l}^{N} \frac{\Delta_{n,i}^k}{k!} \binom{k}{l} (-\lambda)^{k-l} A_K^{(l)}(x) = \sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_K^{(l)}(x) \sum_{k'=0}^{N-l} \Delta_{n,i}^{k'} (-\lambda)^{k'} \frac{l!}{(k'+l)!} \binom{l+k'}{l} =$$

$$= \sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_K^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'} (-\lambda)^{k'}}{k'!}, \tag{1.104}$$

where in the last equality we have used the definition of binomial coefficients. Concerning the rest term in the Dynkin's formula, we use again (1.99) and (1.100) with $\bar{g}$ in place of $f_{i,n}$ and it turns again

$$| \int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} ... \int_{t_i}^{u_N} \mathbb{E}[A^{N+1} \bar{g}(X_{u_{N+1}}) | X_{t_i} = x] du_{N+1} ... du_2 du_1 | \le r(x, x, \Delta_{n,i}^{(1-2\beta)(N+1)}). \tag{1.105}$$

Hence, using (1.103), (2.125) and (2.126) we have the following development:

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_K^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'} (-\lambda)^{k'}}{k'!} + r(x, x, \Delta_{n,i}^{\beta(M+2)}) + r(x, x, \Delta_{n,i}^{(1-2\beta)(N+1)}). \tag{1.106}$$

If $(N+1)(1-2\beta) \ge \beta(M+2)$, it entails

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^l}{l!} A_K^{(l)}(x) \sum_{k'=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^{k'} (-\lambda)^{k'}}{k'!} + r(x, x, \Delta_{n,i}^{\beta(M+2)}).$$

Acting as in the proof of the development of m_θ given in Theorem 3 we can say that it exists $k_0 > 0$ such that, for $|x| \le \Delta_{n,i}^{-k_0}$, the development of $m_{\theta, \Delta_{n,i}}(x)$ is

$$x + \frac{n_{\Delta_{n,i}}(x)}{d_{\Delta_{n,i}}(x)} = x + \sum_{l=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^l}{l!} A_K^{(l)}(x) + r(x, x, \Delta_{n,i}^{\beta(M+2)}). \tag{1.107}$$

The expansion (1.23) follows after remarking that $A_K^{(0)}(x) = 0$. □

1.7.6 Contrast convergence

Before proving the contrast convergence, let us define $r(\theta, x)$ as the particular rest function that turns out from the development of $m_{\theta, \Delta_{n,i}}$:

$$r(\theta, x) := m_{\theta, \Delta_{n,i}}(x) - x - \Delta_{n,i} b(x, \theta) - \Delta_{n,i} \int_{\mathbb{R} \setminus \{0\}} z \gamma(x) [1 - \varphi_{\Delta_{n,i}^\beta}(\gamma(x) z)] F(z) dz. \tag{1.108}$$

We recall that $r(\theta, x)$ is $R(\theta, \Delta_{n,i}^{1+\delta}, x)$ with $\delta > 0$ as defined below equation (1.30). In order to prove the consistency and asymptotic normality of the estimator, the first step is the following Lemma:

Lemma 4. *Suppose that Assumptions 1-5 and A_β are satisfied. Then*

$$\frac{U_n(\theta) - U_n(\theta_0)}{t_n} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\theta) - b(x,\theta_0))^2}{a^2(x)} \pi(dx) \tag{1.109}$$

Proof. By the definition,

$$U_n(\theta) = \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m_\theta(X_{t_i}))^2}{a^2(X_{t_i})\Delta_{n,i}} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}.$$

We want to reformulate the contrast function, in order to compensate for the terms not depending on θ in the difference $U_n(\theta) - U_n(\theta_0)$.
The dynamic of the process X is known and so we can write

$$X_{t_{i+1}} = X_{t_i} + \int_{t_i}^{t_{i+1}} b(X_s, \theta_0) ds + \int_{t_i}^{t_{i+1}} a(X_s) dW_s + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} \gamma(X_{s^-}) z \tilde{\mu}(ds, dz).$$

We have proved the development (2.31) of m_θ, too. We can substitute both of them in $U_n(\theta)$, getting

$$U_n(\theta) = \sum_{i=0}^{n-1} \frac{1}{a^2(X_{t_i})\Delta_{n,i}} [X_{t_i} + \int_{t_i}^{t_{i+1}} b(X_s, \theta_0) ds + \int_{t_i}^{t_{i+1}} a(X_s) dW_s +$$

$$+ \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} \gamma(X_{s^-}) z \tilde{\mu}(ds, dz) - X_{t_i} + \Delta_{n,i}(-b(X_{t_i}, \theta) +$$

$$+ \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(X_{t_i})\, (1 - \varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i}) z))\, F(z) dz) + r(\theta, X_{t_i})]^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}$$

$$= \sum_{i=0}^{n-1} \frac{1}{a^2(X_{t_i})\Delta_{n,i}} (\int_{t_i}^{t_{i+1}} b(X_s, \theta_0) ds - \Delta_{n,i}\, b(X_{t_i}, \theta) + \zeta_i +$$

$$+ r(\theta, X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}},$$

we recall the definition of

$$\zeta_i := \int_{t_i}^{t_{i+1}} a(X_s) dW_s + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}) z \tilde{\mu}(ds, dz) +$$

$$+ \Delta_{n,i} \int_{\mathbb{R}\setminus\{0\}} z\, \gamma(X_{t_i})\, [1 - \varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i}) z)]\, F(z) dz,$$

as in (1.29); we point out that ζ_i does not depend on θ.
In the same way

$$U_n(\theta_0) = \sum_{i=0}^{n-1} \frac{1}{a^2(X_{t_i})\Delta_{n,i}} (\int_{t_i}^{t_{i+1}} b(X_s, \theta_0) ds - \Delta_{n,i}\, b(X_{t_i}, \theta_0) +$$

$$+ \zeta_i + r(\theta_0, X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}$$

and so

$$\frac{U_n(\theta) - U_n(\theta_0)}{t_n} = \frac{1}{t_n} \sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}} [\Delta_{n,i}^2 (b(X_{t_i}, \theta)^2 - b(X_{t_i}, \theta_0)^2) +$$

$$+2\Delta_{n,i}\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds(b(X_{t_i},\theta_0)-b(X_{t_i},\theta))+A_i+B_i+C_i+D_i+E_i], \quad (1.110)$$

with

$$A_i = 2\zeta_i\Delta_{n,i}(b(X_{t_i},\theta_0)-b(X_{t_i},\theta)), \qquad B_i = 2\zeta_i(r(\theta,X_{t_i})-r(\theta_0,X_{t_i})),$$

$$C_i = 2\Delta_{n,i}(r(\theta_0,X_{t_i})b(X_{t_i},\theta_0)-r(\theta,X_{t_i})b(X_{t_i},\theta)), \qquad D_i = r(\theta,X_{t_i})^2-r(\theta_0,X_{t_i})^2,$$

$$E_i = 2\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds(r(\theta,X_{t_i})-r(\theta_0,X_{t_i})).$$

Our goal is to show that the contribution of A_i, B_i, C_i, D_i and E_i go to zero in probability as $n\to\infty$ and to prove that the other terms converge to $\int_{\mathbb{R}}\frac{(b(x,\theta)-b(x,\theta_0))^2}{a^2(x)}\pi(dx)$. We observe that the rest function $r(\theta,x)$ is present in all the terms that have to converge to 0 but A_i, on which we use a different motivation to obtain the convergence:

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}A_i =$$

$$=\frac{1}{t_n}\sum_{i=0}^{n-1}\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})f_{i,n}(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\zeta_i,$$

with $f_{i,n}(X_{t_i},\theta):=\frac{2}{a^2(X_{t_i})}(b(X_{t_i},\theta_0)-b(X_{t_i},\theta))$.
In order to apply Proposition 4 we observe that, by the assumptions done on the coefficients, $f_{i,n}$ has polynomial growth. We therefore get the convergence to zero in probability, using Proposition 4.
We want to show that $\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}B_i\xrightarrow{\mathbb{P}}0$ and so we observe that, by the definition of the function r and by (3.24) we have that

$$r(\theta,X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} = R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i}) = \Delta_{n,i}^{1+\delta}R(\theta,1,X_{t_i}). \quad (1.111)$$

Hence

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}B_i =$$

$$=\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}^{\delta}\zeta_i\frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}(R(\theta,1,X_{t_i})-R(\theta_0,1,X_{t_i}))}{a^2(X_{t_i})}$$

To prove the convergence, we have to show that

$$\frac{1}{t_n}\sum_{i=0}^{n-1}|\mathbb{E}[\Delta_{n,i}^{\delta}f_{i,n}(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\zeta_i\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]|\xrightarrow{\mathbb{P}}0, \quad (1.112)$$

and

$$\frac{1}{(t_n)^2}\sum_{i=0}^{n-1}\mathbb{E}[\Delta_{n,i}^{2\delta}f_{i,n}^2(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\zeta_i^2\varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]\xrightarrow{\mathbb{P}}0,$$

with

$$f_{i,n}(X_{t_i},\theta)=\frac{R(\theta,1,X_{t_i})-R(\theta_0,1,X_{t_i})}{a^2(X_{t_i})}.$$

By the measurability of X_{t_i} with respect to $\mathcal{F}_{t_i}$, by the fact that $|\Delta_{n,i}| \le \Delta_n$ and that $t_n = 0(n\Delta_n)$ we get

$$\frac{1}{t_n}\sum_{i=0}^{n-1} |\mathbb{E}[\Delta_{n,i}^{\delta} f_{i,n}(X_{t_i},\theta) 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}} \zeta_i \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]| =$$

$$= \frac{1}{t_n}\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta} |f_{i,n}(X_{t_i},\theta)||\mathbb{E}[1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}} \zeta_i \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]| \le$$

$$\le \Delta_n^{\delta} \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i},\theta)||\mathbb{E}[1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}} \zeta_i \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]|.$$

We recall that δ is positive. Using (1.31), we get the convergence (1.112) in L^1 and thus in probability.
In the same way,

$$\frac{1}{(t_n)^2}\sum_{i=0}^{n-1} \mathbb{E}[\Delta_{n,i}^{2\delta} f_{i,n}^2(X_{t_i},\theta) 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}} \zeta_i^2 \varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}] \le$$

$$\le \Delta_n^{2\delta} \frac{c}{n^2\Delta_n^2}\sum_{i=0}^{n-1} f_{i,n}^2(X_{t_i},\theta)\mathbb{E}[1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}} \zeta_i^2 \varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}],$$

that goes to zero in probability using (1.32).

$$\frac{1}{t_n}\sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}} C_i = \frac{2}{t_n}\sum_{i=0}^{n-1} \Delta_{n,i}^{1+\delta} f_{i,n}(X_{t_i},\theta)\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}),$$

with $f_{i,n}(X_{t_i},\theta) := \frac{R(\theta_0,1,X_{t_i})b(X_{t_i},\theta_0) - R(\theta,1,X_{t_i})b(X_{t_i},\theta)}{a^2(X_{t_i})} 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}}$, where we have used (1.111).
In module, it is upper bounded by $\Delta_n^{\delta} \frac{c}{n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i},\theta)\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|$.
We observe that the exponent on Δ_n is positive so it goes to zero as $n \to \infty$ and that $|\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})| \le c$. By the polynomial growth of $f_{i,n}$ and the third point of Lemma 26, we get that $\frac{1}{n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i},\theta)|$ is bounded in L^1. It yields the convergence in probability that we were looking for.
Let us consider D_i. Using triangle inequality, we can just prove the convergence of the following:

$$|\frac{1}{t_n}\sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}} r(\theta, X_{t_i})^2| =$$

$$= |\frac{1}{t_n}\sum_{i=0}^{n-1} \Delta_{n,i}^{1+2\delta} \frac{\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}|\le \Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})} R(\theta,1,X_{t_i})^2| \le$$

$$\le \Delta_n^{2\delta} \frac{1}{n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i},\theta)\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|,$$

$f_{i,n}(X_{t_i},\theta) = \frac{R(\theta,1,X_{t_i})^2}{a^2(X_{t_i})}$, using also the indicator is always upper bounded by 1.
Also this time the exponent on Δ_n is positive. We can use the boundedness of

$|\varphi_{\Delta_{n,i}^\beta}|$, the polynomial growth of $f_{i,n}$ and third point of Lemma 26 in order to get that $\frac{1}{n}\sum_{i=0}^{n-1}|f_{i,n}(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|$ is bounded in L^1 . It turns

$$|\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}r(\theta,X_{t_i})^2|\xrightarrow{\mathbb{P}}0.$$

Considering E_i, we use again the triangle inequality in order to prove only the convergence to zero of the following:

$$|\frac{1}{t_n}\sum_{i=0}^{n-1}2\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}\int_{t_i}^{t_{i+1}}b(X_s,\theta_0)ds\,r(\theta,X_{t_i})|. \tag{1.113}$$

In the sequel it will be useful to substitute $\int_{t_i}^{t_{i+1}}b(X_s,\theta_0)ds$ with $\Delta_{n,i}\,b(X_{t_i},\theta_0)$.

$$\int_{t_i}^{t_{i+1}}b(X_s,\theta_0)ds=\int_{t_i}^{t_{i+1}}[b(X_s,\theta_0)-b(X_{t_i},\theta_0)]ds+\Delta_{n,i}b(X_{t_i},\theta_0). \tag{1.114}$$

In order to show that the first term is negligible compared to $\Delta_{n,i}$, we consider the following expected value:

$$\sup_{u\in[0,\Delta_{n,i}]}\mathbb{E}[|b(X_{t_i+u},\theta_0)-b(X_{t_i},\theta_0)||\mathcal{F}_{t_i}]\le\sup_{u\in[0,\Delta_{n,i}]}\mathbb{E}[\left\|\frac{\partial b}{\partial x}\right\|_\infty|X_{t_i+u}-X_{t_i}||\mathcal{F}_{t_i}]\le$$
$$\le c\sup_{u\in[0,\Delta_{n,i}]}\mathbb{E}[|X_{t_i+u}-X_{t_i}||\mathcal{F}_{t_i}].$$

In the last inequality we have used that the derivative of b is supposed bounded. Using Holder inequality we get that it is, for each $p\ge2$, upper bounded by

$$c\sup_{u\in[0,\Delta_{n,i}]}(\mathbb{E}[|X_{t_i+u}-X_{t_i}|^p|\mathcal{F}_{t_i}])^{\frac{1}{p}}\le$$

$$\le c\sup_{u\in[0,\Delta_{n,i}]}(|t_i+u-t_i|(1+|X_{t_i}|^p))^{\frac{1}{p}}=R(\theta,\Delta_{n,i}^{\frac{1}{p}},X_{t_i}). \tag{1.115}$$

Where, in the last inequality, we have used the second point of Lemma 25.
For $p=2$, $\mathbb{E}[|b(X_{t_i+u},\theta_0)-b(X_{t_i},\theta_0)||\mathcal{F}_{t_i}]\le R(\theta,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})$ and therefore

$$\int_{t_i}^{t_{i+1}}\mathbb{E}[|b(X_s,\theta_0)-b(X_{t_i},\theta_0)||\mathcal{F}_{t_i}]ds\le\int_{t_i}^{t_{i+1}}R(\theta,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})ds=R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}), \tag{1.116}$$

negligible compared to $\Delta_{n,i}$, that is the order of the second term of (1.114).
Using (1.111) and (1.114), (1.113) can be reformulated as

$$|\frac{1}{t_n}\sum_{i=0}^{n-1}2\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\Delta_{n,i}^\delta R(\theta,1,X_{t_i})}{a^2(X_{t_i})}[\Delta_{n,i}\,b(X_{t_i},\theta_0)+$$
$$+\int_{t_i}^{t_{i+1}}[b(X_s,\theta_0)-b(X_{t_i},\theta_0)]ds]|. \tag{1.117}$$

The first term is upper bounded by

$$\Delta_n^\delta\frac{1}{n}\sum_{i=0}^{n-1}|f_{i,n}(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|,$$

where $f_{i,n}(X_{t_i}, \theta) = \frac{2b(X_{t_i}, \theta_0) R(\theta, 1, X_{t_i})}{a^2(X_{t_i})}$.
Again, the exponent on Δ_n is positive and $\frac{1}{n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i}, \theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|$ is bounded in L^1 using the boundedness of $\varphi_{\Delta_{n,i}^\beta}$, the polynomial growth of $f_{i,n}$ and the third point of Lemma 26.
Concerning the second term of (1.117), we observe it is upper bounded by

$$\Delta_n^\delta \frac{1}{n\Delta_n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i}, \theta) \int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds \, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|,$$

where $f_{i,n}(X_{t_i}, \theta) = \frac{2R(\theta, 1, X_{t_i})}{a^2(X_{t_i})}$. The exponent on Δ_n is still positive and $\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} |f_{i,n}(X_{t_i}, \theta) \int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds \, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|$ is bounded in L^1. Indeed,

$$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta) \int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds \, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|] \leq$$

$$\leq \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta) \int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds|] =$$

$$= \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta)\mathbb{E}[\int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds \, |\mathcal{F}_{t_i}]|] = \tag{1.118}$$

$$= \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta) R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i})|],$$

where we have used the definition of conditional expectation and (1.116).
From (3.24), we can upper bound (1.118) by $\Delta_n^{\frac{1}{2}} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta) R(\theta, 1, X_{t_i})|]$. The exponent of Δ_n is clearly positive and $\frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[|f_{i,n}(X_{t_i}, \theta) R(\theta, 1, X_{t_i})|]$ is bounded using again the polynomial growth of both $f_{n,i}$ and R and the third point of Lemma 26.
We have obtained the wanted convergence.

Let us now consider the main terms of (1.110): we will show that they converge to $\int_{\mathbb{R}} \frac{(b(x,\theta) - b(x, \theta_0))^2}{a^2(x)} \pi(dx)$.
In order to do it, we want to replace $\int_{t_i}^{t_{i+1}} b(X_s, \theta_0) ds$ with $\Delta_{n,i} \, b(X_{t_i}, \theta_0)$ in (1.110), getting:

$$\Delta_{n,i}^2 [b(X_{t_i}, \theta)^2 - b(X_{t_i}, \theta_0)^2] + 2\Delta_{n,i}^2 b(X_{t_i}, \theta_0)[b(X_{t_i}, \theta_0) - b(X_{t_i}, \theta)] =$$

$$= \Delta_{n,i}^2 [b(X_{t_i}, \theta) - b(X_{t_i}, \theta_0)]^2.$$

Hence, we can reformulate (1.110) adding and subtracting $\Delta_{n,i} b(X_{t_i}, \theta_0)$. We obtain

$$\frac{U_n(\theta) - U_n(\theta_0)}{t_n} = \frac{1}{t_n}\sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})} \Delta_{n,i} [b(X_{t_i}, \theta) - b(X_{t_i}, \theta_0)]^2 +$$

$$+ \frac{1}{t_n}\sum_{i=0}^{n-1} \frac{2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})} (\int_{t_i}^{t_{i+1}} [b(X_s, \theta_0) - b(X_{t_i}, \theta_0)] ds) \times \tag{1.119}$$

$$\times [b(X_{t_i}, \theta_0) - b(X_{t_i}, \theta)] + R_i,$$

where R_i represents the rest terms, for which we have already shown the convergence to 0 in probability. The second term of (1.119) goes to 0 in L^1, in fact

$$\mathbb{E}[|\frac{1}{t_n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})(\int_{t_i}^{t_{i+1}}(b(X_s,\theta_0)ds-b(X_{t_i},\theta_0))ds|]=$$

$$=\mathbb{E}[|\frac{1}{t_n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})(\int_{t_i}^{t_{i+1}}(b(X_s,\theta_0)-b(X_{t_i},\theta_0))ds|\mathcal{F}_{t_i}]|],$$

With $f(X_{t_i},\theta):=\frac{2(b(X_{t_i},\theta_0)-b(X_{t_i},\theta))}{a^2(X_{t_i})}1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}$.
Using that $\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})$ is bounded by a constant and the estimation (1.116), we get that it is upper bounded by

$$\mathbb{E}[|\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)R(\theta,\Delta_{n,i}^{\frac{3}{2}},X_{t_i})|]\le\Delta_n^{\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f(X_{t_i},\theta)R(\theta,1,X_{t_i})|],$$

where in the last inequality we have used (3.24), the triangle inequality and that $|\Delta_{n,i}|\le\Delta_n$. Using the third point of Lemma 26, we obtain that $\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i},\theta)R(\theta,1,X_{t_i})|$ is bounded in L^1 and so the convergence wanted.
To conclude, we use the second point of Proposition 3 on the first term of (1.119). It yields

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})}\Delta_{n,i}(b(X_{t_i},\theta_0)-b(X_{t_i},\theta))^2\xrightarrow{\mathbb{P}}$$

$$\xrightarrow{\mathbb{P}}\int_{\mathbb{R}}\frac{(b(x,\theta)-b(x,\theta_0))^2}{a^2(x)}\pi(dx).$$

Therefore,

$$\frac{U_n(\theta)-U_n(\theta_0)}{t_n}\xrightarrow{\mathbb{P}}\int_{\mathbb{R}}\frac{(b(x,\theta)-b(x,\theta_0))^2}{a^2(x)}\pi(dx).$$

□

Remark 9. *We observe that the contrast function does not converge:* $\forall\theta\in\Theta$

$$\lim_{n\to\infty}\frac{U_n(\theta)}{t_n}=\infty.$$

It happens because, in the expansion

$$X_{t_{i+1}}-m_\theta(X_{t_i})=\zeta_i+\int_{t_i}^{t_{i+1}}b(X_s,\theta_0)ds-\Delta_{n,i}\,b(X_{t_i},\theta)+R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i}),$$

ζ_i *is of the order* $\Delta_n^{\frac{1}{2}}$ *while the order of the part dependent on* θ *is* Δ_n.
That is the reason why we consider the difference between $U_n(\theta)$ *and* $U_n(\theta_0)$*: stressing that* ζ_i *does not depend on* θ*, we get that in the difference it does not contribute anymore.*
The asymptotic behavior of $(U_n(\theta)-U_n(\theta_0))$ *is therefore governed by the part depending on* θ.

1.7.7 Consistency of the estimator

In order to prove the consistency of $\hat{\theta}_n$, we need that the convergence (1.109) takes place in probability uniformly in the parameter θ, we want therefore to show the uniformity of the convergence in θ.
Let $S_n(\theta) := \frac{U_n(\theta)-U_n(\theta_0)}{t_n}$; we regard this as a random element taking values in $(C(\Theta), \|.\|_\infty)$. It suffices to prove the tightness of this sequence, to do it we need an explicit approximation of $\dot{m}_{\theta,h}$. Such an approximation, together with the approximation of $\ddot{m}_{\theta,h}$, will be also useful to study the asymptotic behavior of the derivatives of the contrast function. In the following proposition we study their asymptotic expansions as $\Delta_{n,i} \to 0$:

Proposition 8. *Suppose that Assumptions 1 to 4 and 7 hold, with $\alpha \in (0,2)$, $\alpha \neq 1$ and $\beta \in (0, \frac{1}{1+\alpha} - \epsilon)$. Then, for $|y| \le h^{-k_0}$ (where k_0 is the same as in Theorem 3 or 5, according to $\alpha < 1$ or $\alpha > 1$),*

$$\dot{m}_{\theta,h}(y) = h\dot{b}(y,\theta) + R(\theta, h^{\frac{3}{2}\wedge(2-\alpha\beta-\epsilon-\beta)}, y) \tag{1.120}$$

and

$$\ddot{m}_{\theta,h}(y) = h\ddot{b}(y,\theta) + R(\theta, h^{\frac{3}{2}\wedge(2-\alpha\beta-\epsilon-\beta)}, y). \tag{1.121}$$

Remark 10. *It is also possible to show that*

$$|\dddot{m}_{\theta,h}(y)| = R(\theta, h, y). \tag{1.122}$$

The proposition above will be proved in the Appendix 1.8.1, where we will also justify (1.122). We can now show the tightness of $S_n(\theta)$:

Lemma 5. *Suppose that Assumptions 1 - 8 and A_β are satisfied. Then*

$$S_n(\theta) := \frac{U_n(\theta) - U_n(\theta_0)}{t_n}$$

is a tight sequence in $(C(\Theta), \|.\|_\infty)$.

Proof. In the proof we use the notation of Section 5.3 and especially of the proof of Lemma 4. Since the sum of tight sequences is also tight, we can see $S_n(\theta)$ as $S_{n1}(\theta) + S_{n2}(\theta)$, where

$$S_{n1}(\theta) := \frac{1}{t_n}\sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}}{a^2(X_{t_i})\Delta_{n,i}}[\Delta_{n,i}^2(b(X_{t_i},\theta)^2 - b(X_{t_i},\theta_0)^2)+$$
$$+2\Delta_{n,i}\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds(b(X_{t_i},\theta_0) - b(X_{t_i},\theta)) + C_i + D_i + E_i]+$$
$$+\frac{2}{t_n}\sum_{i=0}^{n-1} \frac{1_{\left\{|X_{t_i}| \le \Delta_{n,i}^{-k}\right\}}\mathbb{E}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]}{a^2(X_{t_i})\Delta_{n,i}}(\Delta_{n,i}(b(X_{t_i},\theta) - b(X_{t_i},\theta_0))+$$
$$+(r(\theta, X_{t_i}) - r(\theta_0, X_{t_i}))),$$

$$S_{n2}(\theta) := \frac{2}{t_n}\sum_{i=0}^{n-1}\frac{1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})+$$

$$-\mathbb{E}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]](\Delta_{n,i}(b(X_{t_i},\theta)-b(X_{t_i},\theta_0))+(r(\theta,X_{t_i})-r(\theta_0,X_{t_i}))),$$

and show the tightness of the two sequences individually, using two different criteria. In order to prove that S_{n1} is tight, we want to show that $\sup_n \mathbb{E}[\sup_{\theta\in\Theta}|\frac{\partial}{\partial\theta}S_{n1}(\theta)|] < \infty$. As concerns $S_{n2}(\theta)$, according to Theorem 20 in Appendix 1 from Ibragimov and Has' Minskii [44], we should verify the following: for some positive constant H independent of n, ,

$$\mathbb{E}[(S_{n2}(\theta))^2] \le H \qquad \forall\theta\in\Theta, \tag{1.123}$$

$$\mathbb{E}[(S_{n2}(\theta_1)-S_{n2}(\theta_2))^2] \le H(\theta_1-\theta_2)^2 \qquad \forall\theta_1,\theta_2\in\Theta. \tag{1.124}$$

The derivative that we want to estimate is, using the expressions of C_i, D_i and E_i,

$$\frac{\partial S_{n1}(\theta)}{\partial\theta} = \frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}[2\Delta_{n,i}^2 b(X_{t_i},\theta)\dot{b}(X_{t_i},\theta)+ \tag{1.125}$$

$$+2\Delta_{n,i}\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds(-\dot{b}(X_{t_i},\theta)) - 2\Delta_{n,i}(\dot{b}r)(X_{t_i},\theta) - 2\Delta_{n,i}(b\dot{r})(\theta,X_{t_i})+$$

$$+2(\dot{r}r)(\theta,X_{t_i}) + 2\dot{r}(\theta,X_{t_i})\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds]+$$

$$+\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{\mathbb{E}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}(\dot{r}(\theta,X_{t_i})+\Delta_{n,i}\dot{b}(X_{t_i},\theta)).$$

Using triangle inequality, we can just estimate each term in L^1 norm.
Using the polynomial growth of both b and $\dot{b}$, the fact that φ and the indicator function are bounded, that a^2 is bigger than a constant from Assumption 5 and that $|\Delta_{n,i}| \le \Delta_n$, we get the first term of (1.125) is upper bounded by

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{n}\sum_{i=0}^{n-1}(1+|X_{t_i}|^c)|],$$

that is bounded by the third point of Lemma 26.
On the second term of (1.125) we can use that φ and the indicator function are bounded, that a^2 is bigger than a constant from Assumption 5, that both b and $\dot{b}$ have polynomial growth, from the integral we get a $|\Delta_{n,i}|$ (using (1.114) and (1.116)) that is smaller than Δ_n and so we have just to use the third point of Lemma 26 in order to say that the moments of X are bounded. Hence

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}2\Delta_{n,i}\int_{t_i}^{t_{i+1}} b(X_s,\theta_0)ds(-\dot{b}(X_{t_i},\theta))|] \le c.$$

Concerning the third and the fourth terms of (1.125), we use again that φ and the indicator function are bounded, that a^2 is bigger than a constant from Assumption 5 and that $\dot{b}$ has polynomial growth. We recall that

$$r(\theta,X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} = R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i}) = \Delta_{n,i}^{1+\delta}R(\theta,1,X_{t_i}), \tag{1.126}$$

using (1.111). By the definition (3.135) and the development (1.120) of $\dot{m}_\theta$ we get also the following estimation:

$$\sup_{\theta\in\Theta}|\dot{r}(\theta,x)| \leq \Delta_{n,i}(1+|x|^c). \tag{1.127}$$

We obtain in this way a $|\Delta_{n,i}|$ that is always smaller than Δ_n and so we can simplify the Δ_n in the denominator. Now we use the third point of Lemma 26 and we get also this time that the expectation is bounded.
Also on the fifth we use that φ and the indicator function are bounded, a^2 is bigger than a constant from Assumption 5, (1.126) and (1.127) on $\dot{r}$. Therefore the fifth term of (1.125) is upper bounded by $\Delta_n^\delta \mathbb{E}[|\frac{1}{n}\sum_{i=0}^{n-1}(1+|X_{t_i}|^c)|]$.
Since the exponent on Δ_n is positive and by the third point of Lemma 26, it is upper bounded by a constant.
As concerns the expected value of the sixth term of (1.125), we use again that φ and the indicator function are both bounded, a^2 is bigger than a constant from Assumption 5 and (1.127) on $\dot{r}$. Moreover, we get a $|\Delta_{n,i}|$ from the integral (using (1.114) and (1.116)). The third point of Lemma 26 is sufficient to assure the boundedness of the considered expectation.
Let us now consider

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})}\dot{b}(X_{t_i},\theta)\mathbb{E}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]|].$$

By the boundedness of φ, the Assumption 5 on a and the polynomial growth of $\dot{b}$, it is upper bounded by

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}](1+|X_{t_i}|^c)|] \leq$$

$$\leq \mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i}^{(1+\delta)\wedge\frac{3}{2}})(1+|X_{t_i}|^c)|] \leq c\Delta_n^{\delta\wedge\frac{1}{2}},$$

where we have used (1.31), $|\Delta_{n,i}|\leq\Delta_n$ and the third point of Lemma 26. Since the exponent on Δ_n is positive, it is bounded by a constant.
In order to conclude the proof of the S_{n1}'s tightness, we observe that by the boundedness of both φ and the indicator function, the Assumption 5 on a and (1.127) on $\dot{r}$ we get

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}\dot{r}(\theta,X_{t_i})\mathbb{E}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]|] \leq$$

$$\leq \mathbb{E}[|\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}](1+|X_{t_i}|^c)],$$

on which we can act exactly like above, getting the wanted boundedness.

Let us now consider S_{n2}. In order to prove (2.47), we observe that

$$\mathbb{E}[(S_{n2}(\theta_1)-S_{n2}(\theta_2))^2] \leq \frac{c}{n^2\Delta_n^2}\mathbb{E}[(\sum_{i=0}^{n-1}\frac{1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}}{a^2(X_{t_i})\Delta_{n,i}}[\zeta_i\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})+ \tag{1.128}$$

$$-\mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]](\Delta_{n,i}(b(X_{t_i}, \theta_2) - b(X_{t_i}, \theta_1)) + r(\theta_1, X_{t_i}) - r(\theta_2, X_{t_i})))^2]$$

By the building the sum is a square integrable martingale. The Pythagoras' theorem on a square integrable martingale yields that (1.128) is equal to

$$\frac{c}{n^2 \Delta_n^2} \sum_{i=0}^{n-1} \mathbb{E}[\frac{1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{a^4(X_{t_i})\Delta_{n,i}^2}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) + \tag{1.129}$$

$$-\mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]]^2(\Delta_{n,i}(b(X_{t_i}, \theta_2) - b(X_{t_i}, \theta_1)) + r(\theta_1, X_{t_i}) - r(\theta_2, X_{t_i}))^2].$$

We now observe that

$$(\Delta_{n,i}(b(X_{t_i}, \theta_2) - b(X_{t_i}, \theta_1)) + r(\theta_1, X_{t_i}) - r(\theta_2, X_{t_i}))^2 \le c\Delta_{n,i}^2(b(X_{t_i}, \theta_2) - b(X_{t_i}, \theta_1))^2 +$$

$$+c(r(\theta_1, X_{t_i}) - r(\theta_2, X_{t_i}))^2 \le c\Delta_{n,i}^2 \dot{b}(X_{t_i}, \theta_u)^2(\theta_1 - \theta_2)^2 + c\dot{r}(\theta_u, X_{t_i})^2(\theta_1 - \theta_2)^2,$$

where $\theta_u \in [\theta_1, \theta_2]$. Using (1.127), it is upper bounded by

$$c\Delta_{n,i}^2[\dot{b}(X_{t_i}, \theta_u)^2 + (1 + |X_{t_i}|^c)^2](\theta_1 - \theta_2)^2. \tag{1.130}$$

Replacing (1.130) in (1.129), using that the indicator function is bounded by a constant, the Assumption 5 on a and that $\dot{b}$ has polynomial growth, we get that (1.129) is upper bounded by

$$\frac{c}{n^2 \Delta_n^2} \sum_{i=0}^{n-1} \mathbb{E}[(\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - \mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}])^2(1 + |X_{t_i}|^c)^2](\theta_1 - \theta_2)^2 =$$

$$= \frac{c}{n^2 \Delta_n^2} \sum_{i=0}^{n-1} \mathbb{E}[\mathbb{E}[(\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - \mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}])^2|\mathcal{F}_{t_i}] \times \tag{1.131}$$

$$\times (1 + |X_{t_i}|^c)^2](\theta_1 - \theta_2)^2,$$

by the definition of conditional expected value and the measurability of X_{t_i}.
We observe that $\mathbb{E}[(\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - \mathbb{E}[\zeta_i \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}])^2|\mathcal{F}_{t_i}]$ is the conditional variance of $\zeta_i \varphi$ and so it is always smaller then $\mathbb{E}[\zeta_i^2 \varphi_{\Delta_{n,i}^\beta}^2(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]$ that is, using (1.32), $R(\theta, \Delta_{n,i}, X_{t_i})$. We get that (1.131) is upper bounded by

$$\frac{1}{n^2 \Delta_n^2} \sum_{i=0}^{n-1} \mathbb{E}[R(\theta, \Delta_{n,i}, X_{t_i})(1 + |X_{t_i}|^c)^2](\theta_1 - \theta_2)^2 \le \frac{1}{n\Delta_n} c(\theta_1 - \theta_2)^2,$$

where in the last inequality we have used (3.24) in order to say that $R(\theta, \Delta_{n,i}, X_{t_i}) = \Delta_{n,i} R(\theta, 1, X_{t_i})$, the fact that $|\Delta_{n,i}| \le \Delta_n$, the natural polynomial growth of the function derived from its definition (3.23) and the third point of Lemma 26 in order to assure the boundedness of the expected value.
Hence, recalling that $n\Delta_n \to \infty$, we get (2.47) since $\frac{1}{n\Delta_n} c(\theta_1 - \theta_2)^2 \le c(\theta_1 - \theta_2)^2$.

Concerning (2.46), we act exactly like we have already done in order to prove (2.47), getting $\mathbb{E}[(S_{n2}(\theta))^2] \le c(\theta - \theta_0)^2$. Θ is a compact set and so Θ's diameter $d := \sup_{\theta_1, \theta_2 \in \Theta} |\theta_1 - \theta_2|$ is $< \infty$. We therefore deduce (2.46): $c(\theta - \theta_0)^2 \le cd^2 \le c$.
The tightness of $S_n(\theta) = \frac{U_n(\theta) - U_n(\theta_0)}{t_n}$ follows. □

We are now ready to show the consistence of the estimator $\hat{\theta}_n := arg\min_{\theta\in\Theta} U_n(\theta)$. We want to prove that $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$ when $n \to \infty$, that is equivalent to show that $\forall \left\{\hat{\theta}_{n_k}\right\} \subset \hat{\theta}_n$, $\exists \left\{\hat{\theta}_{n_{k_j}}\right\} \subset \left\{\hat{\theta}_{n_k}\right\}$ such that $\hat{\theta}_{n_{k_j}} \to \theta_0$ a.s.
Let $\left\{\hat{\theta}_{n_k}\right\}$ be a subsequence of $\left\{\hat{\theta}_n\right\}$. By the uniform convergence in probability of the contrast function given by Lemma 4 and Lemma 5, we get the a.s. convergence along some subsequence of n_k, denoted n_{k_j}:

$$\sup_{\theta\in\Theta} |\frac{U_{n_{k_j}}(\theta) - U_{n_{k_j}}(\theta_0)}{t_{n_{k_j}}} - l(\theta,\theta_0)| \xrightarrow{a.s.} 0, \quad n_{k_j} \to \infty,$$

where $l(\theta,\theta_0) = \int_{\mathbb{R}} \frac{(b(x,\theta)-b(x,\theta_0))^2}{a^2(x)}\pi(dx) \geq 0$.
Now, for fixed $\omega \in \Omega$, thanks to the compactness of Θ, there exists a subsequence of n_{k_j}, that we still denote n_{k_j}, and a θ_∞ such that $\hat{\theta}_{n_{k_j}} \to \theta_\infty$.
Since the mapping $\theta \mapsto l(\theta,\theta_0)$ is continuous, we have $l(\hat{\theta}_{n_{k_j}},\theta_0) \to l(\theta_\infty,\theta_0)$.
Then, by the definition of $\hat{\theta}_n$ as the argmin of $U_n(\theta)$, we have

$$0 \geq \frac{U_{n_{k_j}}(\hat{\theta}_{n_{k_j}}) - U_{n_{k_j}}(\theta_0)}{t_{n_{k_j}}} \to l(\theta_\infty,\theta_0) \geq 0$$

and so $l(\theta_\infty,\theta_0) = 0$. The Assumption 6 of identifiability leads that $\theta_\infty = \theta_0$.
This implies that any convergent subsequence of $\hat{\theta}_n$ tends to θ_0; this means the consistency of $\hat{\theta}_n$.

1.7.8 Contrast's derivatives convergence

We are now ready to show the convergence of the derivative of the contrast function through the following lemma:

Lemma 6. *Suppose that Assumptions 1 - 8 and A_β are satisfied. Then*

$$\frac{\dot{U}_n(\theta_0)}{\sqrt{t_n}} \xrightarrow{\mathcal{L}} N(0, 4\int_{\mathbb{R}} (\frac{\dot{b}(x,\theta_0)}{a(x)})^2\, \pi(dx)).$$

Proof. We recall that

$$U_n(\theta_0) = \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2}{a^2(X_{t_i})\Delta_{n,i}}\, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}},$$

hence

$$\dot{U}_n(\theta_0) = 2\sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\dot{m}_{\theta_0}(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}}\, \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}}. \quad (1.132)$$

It means that

$$\frac{\dot{U}_n(\theta_0)}{\sqrt{t_n}} = \frac{2}{\sqrt{t_n}} \sum_{i=0}^{n-1} (X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) \frac{\dot{b}(X_{t_i},\theta_0)}{a^2(X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} +$$

$$+\frac{2}{\sqrt{t_n}} \sum_{i=0}^{n-1} (X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) \frac{R(\theta_0, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\alpha\beta-\epsilon-\beta)}, X_{t_i})}{a^2(X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}}, \quad (1.133)$$

where we have used the development (1.120) of $\dot{m}_\theta(X_{t_i})$.
We now use Proposition 5 on the first term of (1.133), getting that it converges in distribution to a Gaussian random variable with mean 0 and variance $\int_{\mathbb{R}} \frac{4\dot{b}^2(x,\theta_0)}{a^4(x)} a^2(x)\pi(dx) = \int_{\mathbb{R}} 4(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx)$, as we wanted. In order to get the thesis we want to show that the second term of (1.133) goes to zero in probability as $t_n \to \infty$. In order to do this, we we want to use Lemma 9 of [36] and so we have to prove the following:

$$\frac{2}{\sqrt{t_n}} \sum_{i=0}^{n-1} \mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) \frac{R(\theta_0, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\alpha\beta-\epsilon-\beta)}, X_{t_i})}{a^2(X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} |\mathcal{F}_{t_i}] \to 0 \tag{1.134}$$

$$\frac{4}{t_n} \sum_{i=0}^{n-1} \mathbb{E}[((X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) \frac{R(\theta_0, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\alpha\beta-\epsilon-\beta)}, X_{t_i})}{a^2(X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}})^2 |\mathcal{F}_{t_i}] \to 0 \tag{1.135}$$

Using the measurability and the fact that

$$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}] = 0 \tag{1.136}$$

we get (1.134). Let us consider (1.135). Using the Assumption 5 on a, the measurability of R and the expression (2.26) we can upper bound it with

$$\frac{c}{n\Delta_n} \sum_{i=0}^{n-1} R(\theta_0, \Delta_{n,i}^{1\wedge 2(1-\alpha\beta-\epsilon-\beta)}, X_{t_i}) R(\theta_0, \Delta_{n,i}, X_{t_i}) \le \Delta_n^{1\wedge 2(1-\alpha\beta-\epsilon-\beta)} \frac{1}{n} \sum_{i=0}^{n-1} R(\theta_0, 1, X_{t_i}),$$

that goes to zero in norm 1 by the polynomial growth of R, the third point of Lemma 26 and A_β. Therefore it converges to zero also in probability.
It follows that

$$\frac{2}{\sqrt{t_n}} \sum_{i=0}^{n-1} (X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) \frac{R(\theta_0, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\alpha\beta-\epsilon-\beta)}, X_{t_i})}{a^2(X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} 0,$$

as we wanted. □

Concerning the second derivative of the contrast function, we have the following convergence:

Lemma 7. *Suppose that Assumptions 1 - 8 and A_β hold. Then*

$$\frac{\ddot{U}_n(\theta_0)}{t_n} \xrightarrow{\mathbb{P}} -2 \int_{\mathbb{R}} (\frac{\dot{b}(x,\theta_0)}{a(x)})^2 \, \pi(dx).$$

Proof. Derivating twice the expression of U_n we get

$$\ddot{U}_n(\theta_0) = -2 \sum_{i=0}^{n-1} \frac{\dot{m}_{\theta_0}^2(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} +$$
$$+2 \sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\ddot{m}_{\theta_0}(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}} \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} \tag{1.137}$$

First of all we show that the second term of (1.137), divided by $n\Delta_n$, goes to zero in probability. We use again Lemma 9 of [36]. Hence, our goal is to prove the following:

$$\frac{2}{t_n}\sum_{i=0}^{n-1}\mathbb{E}[\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\ddot{m}_{\theta_0}(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}|\mathcal{F}_{t_i}] \to 0 \tag{1.138}$$

$$\frac{4}{(t_n)^2}\sum_{i=0}^{n-1}\mathbb{E}[(\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\ddot{m}_{\theta_0}(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}})^2|\mathcal{F}_{t_i}] \to 0 \tag{1.139}$$

As we acted in the last proof, we use (1.136) in order to get (1.138).
Concerning (1.139), using Assumption 5 on a, the measurability of R, the development (1.121) of $\ddot{m}_{\theta_0}(X_{t_i})$ and the expression (2.26) we can upper bound it with

$$\frac{c}{n^2\Delta_n^2}\sum_{i=0}^{n-1}[R(\theta_0,\Delta_{n,i},X_{t_i})\frac{\Delta_{n,i}^2\ddot{b}^2(X_{t_i},\theta_0) + R(\theta_0,\Delta_{n,i}^{3\wedge 2(2-\alpha\beta-\epsilon-\beta)},X_{t_i})}{\Delta_{n,i}^2}] \le$$

$$\le \frac{c}{n^2\Delta_n}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}),$$

where in the last inequality we have used the polynomial growth of $\ddot{b}$, the property (3.24) on R and that $|\Delta_{n,i}| \le \Delta_n$. Since $n\Delta_n \to \infty$ and $\frac{1}{n}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})$ is bounded in L^1, we get the convergence en probability wanted.
Let us now consider the first term of (1.137). Using the development (1.120) we get

$$\frac{-2}{t_n}\sum_{i=0}^{n-1}\frac{(\Delta_{n,i}\dot{b}(X_{t_i},\theta_0) + R(\theta_0,\Delta_{n,i}^{\frac{3}{2}\wedge(2-\beta-\beta\alpha-\epsilon)},X_{t_i}))^2}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}. \tag{1.140}$$

Hence, we obtain three terms by expanding the square. Using on the first Proposition 3, we get the convergence

$$\frac{-2}{t_n}\sum_{i=0}^{n-1}\frac{\Delta_{n,i}\dot{b}^2(X_{t_i},\theta_0)}{a^2(X_{t_i})}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} -2\int_{\mathbb{R}}\frac{\dot{b}^2(x,\theta_0)}{a^2(x)}\pi(dx). \tag{1.141}$$

The second term of (1.140) is

$$\frac{-4}{t_n}\sum_{i=0}^{n-1}\frac{2\dot{b}(X_{t_i},\theta_0)R(\theta_0,\Delta_{n,i}^{\frac{3}{2}\wedge(2-\beta-\beta\alpha-\epsilon)},X_{t_i})}{a^2(X_{t_i})}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}.$$

Using Assumption 5 on a, the fact that both φ and the indicator function are bounded, the polynomial growth of both $\dot{b}$ and R and the third point of Lemma 26 we get that its L^1 norm is upper bounded by $c\Delta_n^{\frac{1}{2}\wedge(1-\beta-\beta\alpha-\epsilon)}$. Since the exponent on Δ_n is positive, the convergence in norm L^1 and therefore in probability follows.
Concerning the last term of (1.140), using again Assumption 5 on a, the fact that both φ and the indicator function are bounded, the polynomial growth of R and the third point of Lemma 26 we get that its L^1 norm is upper bounded by $c\Delta_n^{1\wedge(2-2\beta-2\beta\alpha-2\epsilon)}$. Once again, since the exponent on Δ_n is positive, the convergence in norm L^1 and therefore in probability follows.
It yields

$$\frac{\ddot{U}_n(\theta_0)}{t_n} \xrightarrow{\mathbb{P}} -2\int_{\mathbb{R}}\frac{\dot{b}^2(x,\theta_0)}{a^2(x)}\pi(dx).$$

□

1.7.9 Asymptotic normality of the estimator

In order to show the asymptotic normality of the estimator we need the following lemma:

Lemma 8. *Suppose that Assumptions 1 - 8 and A_β hold. Then*

$$\frac{1}{t_n}\sup_{t\in[0,1]}|\ddot{U}_n(\theta_0+t(\hat{\theta}_n-\theta_0))-\ddot{U}_n(\theta_0)|\xrightarrow{\mathbb{P}}0, \tag{1.142}$$

where $\hat{\theta}_n$ is the estimator defined in (2.7).

Proof. Let us define

$$\tilde{\theta}_n:=\theta_0+t(\hat{\theta}_n-\theta_0). \tag{1.143}$$

Using (1.137),

$$\begin{aligned}\frac{\ddot{U}_n(\tilde{\theta}_n)-\ddot{U}_n(\theta_0)}{t_n} &= -\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{(\dot{m}^2_{\tilde{\theta}_n}(X_{t_i})-\dot{m}^2_{\theta_0}(X_{t_i}))}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}+\\ &+\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}}-m_{\theta_0}(X_{t_i}))(\ddot{m}_{\tilde{\theta}_n}(X_{t_i})-\ddot{m}_{\theta_0}(X_{t_i}))}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}+\\ &+\frac{2}{t_n}\sum_{i=0}^{n-1}\frac{(m_{\theta_0}(X_{t_i})-m_{\tilde{\theta}_n}(X_{t_i}))\ddot{m}_{\tilde{\theta}_n}(X_{t_i})}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}.\end{aligned} \tag{1.144}$$

Concerning the first term of (1.144), we use the following estimation:

$$|\dot{m}^2_{\tilde{\theta}_n}(X_{t_i})-\dot{m}^2_{\theta_0}(X_{t_i})|\le 2|\ddot{m}_{\theta_u}(X_{t_i})\dot{m}_{\theta_u}(X_{t_i})(\tilde{\theta}_n-\theta_0)|, \tag{1.145}$$

where $\theta_u\in[\theta_0,\tilde{\theta}_n]$. We replace the development (1.120) and (1.121) of $\dot{m}$ and $\ddot{m}$. Hence the first term of (1.144) is, in module, upper bounded by

$$\begin{aligned}&\frac{2}{n}\sum_{i=0}^{n-1}|2(\dot{b}(X_{t_i},\theta_u)+R(\theta_u,\Delta_{n,i}^{\frac{1}{2}\wedge(1-\beta-\beta\alpha-\epsilon)},X_{t_i}))(\ddot{b}(X_{t_i},\theta_u)+\\ &+R(\theta_u,\Delta_{n,i}^{\frac{1}{2}\wedge(1-\beta-\beta\alpha-\epsilon)},X_{t_i}))||\tilde{\theta}_n-\theta_0|=\\ &=\frac{1}{n}\sum_{i=0}^{n-1}|R(\theta_u,1,X_{t_i})||\tilde{\theta}_n-\theta_0|\le\frac{1}{n}\sum_{i=0}^{n-1}c(1+|X_{t_i}|^c)|\hat{\theta}_n-\theta_0|,\end{aligned} \tag{1.146}$$

where we have used Assumption 5 on a, the boundedness of both φ and the indicator function, the property (3.24) on R that $|\Delta_{n,i}|\le\Delta_n$ and the definition (1.143) of $\tilde{\theta}_n$ joint with the fact that $|t|\le 1$. By the consistency of $\hat{\theta}_n$ that we have already proved, we get that the first term of (1.144) converges to zero in probability uniformly in t, since the right hand side of (1.146) is bounded in L^1 by the third point of Lemma 26 and it does not depend on t.

On the third term of (1.144) we use again the Assumption 5 on a, the fact that both φ and the indicator function are bounded, the development (1.121) of $\ddot{m}_\theta$ and the following estimation: $|m_{\theta_0}(X_{t_i}) - m_{\tilde{\theta}_n}(X_{t_i})| \le |\dot{m}_{\theta_u}(X_{t_i})||\theta_0 - \tilde{\theta}_n|$, on which we can use the development (1.120) of $\dot{m}_\theta$. We can hence upper bound the third term with

$$\frac{2}{n}\sum_{i=0}^{n-1} 2|(\dot{b}(X_{t_i},\theta_u) + R(\theta_u, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\beta-\beta\alpha-\epsilon)}, X_{t_i}))(\ddot{b}(X_{t_i},\tilde{\theta}_n)+$$

$$+R(\tilde{\theta}_n, \Delta_{n,i}^{\frac{1}{2}\wedge(1-\beta-\beta\alpha-\epsilon)}, X_{t_i}))||\theta_0 - \tilde{\theta}_n| = \tag{1.147}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}|R(\theta,1,X_{t_i})||\tilde{\theta}_n - \theta_0| \le \frac{1}{n}\sum_{i=0}^{n-1} c(1+|X_{t_i}|^c)|\hat{\theta}_n - \theta_0|.$$

The consistency of $\hat{\theta}_n$ yields the convergence in probability uniformly in t wanted, by the boundedness in L^1 of the sum, that does not depend on t.
It remains to prove the convergence to zero, uniformly in t, for the second term of (1.144); it is sufficient to prove that the following sequence $S_n(\theta)$ converges to zero uniformly with respect to θ:

$$S_n(\theta) := \frac{2}{t_n}\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))(\ddot{m}_\theta(X_{t_i}) - \ddot{m}_{\theta_0}(X_{t_i}))}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}.$$

The pointwise convergence is already proved (it is enough to repeat the proof of (1.138) and (1.139) with $\ddot{m}_\theta(X_{t_i}) - \ddot{m}_{\theta_0}(X_{t_i})$ in place of $\ddot{m}_{\theta_0}(X_{t_i})$). In order to show that the convergence takes place uniformly in θ, we prove the tightness of $S_n(\theta)$, using the criterion analogues to (2.46) and (2.47).
Let us consider (2.47) first. We observe that

$$\mathbb{E}[(S_n(\theta_1) - S_n(\theta_2))^2] \le$$

$$\le \frac{c}{n^2\Delta_n^2}\mathbb{E}[(\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))(\ddot{m}_{\theta_1}(X_{t_i}) - \ddot{m}_{\theta_2}(X_{t_i}))}{a^2(X_{t_i})\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}})^2]. \tag{1.148}$$

By the building the sum is a square integrable martingale. The Pythagoras' theorem on a square integrable martingale yields that (1.148) is equal to

$$\frac{c}{n^2\Delta_n^2}\mathbb{E}[\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2(\ddot{m}_{\theta_1}(X_{t_i}) - \ddot{m}_{\theta_2}(X_{t_i}))^2}{a^4(X_{t_i})\Delta_{n,i}^2}\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}]. \tag{1.149}$$

We now use the following estimation:

$$|\ddot{m}_{\theta_1}(X_{t_i}) - \ddot{m}_{\theta_2}(X_{t_i})| \le |\dddot{m}_{\theta_u}(X_{t_i})||\theta_1 - \theta_2|. \tag{1.150}$$

Replacing (1.150) in (1.148) and using (1.122) on $\dddot{m}_{\theta_u}(X_{t_i})$, we can upper bound (1.148) with

$$\frac{c}{n^2\Delta_n^2}\mathbb{E}[\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2 R(\theta_u, \Delta_{n,i}^2, X_{t_i})}{a^4(X_{t_i})\Delta_{n,i}^2}\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}](\theta_1 - \theta_2)^2 \le$$

$$\le \frac{c}{n^2\Delta_n^2}\mathbb{E}[\sum_{i=0}^{n-1} f(X_{t_i},\theta_u)\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}|\mathcal{F}_{t_i}]](\theta_1 - \theta_2)^2, \tag{1.151}$$

with $f(X_{t_i}, \theta_u) = \frac{R(\theta_u, 1, X_{t_i})}{a^4(X_{t_i})}$ and where we have used the property (3.24) of the functions R and the definition of conditional expected value.
Using (2.26), the property (3.24) and that $|\Delta_{n,i}| \le \Delta_n$, we can upper bound (1.151) with
$\frac{4}{n^2\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[f(X_{t_i}, \theta_u)R(\theta_0, 1, X_{t_i})](\theta_1 - \theta_2)^2$.
By the Assumption 5 on a and the polynomial growth of R derived by its definition, f has polynomial growth. Using the third point of Lemma 26 we get that the expected value is bounded. Hence, since $n\Delta_n \to \infty$, it yields

$$\frac{4}{n^2\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[f(X_{t_i}, \theta_u)R(\theta_0, 1, X_{t_i})] \le c, \tag{1.152}$$

therefore we obtain (2.47) on S_n.
Concerning (2.46), we can act exactly in the same way, using (1.152) and the compactness of Θ. The tightness of $S_n(\theta)$ follows. □

We are now ready to prove the asymptotic normality of the estimator. Using (1.142) we have that

$$\frac{1}{t_n}\int_0^1 [\ddot{U}_n(\theta_0 + t(\hat{\theta}_n - \theta_0)) - \ddot{U}_n(\theta_0)]dt \xrightarrow{\mathbb{P}} 0. \tag{1.153}$$

We observe that

$$\frac{1}{t_n}\int_0^1 [\ddot{U}_n(\theta_0 + t(\hat{\theta}_n - \theta_0))]dt\sqrt{t_n}(\hat{\theta}_n - \theta_0) =$$
$$= \frac{1}{\sqrt{t_n}}\int_0^1 [\ddot{U}_n(\theta_0 + t(\hat{\theta}_n - \theta_0))]dt\,(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{t_n}}(\dot{U}_n(\hat{\theta}_n) - \dot{U}_n(\theta_0)) = -\frac{\dot{U}_n(\theta_0)}{\sqrt{t_n}}, \tag{1.154}$$

where in the last equality we have used that, on the set $\left\{\hat{\theta}_n \in \mathring{\Theta}\right\}$, $\dot{U}_n(\hat{\theta}_n) = 0$ since $\hat{\theta}_n$ is a minimum.
Hence

$$\sqrt{t_n}(\hat{\theta}_n - \theta_0) = \frac{-\frac{\dot{U}_n(\theta_0)}{\sqrt{t_n}}}{\frac{1}{t_n}\int_0^1[\ddot{U}_n(\theta_0 + t(\hat{\theta}_n - \theta_0))]dt}. \tag{1.155}$$

Using Lemma 6 we have the convergence in distribution of the numerator of (1.155) to $N(0, 4\int_{\mathbb{R}}(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx))$ and, by the equation (1.153), the denominator converges in probability to $-2\int_{\mathbb{R}}(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx)$.
Therefore$\sqrt{t_n}(\hat{\theta}_n - \theta_0)$ converges in distribution to $N(0, \frac{4\int_{\mathbb{R}}(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx)}{4(\int_{\mathbb{R}}(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx))^2})$, i. e. it is $N(0, (\int_{\mathbb{R}}(\frac{\dot{b}(x,\theta_0)}{a(x)})^2\pi(dx))^{-1})$, as we wanted.

1.7.10 Proof of Proposition 1

The proof of the proposition is essentially similar to the the proof of the asymptotic normality of $\hat{\theta}_n$ given in Sections 1.7.6–1.7.9 and we skip it. The main difference comes from the fact that Proposition 5 holds true with $\widetilde{m}_{\theta_0}(X_{t_i})$ replacing $m_{\theta_0}(X_{t_i})$ under the condition that $\sqrt{n}\Delta_n^{\rho-1/2} \to 0$.

1.8 Appendix

In this section we will prove the technical lemmas that we have used in order to show the main theorems.

1.8.1 Proof of expansions of the derivatives of the function $m_{\theta,h}$

In order to prove the explicit approximation of $\dot{m}_{\theta,h}$ and $\ddot{m}_{\theta,h}$ provided in Proposition 8, the following lemma will be useful. We point out that X_t^θ is $X_t^{\theta,x}$ and so the process starts in 0: $X_0^{\theta,x} = x$.

Lemma 9. *Suppose that Assumptions 1 to 4 and 7 hold. Let us define* $\dot{X}_t^{\theta,x} := \frac{\partial X_t^{\theta,x}}{\partial \theta}$ *and* $\ddot{X}_t^{\theta,x} := \frac{\partial^2 X_t^{\theta,x}}{\partial \theta^2}$. *Then, for all* $p \ge 2$ $\exists c > 0$: $\forall h \le \Delta_n$ $\forall x$,

$$\mathbb{E}[|\frac{\dot{X}_h^{\theta,x}}{h}|^p] \le c(1 + |x|^c), \tag{1.156}$$

$$\mathbb{E}[|\frac{\ddot{X}_h^{\theta,x}}{h}|^p] \le c(1 + |x|^c). \tag{1.157}$$

Proof. The dynamic of the process X is known. The same applies to the processes $\dot{X}_t^{\theta,x}$ and $\ddot{X}_t^{\theta,x}$ (cf. (missing citation), section 5).

$$\dot{X}_h^{\theta,x} = \int_0^h (b'(X_s^{\theta,x},\theta)\dot{X}_s^{\theta,x} + \dot{b}(X_s^{\theta,x},\theta))ds + \tag{1.158}$$

$$+ \int_0^h a'(X_s^{\theta,x})\dot{X}_s^{\theta,x} dW_s + \int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^{\theta,x})\dot{X}_s^{\theta,x} z \tilde{\mu}(dz,ds)$$

and

$$\ddot{X}_h^{\theta,x} = \int_0^h (b''(X_s^{\theta,x},\theta)(\dot{X}_s^{\theta,x})^2 + 2\dot{b}'(X_s^{\theta,x},\theta)\dot{X}_s^{\theta,x} + b'(X_s^{\theta,x},\theta)\ddot{X}_s^{\theta,x} + \ddot{b}(X_s^{\theta,x},\theta))ds + \tag{1.159}$$

$$+ \int_0^h (a''(X_s^{\theta,x})(\dot{X}_s^{\theta,x})^2 + a'(X_s^{\theta,x})\ddot{X}_s^{\theta,x})dW_s +$$

$$+ \int_0^h \int_{\mathbb{R}} (\gamma''(X_{s^-}^{\theta,x})(\dot{X}_s^{\theta,x})^2 + \gamma'(X_{s^-}^{\theta,x})\ddot{X}_s^{\theta,x}) z \tilde{\mu}(dz,ds).$$

From now on, we will drop the dependence of the starting point in order to make the notation easier.
Let us start with the proof of (1.156). We observe that, taking the L^p norm of (1.158), we have the following estimation:

$$\mathbb{E}[|\dot{X}_h^\theta|^p] \le c\mathbb{E}[|\int_0^h (b'(X_s^\theta,\theta)\dot{X}_s^\theta + \dot{b}(X_s^\theta,\theta))ds|^p] + c\mathbb{E}[|\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s|^p] + \tag{1.160}$$

$$+ c\mathbb{E}[|\int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z \tilde{\mu}(dz,ds)|^p].$$

Concerning the first term of (1.160),

$$\mathbb{E}[|\int_0^h (b'(X_s^\theta,\theta)\dot{X}_s^\theta + \dot{b}(X_s^\theta,\theta))ds|^p] \le c\mathbb{E}[|\int_0^h b'(X_s^\theta,\theta)\dot{X}_s^\theta ds|^p] + c\mathbb{E}[|\int_0^h \dot{b}(X_s^\theta,\theta)ds|^p].$$

Then, using Jensen inequality on the first, we obtain

$$\mathbb{E}[|\int_0^h b'(X_s^\theta,\theta)\dot{X}_s^\theta ds|^p] = \mathbb{E}[h^p|\frac{1}{h}\int_0^h b'(X_s^\theta,\theta)\dot{X}_s^\theta ds|^p] \le$$

$$\le \mathbb{E}[h^{p-1}\int_0^h |b'(X_s^\theta,\theta)|^p|\dot{X}_s^\theta|^p ds] = h^{p-1}\int_0^h \mathbb{E}[|b'(X_s^\theta,\theta)|^p|\dot{X}_s^\theta|^p]ds.$$

The derivatives of b with respect to x are supposed bounded, it yields

$$\mathbb{E}[|\int_0^h b'(X_s^\theta,\theta)\dot{X}_s^\theta ds|^p] \le ch^{p-1}\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds. \tag{1.161}$$

Let us now consider the second term of (1.160). Using Burkholder-Davis-Gundy and Jensen inequalities we get

$$\mathbb{E}[|\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s|^p] \le c\mathbb{E}[|\int_0^h (a'(X_s^\theta)\dot{X}_s^\theta)^2 ds|^{\frac{p}{2}}] =$$

$$= c\mathbb{E}[h^{\frac{p}{2}}|\frac{1}{h}\int_0^h (a'(X_s^\theta)\dot{X}_s^\theta)^2 ds|^{\frac{p}{2}}] \le ch^{\frac{p}{2}-1}\mathbb{E}[\int_0^h |a'(X_s^\theta)\dot{X}_s^\theta|^p].$$

Therefore

$$\mathbb{E}[|\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s|^p] \le ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds, \tag{1.162}$$

where we have used that the derivatives of a are bounded.
The third term of (1.160) can be estimated using Kunita inequality (cf. the Appendix of [46]):

$$\mathbb{E}[|\int_0^h\int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z\tilde{\mu}(dz,ds)|^p] \le$$

$$\le \mathbb{E}[\int_0^h\int_{\mathbb{R}} |\gamma'(X_s^\theta)\dot{X}_s^\theta|^p|z|^p\bar{\mu}(dz,ds)] + \mathbb{E}[|\int_0^h\int_{\mathbb{R}} (\gamma'(X_s^\theta)\dot{X}_s^\theta)^2 z^2\bar{\mu}(dz,ds)|^{\frac{p}{2}}] \le$$

$$\le \int_0^h \mathbb{E}[|\gamma'(X_s^\theta)|^p|\dot{X}_s^\theta|^p](\int_{\mathbb{R}}|z|^pF(z)dz)ds + \mathbb{E}[|\int_0^h (\gamma'(X_s^\theta)\dot{X}_s^\theta)^2(\int_{\mathbb{R}} z^2F(z)dz)ds|^{\frac{p}{2}}] \le$$

$$\le c\int_0^h \mathbb{E}[|\gamma'(X_s^\theta)|^p|\dot{X}_s^\theta|^p]ds + c\mathbb{E}[|\int_0^h (\gamma'(X_s^\theta)\dot{X}_s^\theta)^2 ds|^{\frac{p}{2}}],$$

where in the last two inequalities we have just used the definition of the compensated measure $\bar{\mu}$ and the third point of Assumption 4.
Since the derivatives of γ are supposed bounded and by the Jensen inequality we get it is upper bounded by

$$c\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds + c\mathbb{E}[h^{\frac{p}{2}-1}\int_0^h |\gamma'(X_s^\theta)|^p|\dot{X}_s^\theta|^p ds] \le$$

$$\le c\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds + ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds.$$

Hence

$$\mathbb{E}[|\int_0^h\int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z\tilde{\mu}(dz,ds)|^p] \le c(1+h^{\frac{p}{2}-1})\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds. \tag{1.163}$$

From (1.161), (1.162) and (2.97), we obtain

$$\mathbb{E}[|\dot{X}_h^\theta|^p] \leq c\mathbb{E}[|\int_0^h \dot{b}(X_s^\theta, \theta)ds|^p] + c(1 + h^{\frac{p}{2}-1} + h^{p-1})\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds.$$

Let M_h be $\mathbb{E}[|\dot{X}_h^\theta|^p]$, then the equation above can be seen as

$$M_h \leq c\mathbb{E}[|\int_0^h \dot{b}(X_s^\theta, \theta)ds|^p] + c(1 + h^{\frac{p}{2}-1} + h^{p-1})\int_0^h M_s ds.$$

Using Gronwall lemma, it yields $M_h \leq c\mathbb{E}[|\int_0^h \dot{b}(X_s^\theta, \theta)ds|^p]e^{ch(1+h^{\frac{p}{2}-1}+h^{p-1})}$.
By the polynomial growth of $\dot{b}$ and the third point of Lemma 1,

$$\mathbb{E}[|\int_0^h \dot{b}(X_s^\theta, \theta)ds|^p] \leq ch^p(1 + |X_0^{0,x}|^c) = ch^p(1 + |x|^c).$$

Hence $\mathbb{E}[|\dot{X}_h^\theta|^p] \leq ch^p(1 + |x|^c)$.
Our goal is now to prove (1.157). In order to do it, we take the L^p norm of (1.159), getting the following estimation:

$$\mathbb{E}[|\ddot{X}_h^\theta|^p] \leq \mathbb{E}[|\int_0^h (b''(X_s^\theta, \theta)(\dot{X}_s^\theta)^2 + 2\dot{b}'(X_s^\theta, \theta)\dot{X}_s^\theta + b'(X_s^\theta, \theta)\ddot{X}_s^\theta +$$

$$+\ddot{b}(X_s^\theta, \theta))ds|^p] + \mathbb{E}[|\int_0^h (a''(X_s^\theta)(\dot{X}_s^\theta)^2 + a'(X_s^\theta)\ddot{X}_s^\theta)dW_s|^p] + \tag{1.164}$$

$$+\mathbb{E}[|\int_0^h \int_{\mathbb{R}} (\gamma''(X_{s-}^\theta)(\dot{X}_s^\theta)^2 + \gamma'(X_{s-}^\theta)\ddot{X}_s^\theta)z\tilde{\mu}(dz, ds)|^p]$$

The first term of (1.164) is upper bounded by

$$\mathbb{E}[|\int_0^h (b''(X_s^\theta, \theta)(\dot{X}_s^\theta)^2 ds|^p] + \mathbb{E}[|\int_0^h 2\dot{b}'(X_s^\theta, \theta)\dot{X}_s^\theta ds|^p] + \mathbb{E}[|\int_0^h b'(X_s^\theta, \theta)\ddot{X}_s^\theta ds|^p] +$$

$$+\mathbb{E}[|\int_0^h \ddot{b}(X_s^\theta, \theta)ds|^p] \leq ch^{p-1}\int_0^h \mathbb{E}[|\dot{X}_s^\theta|^{2p}]ds + ch^{p-1}\int_0^h \mathbb{E}[|\dot{b}'(X_s^\theta, \theta)|^p|\dot{X}_s^\theta|^p]ds +$$

$$+ch^{p-1}\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds + \mathbb{E}[|\int_0^h \ddot{b}(X_s^\theta, \theta)ds|^p], \tag{1.165}$$

where we have used Jensen inequality and that the derivatives of b with respect to x are supposed bounded.
By Holder inequality

$$\mathbb{E}[|\dot{b}'(X_s^\theta, \theta)|^p|\dot{X}_s^\theta|^p] \leq (\mathbb{E}[|\dot{b}'(X_s^\theta, \theta)|^{pp_1}])^{\frac{1}{p_1}}(\mathbb{E}[|\dot{X}_s^\theta|^{pp_2}]^{\frac{1}{p_2}} \leq$$

$$\leq c(h^{pp_2})^{\frac{1}{p_2}}(1 + |x|^c) = ch^p(1 + |x|^c),$$

where in the last inequality we have used the boundedness of $\dot{b}'$ and (1.156). Since $\ddot{b}$ has polynomial growth and by the third point of Lemma 25, $\mathbb{E}[|\int_0^h \ddot{b}(X_s^\theta, \theta))ds|^p] \leq ch^p(1+|x|^c)$. Replacing in (1.165) and using also on its first term (1.156) we obtain it is upper bounded by

$$\mathbb{E}[|\int_0^h (b''(X_s^\theta, \theta)(\dot{X}_s^\theta)^2 + 2\dot{b}'(X_s^\theta, \theta)\dot{X}_s^\theta + b'(X_s^\theta, \theta)\ddot{X}_s^\theta + \ddot{b}(X_s^\theta, \theta))ds|^p] \leq \tag{1.166}$$

$$\leq c(1+|x|^c)(h^{3p}+h^{2p}+h^p)+ch^{p-1}\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds.$$

Let us now consider the second term of (1.164). By Burkholder-Davis-Gundy and Jensen inequalities we get

$$\mathbb{E}[|\int_0^h (a''(X_s^\theta)(\dot{X}_s^\theta)^2+a'(X_s^\theta)\ddot{X}_s^\theta)dW_s|^p] \leq \mathbb{E}[|\int_0^h (a''(X_s^\theta)(\dot{X}_s^\theta)^2+a'(X_s^\theta)\ddot{X}_s^\theta)^2ds|^{\frac{p}{2}}] \leq$$

$$\leq h^{\frac{p}{2}-1}\mathbb{E}[\int_0^h |a''(X_s^\theta)(\dot{X}_s^\theta)^2|^p+|a'(X_s^\theta)\ddot{X}_s^\theta|^pds] \leq ch^{\frac{p}{2}+2p}(1+|x|^c)+ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds, \tag{1.167}$$

where in the last inequality we have used that the derivatives of a are supposed bounded and (1.156).
Concerning the last term of (1.164), by Kunita inequality it is upper bounded by

$$\mathbb{E}[\int_0^h \int_{\mathbb{R}} |\gamma''(X_s^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_s^\theta)\ddot{X}_s^\theta|^p|z|^p\bar{\mu}(dz,ds)]+$$

$$+\mathbb{E}[|\int_0^h \int_{\mathbb{R}} (\gamma''(X_s^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_s^\theta)\ddot{X}_s^\theta)^2z^2\bar{\mu}(dz,ds)|^{\frac{p}{2}}] \leq$$

$$\leq c\int_0^h \mathbb{E}[|\gamma''(X_s^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_s^\theta)\ddot{X}_s^\theta|^pds]+h^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\gamma''(X_s^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_s^\theta)\ddot{X}_s^\theta|^pds],$$

having used Jensen inequality and the third point of Assumption 4 in order to say that $\int_{\mathbb{R}}|z|^pF(z)dz < c$.
Using (1.156) and the boundedness of the derivatives of γ, it is upper bounded by

$$c(1+|x|^c)h^{2p+1}+c\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds+ch^{\frac{p}{2}+2p}(1+|x|^c)+ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds. \tag{1.168}$$

From (1.166), (1.167) and (1.168) we get

$$\mathbb{E}[|\ddot{X}_h^\theta|^p] \leq c(1+|x|^c)h^p(1+h^p+h^{2p}+h^{p+\frac{p}{2}}+h^{p+1})+c(1+h^{p-1}+h^{\frac{p}{2}-1})\int_0^h \mathbb{E}[|\ddot{X}_s^\theta|^p]ds.$$

Using Gronwall Lemma we obtain $\mathbb{E}[|\ddot{X}_h^\theta|^p] \leq c(1+|x|^c)h^p(1+h^p+h^{2p}+h^{p+\frac{p}{2}}+h^{p+1})$ and so $\mathbb{E}[|\ddot{X}_h^\theta|^p] \leq c(1+|x|^c)h^p$, as we wanted. ☐

Remark 11. *Supposing that the same assumptions as in Lemma* 5 *hold and acting as we have done in order to get the estimations* (1.156) *and* (1.157) *it is possible to prove that, for all* $p \geq 2$ $\exists c > 0$*:* $\forall h \leq \Delta_n$, $\forall x$,

$$\mathbb{E}[|\frac{\partial^3}{\partial\theta^3}X_h^{\theta,x}|^p\frac{1}{h^p}] \leq c(1+|x|^c). \tag{1.169}$$

1.8.1.1 Proof of Proposition 8

Proof. As in the proof of Lemma 9, we drop the dependence on the starting point in order to make the notation easier.
We recall the definition of $m_{\theta,h}(x)$:

$$m_{\theta,h}(x) := \frac{\mathbb{E}[X_h^\theta\varphi_{h^\beta}(X_h^\theta-X_0^\theta)|X_0^\theta=x]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta-X_0^\theta)|X_0^\theta=x]} = \frac{\mathbb{E}[X_h^\theta\varphi_{h^\beta}(X_h^\theta-x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta-x)]}.$$

Its derivative with respect to θ is

$$\frac{\mathbb{E}[\dot{X}_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] + \mathbb{E}[X_h^\theta h^{-\beta} \dot{X}_h^\theta \varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} - m_{\theta,h}(x) \frac{\mathbb{E}[h^{-\beta} \dot{X}_h^\theta \varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]}. \tag{1.170}$$

On the second and on the third term of (1.170) we divide and we multiply by h and then we use Proposition 7, taking $Z_1 = \frac{\dot{X}_h^\theta}{h} X_h^\theta$ and $Z_2 = \frac{\dot{X}_h^\theta}{h}$, respectively. We are allowed to do that because they are both bounded in L^p, with p arbitrary high, since we can use (1.156) on Z_2 and Holder inequality, (1.156) and the third point of Lemma 25 on Z_1. For $|x| \le h^{-k_0}$ we have

$$m_{\theta,h}(x) = x + \frac{\mathbb{E}[(X_h^\theta - x)\varphi_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} = R(\theta, 1, x), \tag{1.171}$$

where we have used that k_0 turns out in the proof of theorems 3 and 5, hence it has been chosen such that, for $|x| \le h^{-k_0}$ we have that $\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)] \ge \frac{1}{2}$. Moreover the expected value is bounded as a result of the boundedness of φ and the third point of Lemma 25. It yields, for $\epsilon > 0$ arbitrary small,

$$\dot{m}_{\theta,h} = \frac{\mathbb{E}[\dot{X}_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] + R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]}. \tag{1.172}$$

Let us now consider the first term. Replacing the dynamic of the process $\dot{X}_h^\theta$, we get

$$\begin{aligned} \mathbb{E}[\int_0^h (b'(X_s^\theta, \theta)\dot{X}_s^\theta + \dot{b}(X_s^\theta, \theta)) ds\, \varphi_{h^\beta}(X_h^\theta - x)] + \mathbb{E}[(\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s + \\ + \int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z \tilde{\mu}(dz, ds))\, \varphi_{h^\beta}(X_h^\theta - x)] = \\ = \mathbb{E}[\int_0^h \dot{b}(X_s^\theta, \theta) ds\, \varphi_{h^\beta}(X_h^\theta - x)] + R(\theta, h^{\frac{3}{2}}, x). \end{aligned} \tag{1.173}$$

In fact, using Holder inequality,

$$|\mathbb{E}[\int_0^h b'(X_s^\theta, \theta)\dot{X}_s^\theta ds\, \varphi_{h^\beta}(X_h^\theta - x)]| \le$$

$$\le (\mathbb{E}[|\int_0^h b'(X_s^\theta, \theta)\dot{X}_s^\theta ds|^p])^{\frac{1}{p}} (\mathbb{E}[\varphi_{h^\beta}^q(X_h^\theta - x)])^{\frac{1}{q}} \le (ch^{p-1} \int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p] ds)^{\frac{1}{p}},$$

where in the last inequality we have used that φ is bounded and (1.161). By (1.156), it is upper bounded by $(ch^{2p}(1 + |x|^c))^{\frac{1}{p}}$. It turns

$$\mathbb{E}[\int_0^h b'(X_s^\theta, \theta)\dot{X}_s^\theta ds\, \varphi_{h^\beta}(X_h^\theta - x)] = R(\theta, h^2, x). \tag{1.174}$$

In the same way, from Holder inequality, (1.162) and the fact that φ is bounded, we get $|\mathbb{E}[\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s\, \varphi_{h^\beta}(X_h^\theta - x)]| \le (ch^{\frac{p}{2}-1} \int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p] ds)^{\frac{1}{p}}$. Using (1.156), it yields

$$\mathbb{E}[\int_0^h a'(X_s^\theta)\dot{X}_s^\theta dW_s\, \varphi_{h^\beta}(X_h^\theta - x)] = R(\theta, h^{\frac{3}{2}}, x). \tag{1.175}$$

Using again Holder inequality, the fact that φ is bounded and (2.97) we obtain

$$|\mathbb{E}[\int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z\tilde{\mu}(dz,ds)\, \varphi_{h^\beta}(X_h^\theta - x)]| \le (c(1+h^{\frac{p}{2}-1}) \int_0^h \mathbb{E}[|\dot{X}_s^\theta|^p]ds)^{\frac{1}{p}}.$$

Using (1.156), we obtain

$$\mathbb{E}[\int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z\tilde{\mu}(dz,ds)\, \varphi_{h^\beta}(X_h^\theta - x)] = R(\theta, h^{1+\frac{1}{p}}, x),$$

where p turns out from Holder inequality. We can choose $p = 2$, getting

$$\mathbb{E}[\int_0^h \int_{\mathbb{R}} \gamma'(X_{s^-}^\theta)\dot{X}_s^\theta z\tilde{\mu}(dz,ds)\, \varphi_{h^\beta}(X_h^\theta - x)] = R(\theta, h^{\frac{3}{2}}, x). \tag{1.176}$$

Using (1.174), (1.175) and (1.176) we have (1.173), as we wanted.
The first term of (1.173) can be seen as

$$\mathbb{E}[\int_0^h (\dot{b}(X_s^\theta,\theta) - \dot{b}(x,\theta))ds\, \varphi_{h^\beta}(X_h^\theta - x)] + \mathbb{E}[\int_0^h \dot{b}(x,\theta)ds\, \varphi_{h^\beta}(X_h^\theta - x)].$$

Using Holder inequality and the fact that φ is bounded we get

$$\mathbb{E}[\int_0^h (\dot{b}(X_s^\theta,\theta) - \dot{b}(x,\theta))ds\, \varphi_{h^\beta}(X_h^\theta - x)] \le$$

$$\le c(\mathbb{E}[(\int_0^h (\dot{b}(X_s^\theta,\theta) - \dot{b}(x,\theta))ds)^p])^{\frac{1}{p}} \le c(\mathbb{E}[(\int_0^h \left\| \frac{\partial \dot{b}}{\partial x} \right\|_\infty |X_s^\theta - x| ds)^p])^{\frac{1}{p}}.$$

From Jensen inequality we get it is upper bounded by $c(h^{p-1} \int_0^h \mathbb{E}[|X_s^\theta - x|^p]ds)^{\frac{1}{p}} \le c(h^{p+1}(1+|x|^p))^{\frac{1}{p}}$, where we have used the second point of Lemma 25. It yields

$$\mathbb{E}[\int_0^h (\dot{b}(X_s^\theta,\theta) - \dot{b}(x,\theta))ds\, \varphi_{h^\beta}(X_h^\theta - x)] = R(\theta, h^{1+\frac{1}{p}}, x).$$

Taking $p = 2$, the equation (1.173) becomes

$$\mathbb{E}[\int_0^h \dot{b}(x,\theta)ds\, \varphi_{h^\beta}(X_h^\theta - x)] + R(\theta, h^{\frac{3}{2}}, x) + R(\theta, h^{\frac{3}{2}}, x) =$$

$$= \mathbb{E}[h\dot{b}(x,\theta)\varphi_{h^\beta}(X_h^\theta - x)] + R(\theta, h^{\frac{3}{2}}, x). \tag{1.177}$$

Replacing in (1.172), we get

$$\dot{m}_{\theta,h}(x) = \frac{\mathbb{E}[h\dot{b}(x,\theta)\varphi_{h^\beta}(X_h^\theta - x)] + R(\theta, h^{\frac{3}{2}}, x) + R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} =$$

$$= h\dot{b}(x,\theta) + \frac{R(\theta, h^{\frac{3}{2} \wedge (2-\alpha\beta-\epsilon-\beta)}, x)}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]}.$$

We use the developments (1.12) and (1.15) on the denominator; in both of them the function R is negligible compared to 1 without any condition on α and β. Hence for $|x| \le h^{-k_0}$ we get the expression (1.120).

In order to prove (1.121), we have to compute the second derivative of $m_{\theta,h}(x)$. From now on we will write only $\varphi^{(k)}$ for $\varphi^{(k)}_{h^\beta}(X_h^\theta - x)$, $k \geq 0$.

$$\ddot{m}_{\theta,h}(x) = \frac{\mathbb{E}[\ddot{X}_h^\theta \varphi] + h^{-\beta}\mathbb{E}[(\dot{X}_h^\theta)^2\varphi']}{\mathbb{E}[\varphi]} - \frac{h^{-\beta}\mathbb{E}[\dot{X}_h^\theta \varphi]\mathbb{E}[\dot{X}_h^\theta \varphi']}{(\mathbb{E}[\varphi])^2} +$$
$$+ h^{-\beta}\frac{\mathbb{E}[(\dot{X}_h^\theta)^2\varphi'] + \mathbb{E}[(\dot{X}_h^\theta)^2\varphi'' X_h^\theta h^{-\beta}] + \mathbb{E}[X_h^\theta \varphi' \ddot{X}_h^\theta]}{\mathbb{E}[\varphi]} +$$
$$+ h^{-2\beta}\mathbb{E}[\dot{X}_h^\theta \varphi'] \frac{m_{\theta,h}(x)\mathbb{E}[\dot{X}_h^\theta \varphi'] - \mathbb{E}[X_h^\theta \dot{X}_h^\theta \varphi']}{(\mathbb{E}[\varphi])^2} +$$
$$- h^{-\beta}\frac{\dot{m}_{\theta,h}(x)\mathbb{E}[\dot{X}_h^\theta \varphi'] + h^{-\beta} m_{\theta,h}(x)\mathbb{E}[(\dot{X}_h^\theta)^2\varphi''] + m_{\theta,h}(x)\mathbb{E}[\ddot{X}_h^\theta \varphi']}{\mathbb{E}[\varphi]}.$$

As for the study of (1.170), we want to rely on Proposition 6 to treat each term of the form $\mathbb{E}[Z\varphi^{(k)}]$, with $k \geq 1$, where Z is bounded in L^p and use $|\mathbb{E}[Z\varphi^{(k)}]| \leq \mathbb{E}[|Z||\varphi^{(k)}|] = R(\theta, h^{1-\alpha\beta-\epsilon}, x)$.
We take successively the following variables as choice for Z: $(\frac{\dot{X}_h^\theta}{h})^2$, $\frac{\dot{X}_h^\theta}{h}$, $(\frac{\dot{X}_h^\theta}{h})^2$, $(\frac{\dot{X}_h^\theta}{h})^2 X_h^\theta$, $\frac{\ddot{X}_h^\theta}{h} X_h^\theta$, $\frac{\dot{X}_h^\theta}{h}$, $\frac{\dot{X}_h^\theta}{h} X_h^\theta$, $\frac{\dot{X}_h^\theta}{h}$, $(\frac{\dot{X}_h^\theta}{h})^2$, $\frac{\ddot{X}_h^\theta}{h}$.
All those variable Z are bounded in L^p for $p \geq 2$ by (1.156) - (1.157), the third point of Lemma 1 and Holder inequality. We deduce

$$\ddot{m}_{\theta,h}(x) = \frac{\mathbb{E}[\ddot{X}_h^\theta \varphi] + R(\theta, h^{3-\alpha\beta-\epsilon-\beta}, x)}{\mathbb{E}[\varphi]} - \frac{\mathbb{E}[\dot{X}_h^\theta \varphi] R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)}{(\mathbb{E}[\varphi])^2} + \tag{1.178}$$
$$+ \frac{R(\theta, h^{3-\alpha\beta-\epsilon-\beta}, x) + R(\theta, h^{3-\alpha\beta-\epsilon-2\beta}, x) + R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)}{\mathbb{E}[\varphi]} + \frac{R(\theta, h^{4-2\alpha\beta-\epsilon-2\beta}, x)}{(\mathbb{E}[\varphi])^2} +$$
$$- \frac{R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)\dot{m}_{\theta,h}(x) + R(\theta, h^{3-\alpha\beta-\epsilon-2\beta}, x) + R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)}{\mathbb{E}[\varphi]}.$$

We are no longer considering $m_{\theta,h}(x)$ because, by the expression (1.171), we can include it in the function R.
Using (1.173) and (1.177), $\mathbb{E}[\dot{X}_h^\theta \varphi] = h\dot{b}(x,\theta)\mathbb{E}[\varphi] + R(\theta, h^{\frac{3}{2}}, x)$.
Hence $\mathbb{E}[\dot{X}_h^\theta \varphi] R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x) = R(\theta, h^{3-\alpha\beta-\epsilon-\beta}, x)$, by the definition of rest function R.
We have already proved (1.120), so

$$R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x)\dot{m}_{\theta,h}(x) = R(\theta, h^{3-\alpha\beta-\epsilon-\beta}, x).$$

Let us now consider $\mathbb{E}[\ddot{X}_h^\theta \varphi]$. Replacing the dynamic of $\ddot{X}_h^\theta$ by (1.159), it is

$$\mathbb{E}[\varphi \int_0^h \ddot{b}(X_s^\theta, \theta) ds] + \mathbb{E}[\varphi \int_0^h (b''(X_s^\theta, \theta)(\dot{X}_s^\theta)^2 + 2\dot{b}'(X_s^\theta, \theta)\dot{X}_s^\theta + b'(X_s^\theta, \theta)\ddot{X}_s^\theta) ds] +$$
$$+ \mathbb{E}[\varphi \int_0^h (a''(X_s^\theta)(\dot{X}_s^\theta)^2 + a'(X_s^\theta)\ddot{X}_s^\theta dW_s] +$$
$$+ \mathbb{E}[\varphi \int_0^h \int_{\mathbb{R}} (\gamma''(X_{s^-}^\theta)(\dot{X}_s^\theta)^2 + \gamma'(X_{s^-}^\theta)\ddot{X}_s^\theta) z\tilde{\mu}(dz, ds)] =$$
$$= \mathbb{E}[\varphi \int_0^h \ddot{b}(X_s^\theta, \theta) ds] + R(\theta, h^{\frac{3}{2}}, x). \tag{1.179}$$

Indeed, using Holder inequality,

$$|\mathbb{E}[\varphi\int_0^h(b''(X_s^\theta,\theta)(\dot{X}_s^\theta)^2+2\dot{b}'(X_s^\theta,\theta)\dot{X}_s^\theta+b'(X_s^\theta,\theta)\ddot{X}_s^\theta)ds]|\le$$

$$\le(\mathbb{E}[\varphi^q])^{\frac{1}{q}}(\mathbb{E}[(\int_0^h(b''(X_s^\theta,\theta)(\dot{X}_s^\theta)^2+2\dot{b}'(X_s^\theta,\theta)\dot{X}_s^\theta+b'(X_s^\theta,\theta)\ddot{X}_s^\theta)ds)^p])^{\frac{1}{p}}\le$$

$$\le(c(1+|x|^c)h^{3p}+c(1+|x|^c)h^{2p}+ch^{p-1}\int_0^h\mathbb{E}[|\ddot{X}_s|^p]ds)^{\frac{1}{p}},$$

where in the last inequality we have used that φ is bounded and we acted as in (1.166). By (1.157), it is upper bounded by $(ch^{3p}+ch^{2p})^{\frac{1}{p}}(1+|x|^c)$. It turns

$$\mathbb{E}[\varphi\int_0^h(b''(X_s^\theta,\theta)(\dot{X}_s^\theta)^2+2\dot{b}'(X_s^\theta,\theta)\dot{X}_s^\theta+b'(X_s^\theta,\theta)\ddot{X}_s^\theta)ds]=R(\theta,h^2,x). \quad (1.180)$$

In the same way, from Holder inequality, (1.167) and the fact that φ is bounded we get

$$|\mathbb{E}[\varphi\int_0^h(a''(X_s^\theta)(\dot{X}_s^\theta)^2+a'(X_s^\theta)\ddot{X}_s^\theta)dW_s]|\le(c(1+|x|^c)h^{2p+\frac{p}{2}}+ch^{\frac{p}{2}-1}\int_0^h\mathbb{E}[|\ddot{X}_s|^p]ds)^{\frac{1}{p}}.$$

Using (1.157), it is upper bounded by $\le(ch^{2p+\frac{p}{2}}+ch^{p+\frac{p}{2}})^{\frac{1}{p}}(1+|x|^c)$ and so we obtain

$$|\mathbb{E}[\varphi\int_0^h(a''(X_s^\theta)(\dot{X}_s^\theta)^2+a'(X_s^\theta)\ddot{X}_s^\theta)dW_s]|=R(\theta,h^{\frac{3}{2}},x). \quad (1.181)$$

Using again Holder inequality, (1.168), the fact that φ is bounded and (1.157), we have

$$|\mathbb{E}[\varphi\int_0^h\int_{\mathbb{R}}(\gamma''(X_{s^-}^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_{s^-}^\theta)\ddot{X}_s^\theta)z\tilde{\mu}(dz,ds)]|\le$$

$$\le c(h^{2p+1}+h^{p+1}+h^{2p+\frac{p}{2}}+h^{p+\frac{p}{2}})^{\frac{1}{p}}(1+|x|^c).$$

Hence, since $p\ge 2$,

$$|\mathbb{E}[\varphi\int_0^h\int_{\mathbb{R}}(\gamma''(X_{s^-}^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_{s^-}^\theta)\ddot{X}_s^\theta)z\tilde{\mu}(dz,ds)]|=R(\theta,h^{1+\frac{1}{p}},x).$$

Since p turns out from Holder inequality and on which we have only the constraint $p\ge 2$, we can choose $p=2$, getting

$$|\mathbb{E}[\varphi\int_0^h\int_{\mathbb{R}}(\gamma''(X_{s^-}^\theta)(\dot{X}_s^\theta)^2+\gamma'(X_{s^-}^\theta)\ddot{X}_s^\theta)z\tilde{\mu}(dz,ds)]|=R(\theta,h^{\frac{3}{2}},x). \quad (1.182)$$

From (1.180), (1.181) and (1.182) we have (1.179) as we wanted.
The first term of (1.179) can be seen as $\mathbb{E}[\varphi\int_0^h(\ddot{b}(X_s^\theta,\theta)-\ddot{b}(x,\theta))ds]+\mathbb{E}[\varphi\int_0^h\ddot{b}(x,\theta)ds]$. Using Holder inequality and the fact that φ is bounded we get

$$\mathbb{E}[\int_0^h(\ddot{b}(X_s^\theta,\theta)-\ddot{b}(x,\theta))ds\,\varphi_{h^\beta}(X_h^\theta-x)]\le$$

$$\le c(\mathbb{E}[(\int_0^h(\ddot{b}(X_s^\theta,\theta)-\ddot{b}(x,\theta))ds)^p])^{\frac{1}{p}}\le c(\mathbb{E}[(\int_0^h\left\|\frac{\partial\ddot{b}}{\partial x}\right\|_\infty|X_s^\theta-x|ds)^p])^{\frac{1}{p}}.$$

From Jensen inequality we get it is upper bounded by $c(h^{p-1}\int_0^h \mathbb{E}[|X_s^\theta - x|^p]ds)^{\frac{1}{p}} \le c(h^{p+1}(1+|x|^p))^{\frac{1}{p}}$, where we have used the second point of Lemma 25. It yields

$$\mathbb{E}[\varphi\int_0^h(\ddot{b}(X_s^\theta,\theta)-\ddot{b}(x,\theta))ds] = R(\theta, h^{1+\frac{1}{p}}, x).$$

Therefore, considering $p=2$, (1.179) becomes $\mathbb{E}[\varphi\,\ddot{X}_h^\theta] = \mathbb{E}[\varphi\,\ddot{b}(x,\theta)h] + R(\theta, h^{\frac{3}{2}}, x)$. Replacing in (1.178) and using the development (1.12) or (1.15) of the denominator we obtain, for $|x| \le h^{-k_0}$,

$$\ddot{m}_{\theta,h}(x) = h\ddot{b}(x,\theta) + R(\theta, h^{\frac{3}{2}}, x) + R(\theta, h^{3-\alpha\beta-\epsilon-\beta}, x) + R(\theta, h^{3-\alpha\beta-\epsilon-2\beta}, x) +$$

$$+R(\theta, h^{2-\alpha\beta-\epsilon-\beta}, x) + R(\theta, h^{4-2\alpha\beta-\epsilon-2\beta}, x) = h\ddot{b}(x,\theta) + R(\theta, h^{\frac{3}{2}\wedge(2-\alpha\beta-\epsilon-\beta)}, x).$$

□

We want now to justify (1.122).
In the expression of $\dddot{m}_{\theta,h}(y)$, the numerator is the sum of product of terms with the following form:
$\mathbb{E}[\varphi^{(k)}X_h^{h_0}\dot{X}_h^{h_1}\ddot{X}_h^{h_2}\dddot{X}_h^{h_3}]h^{-\beta k}$, where $k \ge 1$ and $h_1+h_2+h_3 \ge k$.
The only term with a different form is $\mathbb{E}[\varphi\dddot{X}]$, that is $R(\theta,h,y)$ by the boundedness of φ and the equation (1.169).
We observe that, using Proposition 7 defining $Z = \frac{X_h^{h_0}\dot{X}_h^{h_1}\ddot{X}_h^{h_2}\dddot{X}_h^{h_3}}{h^{h_1+h_2+h_3}}$, we get

$$|\mathbb{E}[\varphi^{(k)}X_h^{h_0}\dot{X}_h^{h_1}\ddot{X}_h^{h_2}\dddot{X}_h^{h_3}]|h^{-\beta k} \le h^{-\beta k + h_1+h_2+h_3+1-\alpha\beta-\epsilon} \le h^{(1-\beta)k+1-\alpha\beta-\epsilon}.$$

We observe that the exponent on h is more then 1 if and only if $\beta < \frac{k}{k+\alpha} - \frac{\epsilon}{k+\alpha}$, with $k \ge 1$. Since $\frac{1}{1+\alpha} - \frac{\epsilon}{1+\alpha}$ is the smallest, the Assumption $\beta < \frac{1}{1+\alpha} - \frac{\epsilon}{1+\alpha}$ that we added in Proposition 8 assures that $|\dddot{m}_{\theta,h}(y)| = R(\theta,h,y)$, as we wanted.

1.8.2 Proof of limit theorems

In this subsection we prove the theorems stated in Section 1.6.

1.8.2.1 Proof of Proposition 3

Proof. (i) follows from Lemma 4.4 in [38], ergodic theorem and the L^1 convergence to zero of $\frac{1}{n}\sum_{i=0}^{n-1}(1+|X_{t_i}|)^c 1_{\left\{|X_{t_i}|>\Delta_{n,i}^{-k}\right\}}$, which is a consequence of the third point of Lemma 26. Remark that in [38] the Lemma 4.4 is stated the for $\alpha \in (0,1)$ only. However an inspection of the proof shows that it is valid for $\alpha \in (0,2)$.
Concerning (ii), we can see $\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}$ as

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}+\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})-1)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}. \tag{1.183}$$

We have already showed in (i) that on the first term of (1.183) we have the convergence wanted and so, in order to get the thesis, it is enough to prove the following:

$$\sup_{\theta\in\Theta}|\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})-1)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}| \xrightarrow{\mathbb{P}} 0 \tag{1.184}$$

We observe that

$$|\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)(\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})-1)1_{\left\{|X_{t_i}|\le\le\Delta_{n,i}^{-k}\right\}}|\le$$

$$\le|\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)(\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})-1)|.$$

By the definition of φ, it is different from zero only if $|\Delta X_i|>\Delta_{n,i}^{\beta}$. Using Markov inequality and Lemma 25,

$$\mathbb{P}(|X_{t_{i+1}}-X_{t_i}|>\Delta_{n,i}^{\beta})\le\mathbb{E}[|X_{t_{i+1}}-X_{t_i}|^2]\Delta_{n,i}^{-2\beta}\le c\Delta_{n,i}^{1-2\beta}. \tag{1.185}$$

It means that the left hand side of (1.184) converges to zero in L^1 and so in probability, indeed

$$\mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{t_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)(\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})-1)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|]\le$$

$$\le\mathbb{E}[\sup_{\theta\in\Theta}|\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\Delta_{n,i}f(X_{t_i},\theta)1_{\left\{|X_{t_{i+1}}-X_{t_i}|>\Delta_{n,i}^{\beta}\right\}}|]\le$$

$$\le\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\Delta_{n,i}\mathbb{E}[\sup_{\theta\in\Theta}|f(X_{t_i},\theta)|^2]^{\frac{1}{2}}\mathbb{E}[|1_{\left\{|X_{t_{i+1}}-X_{t_i}|>\Delta_{n,i}^{\beta}\right\}}|^2]^{\frac{1}{2}}\le$$

$$\le\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}\Delta_{n,i}\mathbb{P}(|X_{t_{i+1}}-X_{t_i}|>\Delta_{n,i}^{\beta})^{\frac{1}{2}}\le c\Delta_{n,i}^{\frac{1}{2}-\beta},$$

where we have first used Cauchy-Schwarz inequality and then the polynomial growth of $|\sup_{\theta\in\Theta}f|$and the third point of Lemma 26 and (1.185). Since the exponent on $\Delta_{n,i}$ is positive we get the thesis. □

1.8.2.2 Proof of Proposition 4 and Lemma 3

Proof of Proposition 4.
In order to show that $\frac{1}{t_n}\sum_{i=0}^{n-1}f_{i,n}(X_{t_i},\theta)\zeta_i\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}$ converges to zero in probability, we want to use the Lemma 9 of (missing citation) and so we have to show the following:

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\mathbb{E}[f_{i,n}(X_{t_i},\theta)\zeta_i\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}]\to0, \tag{1.186}$$

$$\frac{1}{(t_n)^2}\sum_{i=0}^{n-1}\mathbb{E}[f_{i,n}^2(X_{t_i},\theta)\zeta_i^2\varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}]\to0. \tag{1.187}$$

If Lemma 3 holds we have that, using (1.31), the left hand side of (1.186) results upper bounded by $\Delta_n^{\delta\wedge\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}|f_{i,n}(X_{t_i},\theta)|R(\theta,1,X_{t_i})$, where we have used the property (3.24) on R and the fact that $|\Delta_{n,i}|\le\Delta_n$. Since the exponent on Δ_n is positive and $\frac{1}{n}\sum_{i=0}^{n-1}f_{i,n}(X_{t_i},\theta)R(\theta,1,X_{t_i})$ is bounded in L^1 using the polynomial growth of both $f_{i,n}$ and R and the third point of Lemma 26, we get the convergence in probability (1.186).
Concerning (1.187), if Lemma 3 holds we can use (1.32) getting that (1.187) is

$\frac{1}{n^2\Delta_n}\sum_{i=0}^{n-1} f_{i,n}^2(X_{t_i},\theta)R(\theta,1,X_{t_i})$, where we have used also the property (3.24) on R and the fact that $|\Delta_{n,i}| \le \Delta_n$. Since $n\Delta_n \to \infty$ and $\frac{1}{n}\sum_{i=0}^{n-1} f_{i,n}^2(X_{t_i},\theta)R(\theta,1,X_{t_i})$ is bounded in L^1 by the polynomial growth of both $f_{i,n}$ and R and the third point of Lemma 26, we get the convergence (1.187) as we wanted.
Hence, if Lemma 3 holds, then Proposition 4 is proved. □

Proof of Lemma 3.
By the definition (1.29) of ζ_i and the dynamic of the process X, we get

$$\zeta_i = X_{t_{i+1}} - X_{t_i} - \int_{t_i}^{t_{i+1}} b(\theta_0, X_s)ds + \Delta_{n,i}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{t_i})\left[1-\varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i})z)\right]F(z)dz. \tag{1.188}$$

We write the left hand side of (1.31) by using the last equation and adding and subtracting $m_{\theta_0,\Delta_{n,i}}(X_{t_i})$:

$$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0,\Delta_{n,i}}(X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}]+ \tag{1.189}$$

$$+\mathbb{E}[(m_{\theta_0,\Delta_{n,i}}(X_{t_i}) - X_{t_i} - \int_{t_i}^{t_{i+1}} b(\theta_0,X_s)ds+$$

$$+\Delta_{n,i}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{t_i})\left[1-\varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i})z)\right]F(z)dz)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}].$$

By the $\mathcal{F}_{t_i}$-measurability of X_{t_i}, the first term of (1.189) is equal to

$$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0,\Delta_{n,i}}(X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}]1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}},$$

that is zero by the definition of $m_{\theta_0,\Delta_{n,i}}$.
On the second term of (1.189) we use the development (2.31) or (1.17), respectively for $\alpha<1$ and $\alpha>1$. Hence, we obtain

$$\mathbb{E}[(\int_{t_i}^{t_{i+1}}(b(\theta_0,X_{t_i})-b(\theta_0,X_s))ds+R(\theta_0,\Delta_{n,i}^{1+\delta},X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}], \tag{1.190}$$

where $\delta>0$ is defined below equation (1.30). Using the boundedness of both φ and the indicator function and (1.116) on the first term of (1.190), we get that (1.190) is upper bounded by

$$R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+\delta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{(1+\delta)\wedge\frac{3}{2}},X_{t_i}),$$

as we wanted.
Concerning the second point of Lemma 3, we use (1.29) in order to say that

$$\zeta_i^2 \le c(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2 + c(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz))^2+ \tag{1.191}$$

$$+c\Delta_{n,i}^2(\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{t_i})\left[1-\varphi_{\Delta_{n,i}^\beta}(\gamma(X_{t_i})z)\right]F(z)dz)^2.$$

Using this estimation in the left hand side of (1.32) we obtain three terms, the first is

$$\mathbb{E}[c(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2\varphi_{\Delta_{n,i}^\beta}^2(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}] \le \mathbb{E}[c(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2|\mathcal{F}_{t_i}],$$

by the boundedness of both φ and the indicator function. Using the conditional form of Ito's isometry it is

$$c\mathbb{E}[\int_{t_i}^{t_{i+1}} a^2(X_s)ds|\mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}, X_{t_i}), \tag{1.192}$$

by the polynomial growth of a, the third point of Lemma 25 and the definition of the function R.
We can upper bound the second term of (1.191) using first of all the boundedness of both φ and the indicator function, and then Kunita's inequality in the conditional form (Appendix of [46]). We get the following estimation:

$$\mathbb{E}[c(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz))^2 \varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \le \Delta^{-k}_{n,i}\}}|\mathcal{F}_{t_i}] \le$$

$$\le \mathbb{E}[(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} cz\,\gamma(X_{s^-})\tilde{\mu}(ds,dz))^2|\mathcal{F}_{t_i}] \le$$

$$\le \mathbb{E}[\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} c|z|^2\,\gamma^2(X_{s^-})\bar{\mu}(ds,dz)|\mathcal{F}_{t_i}] \le$$

$$\le c\mathbb{E}[\int_{t_i}^{t_{i+1}} \gamma^2(X_{s^-})ds|\mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}, X_{t_i}), \tag{1.193}$$

where in the last inequality and equality we have used, respectively, the definition of the compensator measure $\bar{\mu}$ and the polynomial growth of γ and the third point of Lemma 25.
Concerning the third term of (1.191), we have already showed in Remark 3 an estimation, depending on α, that is at most $\Delta^{\frac{1}{2}}_{n,i}$. Its square is therefore at least a $R(\theta, \Delta_{n,i}, X_{t_i})$ function, it follows that (1.32) holds.
We now want to prove (2.26). Using (1.30),

$$(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2 \le c\zeta_i^2 + c(\int_{t_i}^{t_{i+1}} b(X_s, \theta_0)ds - \Delta_{n,i}b(X_{t_i}, \theta_0))^2 + R(\theta_0, \Delta^{2+2\delta}_{n,i}, X_{t_i}). \tag{1.194}$$

We can replace it in (2.26), getting three terms that are of magnitude at most $\Delta_{n,i}$. Indeed, on the first we can use (1.32).
On the second term we can use the boundedness of both φ and the indicator function and Jensen inequality, getting

$$c\mathbb{E}[(\int_{t_i}^{t_{i+1}} b(X_s, \theta_0)ds - \Delta_{n,i}b(X_{t_i}, \theta_0))^2 \varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \le \Delta^{-k}_{n,i}\}}|\mathcal{F}_{t_i}] \le$$

$$c\Delta_{n,i}\mathbb{E}[\int_{t_i}^{t_{i+1}} (b(X_s, \theta_0)ds - b(X_{t_i}, \theta_0))^2|\mathcal{F}_{t_i}] \le$$

$$\le c\Delta_{n,i}\mathbb{E}[\int_{t_i}^{t_{i+1}} b^2(X_s, \theta_0)ds|\mathcal{F}_{t_i}] + c\Delta^2_{n,i}\mathbb{E}[b^2(X_{t_i}, \theta_0)|\mathcal{F}_{t_i}] = R(\theta_0, \Delta^2_{n,i}, X_{t_i}), \tag{1.195}$$

where in the last equality we have used the polynomial growth of b on both of the two terms and moreover the third point of Lemma 25 on the first term.
In conclusion, we obtain

$$\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2 \varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \le \Delta^{-k}_{n,i}\}}|\mathcal{F}_{t_i}] =$$

$$= R(\theta_0, \Delta_{n,i}, X_{t_i}) + R(\theta_0, \Delta^2_{n,i}, X_{t_i}) + R(\theta_0, \Delta^{2+2\delta}_{n,i}, X_{t_i}) = R(\theta_0, \Delta_{n,i}, X_{t_i}).$$

Hence, we have the thesis. □

1.8.2.3 Proof of Proposition 5.

In order to prove Proposition 5, the following lemma will be useful:

Lemma 10. *Let us denote by $\tilde{X}^J$ the jump part of X given by*

$$\tilde{X}_t^J := \int_0^t \int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz), \qquad t \geq 0 \tag{1.196}$$

and $\Delta_i \tilde{X}^J := \tilde{X}^J_{t_{i+1}} - \tilde{X}^J_{t_i}$.
Then, for each $q \geq 2$, $\exists \epsilon > 0$ such that

$$\mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q | \mathcal{F}_{t_i}] = R(\theta_0, \Delta_{n,i}^{1+\beta(q-\alpha)}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{1+\epsilon}, X_{t_i}). \tag{1.197}$$

Proof of Lemma 10.
For all $n \in \mathbb{N}$ and $i \in \mathbb{N}$ we define the set on which all the jumps of L on the interval $(t_i, t_{i+1}]$ are small:

$$N_n^i := \left\{ |\Delta L_s| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}; \quad \forall s \in (t_i, t_{i+1}] \right\}, \tag{1.198}$$

where $\Delta L_s := L_s - L_{s^-}$. We hence split the left hand side of (1.197) as

$$\mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{N_n^i} | \mathcal{F}_{t_i}] + \mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{(N_n^i)^c} | \mathcal{F}_{t_i}]. \tag{1.199}$$

We now observe that, by the definition of N_n^i,

$$|\mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{N_n^i} | \mathcal{F}_{t_i}]| \leq$$

$$\leq c\mathbb{E}[|\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz)|^q + |\int_{t_i}^{t_{i+1}} \int_{|z| \geq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|\,|\gamma(X_{s^-})|\bar{\mu}(ds,dz)|^q | \mathcal{F}_{t_i}]. \tag{1.200}$$

We observe that the order of the second term depend on α. Acting as in Remark 3, we get that its order is $\Delta_{n,i}^q$ if $\alpha \in (0,1)$ while it is $\Delta_{n,i}^{q+q\beta(1-\alpha)}$ if $\alpha \in (1,2)$. Since q is more than $q + q\beta(1-\alpha)$ if and only if $\alpha > 1$, we can say that the second term of (1.200) is upper bounded by $c\Delta_{n,i}^{q \wedge (q+q\beta(1-\alpha))}$. The first term of (1.200) is instead upper bounded by

$$c\mathbb{E}[|\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|^2\,|\gamma(X_{s^-})|^2 \bar{\mu}(ds,dz)|^{\frac{q}{2}} | \mathcal{F}_{t_i}] +$$

$$+ c\mathbb{E}[\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|^q\,|\gamma(X_{s^-})|^q \bar{\mu}(ds,dz) | \mathcal{F}_{t_i}] \leq$$

$$\leq c\,\|\gamma\|_\infty^q (\mathbb{E}[|\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|^{1-\alpha} dz ds|^{\frac{q}{2}} | \mathcal{F}_{t_i}] + \mathbb{E}[\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|^{q-1-\alpha} dz ds | \mathcal{F}_{t_i}]) \leq$$

$$\leq c(\int_{t_i}^{t_{i+1}} \Delta_{n,i}^{(2-\alpha)\beta} ds)^{\frac{q}{2}} + c\int_{t_i}^{t_{i+1}} \Delta_{n,i}^{(q-\alpha)\beta} ds + \Delta_{n,i}^q \leq c(\Delta_{n,i}^{(1+(2-\alpha)\beta)\frac{q}{2}} + \Delta_{n,i}^{(q-\alpha)\beta+1}) = \tag{1.201}$$

$$= R(\theta_0, \Delta_{n,i}^{(q-\alpha)\beta+1}, X_{t_i}),$$

where we have used Kunita inequality, the definition of $\bar{\mu}$ and the second point of Assumption 4. Using the consideration below equation (1.200) and (1.201) we get

$$|\mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{N_n^i} | \mathcal{F}_{t_i}]| \le R(\theta_0, \Delta_{n,i}^{(q-\alpha)\beta+1}, X_{t_i}) + \tag{1.202}$$

$$+R(\theta_0, \Delta_{n,i}^{q\wedge(q+q\beta(1-\alpha))}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{(q-\alpha)\beta+1}, X_{t_i}).$$

For $\alpha \in (0,2)$, $\alpha \neq 1$ and $\beta \in (0, \frac{1}{2})$ the exponent on $\Delta_{n,i}$ can be seen as $1+\epsilon$, with $\epsilon > 0$.
Concerning the second term of (2.86), we have

$$\mathbb{E}[|\Delta_i \tilde{X}^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{(N_n^i)^c} | \mathcal{F}_{t_i}] \le \tag{1.203}$$

$$\le c\mathbb{E}[(|\Delta_i X|^q + |\Delta X_i^c|^q)|\varphi_{\Delta_{n,i}^\beta}^q(X_{t_{i+1}} - X_{t_i})| 1_{(N_n^i)^c} | \mathcal{F}_{t_i}],$$

where $|\Delta_i X| := |X_{t_{i+1}} - X_{t_i}|$ and ΔX_i^c is the increment of the continuous part of X in the interval $(t_i, t_{i+1}]$. We observe that, by the definition of $\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})$, the first term in the right hand side is different from zero only if $|\Delta_i X|^q \le \Delta_{n,i}^{\beta q}$. Therefore

$$\mathbb{E}[|\Delta_i X|^q \varphi_{\Delta_{n,i}^\beta}^q(X_{t_{i+1}} - X_{t_i}) 1_{(N_n^i)^c} | \mathcal{F}_{t_i}] \le \Delta_{n,i}^{\beta q} \mathbb{P}_i((N_n^i)^c) \le c\Delta_{n,i}^{\beta q + 1 - \alpha\beta}. \tag{1.204}$$

Indeed

$$\mathbb{P}_i((N_n^i)^c) = \mathbb{P}_i(\exists s \in (t_i, t_{i+1}] \, : \, |\Delta L_s| > \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}) \le c \int_{t_i}^{t_{i+1}} \int_{\frac{4\Delta_{n,i}^\beta}{\gamma_{min}}}^\infty F(z) dz ds \le c\Delta_{n,i}^{1-\alpha\beta}, \tag{1.205}$$

where we have used the third point of Assumption 4. Since $q \ge 2$, $\beta q + 1 - \alpha\beta$ is always more than 1.
In the same way

$$\mathbb{E}[|\Delta X_i^c|^q \varphi_{\Delta_{n,i}^\beta}^q(X_{t_{i+1}} - X_{t_i}) 1_{(N_n^i)^c} | \mathcal{F}_{t_i}] \le c\Delta_{n,i}^{\frac{1}{2}q} \Delta_{n,i}^{1-\alpha\beta}(1 + |X_{t_i}|^c), \tag{1.206}$$

that is again more than 1. Using (2.86), (1.202), (1.204) and (1.206) we get the thesis. □

We can now prove Proposition 5.

Proof of Proposition 5.
We denote

$$s_i^n := \frac{1}{\sqrt{t_n}}(X_{t_{i+1}} - m_{\theta_0}(X_{t_i})) f(X_{t_i}, \theta) \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}. \tag{1.207}$$

In order to show the asymptotic normality we have to prove that s^n is a martingale difference array such that

$$\sum_{i=0}^{n-1} \mathbb{E}[|s_i^n|^{2+r} | \mathcal{F}_{t_i}] \xrightarrow{\mathbb{P}} 0, \tag{1.208}$$

for a constant $\delta > 0$, and

$$\sum_{i=0}^{n-1} \mathbb{E}[|s_i^n|^2|\mathcal{F}_{t_i}] \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f^2(x,\theta)a^2(x)\pi(dx), \tag{1.209}$$

c.f. Theorem A2 in the Appendix of [88].
We observe that s_i^n is a martingale difference array since, $\forall i \geq 0$,

$$\mathbb{E}[s_i^n|\mathcal{F}_{t_i}] = \frac{f(X_{t_i},\theta)1_{\{|X_{t_i}|\leq \Delta_{n,i}^{-k}\}}}{\sqrt{t_n}}\mathbb{E}[(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}] = 0,$$

by the measurability of f and the indicator function and the definition of $m_{\theta_0}(X_{t_i})$.
We now want to prove (1.209). Using (1.30) and the definition of ζ_i we have that

$$(X_{t_{i+1}} - m_{\theta_0}(X_{t_i}))^2 = (\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2 + 2B_{i,n}\int_{t_i}^{t_{i+1}} a(X_s)dW_s + B_{i,n}^2, \tag{1.210}$$

where

$$B_{i,n} := \int_{t_i}^{t_{i+1}} (b(X_s,\theta_0) - b(X_{t_i},\theta_0))ds + R(\theta_0, \Delta_{n,i}^{1+\delta}, X_{t_i}) +$$

$$+\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz) + \Delta_{n,i}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})[1-\varphi_{\Delta_{n,i}^{\beta}}(\gamma(X_{t_i})z)]F(z)dz.$$

Replacing (1.210) in the definition (1.207) of s_i^n we get three terms. We start proving that

$$\frac{1}{t_n}\sum_{i=0}^{n-1}\mathbb{E}[B_{i,n}^2 f^2(X_{t_i},\theta)\varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}|\leq \Delta_{n,i}^{-k}\}}|\mathcal{F}_{t_i}] \xrightarrow{\mathbb{P}} 0. \tag{1.211}$$

Indeed,

$$\mathbb{E}[B_{i,n}^2\varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}] \leq c\mathbb{E}[[(\int_{t_i}^{t_{i+1}}(b(X_s,\theta_0) - b(X_{t_i},\theta_0))ds)^2 +$$

$$+R(\theta_0,\Delta_{n,i}^{2+2\delta}, X_{t_i}) + (\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz))^2 + \tag{1.212}$$

$$+(\Delta_{n,i}\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})[1-\varphi_{\Delta_{n,i}^{\beta}}(\gamma(X_{t_i})z)]F(z)dz)^2]\varphi_{\Delta_{n,i}^{\beta}}^2(X_{t_{i+1}} - X_{t_i})|\mathcal{F}_{t_i}] \leq$$

$$\leq R(\theta_0,\Delta_{n,i}^2, X_{t_i}) + R(\theta_0,\Delta_{n,i}^{2+2\delta}, X_{t_i}) + R(\theta_0,\Delta_{n,i}^{(1+\beta(q-\alpha))\wedge(1+\epsilon)}, X_{t_i}), \tag{1.213}$$

where we have used (1.195) on the first term of (1.212), (1.197) of the previous lemma on the third and Remark 3 on the fourth. Indeed, in Remark 3, we found that the last term in less than $R(\theta_0,\Delta_{n,i}^2, X_{t_i})$ if $\alpha \leq 1$ and less than $R(\theta_0,\Delta_{n,i}^{2+2\beta(1-\alpha)}, X_{t_i})$ if $\alpha > 1$; in both cases the exponent on $\Delta_{n,i}$ is always more than 1, hence we can write it as $1+\epsilon$.
We can upper bound with (1.213) the left hand side of (1.211) getting $\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)R(\theta_0,\Delta_{n,i}^{1+\epsilon}, X_{t_i})1_{\{|X_{t_i}|\leq \Delta_{n,i}^{-k}\}}$, that converges to 0 in norm L^1 by the polynomial growth of both f and R and the third point of Lemma 26 and using that $|\Delta_{n,i}| \leq \Delta_n$. We obtain therefore the convergence in probability (1.211) wanted.
Let us now consider the contribution of the first term of (1.210) for the proof of (1.209). We can see it as

$$\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2|\mathcal{F}_{t_i}]1_{\{|X_{t_i}|\leq \Delta_{n,i}^{-k}\}} + \tag{1.214}$$

$$+\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2(\varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}}-X_{t_i})-1)|\mathcal{F}_{t_i}]1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}.$$

On the first term of (1.214) we use Ito's isometry, getting

$$\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[\int_{t_i}^{t_{i+1}} a(X_s)^2 ds|\mathcal{F}_{t_i}]1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} =$$

$$=\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)(\Delta_{n,i}a^2(X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}))1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}, \tag{1.215}$$

where we have used (1.116) with a^2 in place of b. Using the first point of Proposition 1 we get that

$$\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\Delta_{n,i}a^2(X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f^2(x,\theta)a^2(x)\pi(dx), \tag{1.216}$$

while $\frac{1}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}$ goes to zero in norm L^1 and therefore in probability.
Let us now consider the second term of (1.214). Using Cauchy- Schwarz inequality we get it is upper bounded by

$$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^4|\mathcal{F}_{t_i}]^{\frac{1}{2}}\mathbb{E}[|\varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}}-X_{t_i})-1|^2|\mathcal{F}_{t_i}]^{\frac{1}{2}} \le$$

$$\le \frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[|\int_{t_i}^{t_{i+1}} a^2(X_s)ds|^2|\mathcal{F}_{t_i}]^{\frac{1}{2}}\mathbb{E}[1_{\left\{|X_{t_{i+1}}-X_{t_i}|>\Delta^{\beta}_{n,i}\right\}}|\mathcal{F}_{t_i}]^{\frac{1}{2}},$$

where we have used Burkholder Davis Gundy inequality and the fact that, by the definition of φ, it is different from 0 only if $|X_{t_{i+1}}-X_{t_i}|>\Delta^{\beta}_{n,i}$. Using Jensen inequality and (1.185) in the conditional form we can upper bound it with

$$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[\Delta^2_{n,i}|\frac{1}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}} a^2(X_s)ds|^2|\mathcal{F}_{t_i}]^{\frac{1}{2}}\mathbb{P}(|X_{t_{i+1}}-X_{t_i}|>\Delta^{\beta}_{n,i}|\mathcal{F}_{t_i})^{\frac{1}{2}} \le$$

$$\frac{c}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[\Delta_{n,i}\int_{t_i}^{t_{i+1}} a^4(X_s)ds|\mathcal{F}_{t_i}]^{\frac{1}{2}}R(\theta_0,\Delta_{n,i}^{\frac{1}{2}-\beta},X_{t_i}) \le$$

$$\le \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\Delta_{n,i}^{\frac{1}{2}}[\Delta_{n,i}a^4(X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i})]^{\frac{1}{2}}R(\theta_0,\Delta_{n,i}^{\frac{1}{2}-\beta},X_{t_i}), \tag{1.217}$$

where we have also used (1.116) with a^4 in place of b. We observe that (1.217) goes to 0 in L^1 and therefore in probability, indeed its L^1 norm is upper bounded by

$$\le \Delta_n^{\frac{1}{2}-\beta}\frac{c}{n}\sum_{i=0}^{n-1}\mathbb{E}[f^2(X_{t_i},\theta)R(\theta_0,1,X_{t_i})(a^2(X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{3}{4}},X_{t_i}))],$$

that goes to 0 by the polynomial growth of f, R and a and the third point of Lemma 25 and since $\beta<\frac{1}{2}$.
Let us now consider the second term of (1.210) for the proof of (1.209). Using Cauchy-Schwarz inequality, (1.213) and Ito's isometry we get

$$\frac{2}{t_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)\mathbb{E}[B_{i,n}\int_{t_i}^{t_{i+1}} a(X_s)dW_s\,\varphi^2_{\Delta^{\beta}_{n,i}}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}|\mathcal{F}_{t_i}] \le$$

$$\leq \frac{c}{n\Delta_n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)R(\theta_0,\Delta_{n,i}^{1+\epsilon},X_{t_i})^{\frac{1}{2}}\mathbb{E}[\int_{t_i}^{t_{i+1}} a(X_s)^2 ds|\mathcal{F}_{t_i}]^{\frac{1}{2}} \leq$$

$$\leq \Delta_n^{\frac{\epsilon}{2}}\frac{c}{n}\sum_{i=0}^{n-1} f^2(X_{t_i},\theta)R(\theta_0,1,X_{t_i})(a^2(X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{1}{4}},X_{t_i})), \tag{1.218}$$

where in the last inequality we have used the property (3.24) of R and (1.215) with the trivial estimation $|\Delta_{n,i}| \leq \Delta_n$. By the polynomial growth of both a, f and R and the fact that the exponent on Δ_n is positive we have that (1.218) converges to 0 en norm L^1. Hence it converges to 0 in probability, (1.209) follows.
Our goal is now to prove (1.208). Using (1.210) we have that

$$\sum_{i=0}^{n-1}\mathbb{E}[|s_i^n|^{2+r}|\mathcal{F}_{t_i}] \leq$$

$$\leq c\frac{1}{(n\Delta_n)^{1+\frac{r}{2}}}\sum_{i=0}^{n-1} f^{2+r}(X_{t_i},\theta)(\mathbb{E}[B_{i,n}^{2+r}\varphi^{2+r}_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})|\mathcal{F}_{t_i}]+ \tag{1.219}$$

$$+\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^{2+r}|\mathcal{F}_{t_i}]).$$

We act as we have already done in the proof of (1.209) on the first term of (1.219): using (1.197) we get it is upper bounded by

$$\frac{c}{(n\Delta_n)^{1+\frac{r}{2}}}\sum_{i=0}^{n-1} f^{2+r}(X_{t_i},\theta)R(\theta_0,\Delta_{n,i}^{1+\epsilon},X_{t_i}) \leq \Delta_n^{\epsilon}\frac{c}{(n\Delta_n)^{\frac{r}{2}}}\frac{1}{n}\sum_{i=0}^{n-1} f^{2+r}(X_{t_i},\theta)R(\theta_0,1,X_{t_i}),$$

that converges to 0 in norm L^1 (and therefore in probability) since $\epsilon > 0$ and $n\Delta_n \to \infty$ for $n \to \infty$. Concerning the second term of (1.219), using Burkholder-Davis-Gundy inequality and (1.215) we have

$$\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^{2+r}|\mathcal{F}_{t_i}]) \leq R(\theta_0,\Delta_{n,i}^{1+\frac{r}{2}},X_{t_i}). \tag{1.220}$$

Using (1.220) we get that the second term of (1.219) is upper bounded by $\frac{c}{n^{\frac{r}{2}}}\frac{1}{n}\sum_{i=0}^{n-1} f^{2+r}(X_{t_i},\theta)R(\theta_0,1,X_{t_i})$, that converges to 0 in norm L^1 and hence in probability since $n^{\frac{r}{2}} \to \infty$. We deduce (1.208) and therefore the wanted asymptotic normality. □

1.8.3 Proof of Propositions 19 and 7.

Since Proposition 19 is a consequence of Proposition 7, let us start with the proof of Proposition 7. To lighten the notation we forget the dependence on θ of X^θ and Z_θ.
Proof of Proposition 7.
Using $\tilde{X}_t^J$ defined in (1.196), we introduce the event

$$E_h := \left\{\tilde{X}_h^J := \int_0^t\int_{\mathbb{R}\setminus\{0\}} z\,\gamma(X_{s^-})\tilde{\mu}(ds,dz) \in [\frac{1}{2}h^\beta, 4h^\beta]\right\}. \tag{1.221}$$

We have that

$$\mathbb{E}[|Z\varphi_{h^\beta}^{(k)}(X_h - x)|] = \mathbb{E}[|Z\varphi_{h^\beta}^{(k)}(X_h - x)|1_{E_h}] + \mathbb{E}[|Z\varphi_{h^\beta}^{(k)}(X_h - x)|1_{E_h^c}]. \quad (1.222)$$

We observe that, by its definition, $\varphi_{h^\beta}^{(k)}(X_h - x)$ is different from 0 only if $|X_h - x| \in [h^\beta, 2h^\beta]$. But $\Delta_h X := |X_h - x| = |X_h^c - x + \tilde{X}_h^J|$ hence on E_h^c, where $\tilde{X}_h^J \notin [\frac{1}{2}h^\beta, 4h^\beta]$, from $|X_h - x| \in [h^\beta, 2h^\beta]$ we deduce that it must be $|X_h^c - x| \geq \frac{1}{2}h^\beta$. Using this observation and Holder inequality we have that the second term on the right hand side of (1.222) is upper bounded by

$$(\mathbb{E}[|Z|^p])^{\frac{1}{p}}(\mathbb{E}[|\varphi_{h^\beta}^{(k)}(X_h - x)|^q 1_{E_h^c}])^{\frac{1}{q}} \leq c(\mathbb{P}(|X_h^c - x| \geq \frac{1}{2}h^\beta))^{\frac{1}{q}} \leq ch^{\frac{r}{q}(\frac{1}{2}-\beta)}$$

$\forall r > 1$, where we have also used that Z is bounded in L^p and Remark 2 in [38].
In order to estimate the first term on the right hand side of (1.222) we need the following lemma that we will prove at the end of the section:

Lemma 11. *Let us consider E_h, the set defined in* (1.221). *We have*

$$\mathbb{P}(E_h) \leq R(\theta, h^{1-\beta\alpha}, x). \quad (1.223)$$

If $Z \in \mathcal{Z}_{h,c,p}$, then using Holder inequality, the estimation (1.223) and the boundedness of Z in L^p we get

$$\mathbb{E}[|Z\varphi_{h^\beta}^{(k)}(X_h - x)|1_{E_h}] \leq (\mathbb{E}[|Z|^p]^{\frac{1}{p}})(\mathbb{E}[|\varphi_{h^\beta}^{(k)}(X_h - x)|^q 1_{E_h}])^{\frac{1}{q}} \leq$$

$$\leq cR(\theta, h^{1-\beta\alpha}, x)^{\frac{1}{q}} = cR(\theta, h^{\frac{1-\beta\alpha}{q}}, x),$$

with $\frac{1}{p} + \frac{1}{q} = 1$. Hence, we get the Proposition 7. □

Proposition 19 is a consequence of Proposition 7, observing that $(h(X_u, \theta))_{\theta \in \Theta} \in \mathcal{Z}_{t_{i+1}-t_i,c,p}$, for $u \in [t_i, t_{i+1}]$, and the Markov property.

In conclusion, we prove Lemma 11.

Proof of Lemma 11.
We use again the set N_n^i defined in (1.198). We have

$$\mathbb{P}(E_h) = \mathbb{P}(E_h \cap N_n^i) + \mathbb{P}(E_h \cap (N_n^i)^c). \quad (1.224)$$

On the second term of (1.224) we use (1.205), getting

$$\mathbb{P}(E_h \cap (N_n^i)^c) \leq \mathbb{P}((N_n^i)^c) \leq ch^{1-\alpha\beta}. \quad (1.225)$$

Concerning the set $E_h \cap N_n^i$, we use Markov inequality and we obtain, $\forall r > 1$,

$$\mathbb{P}(E_h \cap N_n^i) \leq c\mathbb{E}[|\tilde{X}_h^J|^r 1_{N_n^i}]h^{-\beta r} \leq ch^{-\beta r}h^{1+\beta(r-\alpha)} = ch^{1-\beta\alpha}, \quad (1.226)$$

where in the last inequality we used (1.202).
Using (1.224), (1.225) and (1.226) we get the Lemma 11. □

Chapter 2

Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function.

Abstract :

In this chapter we consider an ergodic diffusion process with jumps whose drift coefficient depends on μ and volatility coefficient depends on σ, two unknown parameters. We suppose that the process is discretely observed at the instants $(t_i^n)_{i=0,\dots,n}$ with $\Delta_n = \sup_{i=0,\dots,n-1}(t_{i+1}^n - t_i^n) \to 0$.
We introduce an estimator of $\theta := (\mu, \sigma)$, based on a contrast function, which is asymptotically gaussian without requiring any conditions on the rate at which $\Delta_n \to 0$, assuming a finite jump activity.
This extends earlier results where a condition on the step discretization was needed (see [38],[88]) or where only the estimation of the drift parameter was considered (see [38] and previous chapter).
In general situations, our contrast function is not explicit and in practise one has to resort to some approximation. We propose explicit approximations of the contrast function, such that the estimation of θ is feasible under the condition that $n\Delta_n^k \to 0$ where $k > 0$ can be arbitrarily large. This extends the results obtained by Kessler [51] in the case of continuous processes.

Keys words : Efficient drift estimation, volatility estimation, ergodic properties, high frequency data, Lévy-driven SDE, thresholding methods.

2.1 Introduction

Recently, diffusion processes with jumps are becoming powerful tools to model various stochastic phenomena in many areas, for example, physics, biology, medical sciences, social sciences, economics, and so on. In finance, jump-processes were introduced to model the dynamic of exchange rates ([11]), asset prices ([70],[56]), or volatility processes ([8],[31]). Utilization of jump-processes in neuroscience, instead, can be found for instance in [26]. Therefore, inference problems for such models from various types of data should be studied, in particular, inference from discrete observation should be desired since the actual data may be obtained discretely In this chapter, our aim is to estimate jointly the drift and the volatility parameter $(\mu, \sigma) =: \theta$ from a discrete sampling of the process X^θ solution to

$$X_t^\theta = X_0^\theta + \int_0^t b(\mu, X_s^\theta)ds + \int_0^t a(\sigma, X_s^\theta)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s-}^\theta) z \tilde{\mu}(ds, dz),$$

where W is a one dimensional Brownian motion and $\tilde{\mu}$ a compensated Poisson random measure, with a finite jump activity. We assume that the process is sampled at the times $(t_i^n)_{i=0,\dots,n}$ where the sampling step $\Delta_n := \sup_{i=0,\dots,n-1} t_{i+1}^n - t_i^n$ goes to zero. A crucial point for applications in the high frequency setting is to impose minimal conditions on the sampling step size. This will be one of our main objectives in this chapter, for the joint estimation of μ and σ.
It is known that, as a consequence of the presence of a Gaussian component, it is impossible to estimate the drift parameter on a finite horizon time; we therefore assume that $t_n^n \to \infty$ and we suppose to have an ergodic process X^θ.

The topic of high frequency estimation for discretely observed diffusions in the case without jumps is well developed, by now. Florens-Zmirou has introduced, in [33], an estimator for both the drift and the diffusion parameters under the fast sampling assumption $n\Delta_n^2 \to 0$. Yoshida [95] has then suggested a correction of the contrast function of [33] that releases the condition on the step discretization to $n\Delta_n^3 \to 0$. In Kessler [51], the author proposes an explicit modification of the Euler scheme contrast that allows him to build an estimator which is asymptotically normal under the condition $n\Delta_n^k \to 0$ where $k \geq 2$ is arbitrarily large. The result found by Kessler, therefore, holds for any arbitrarily slow polynomial decay to zero of the sampling step.

When a jump component is added, less results are known. Shimizu [87] proposes parametric estimation of drift, diffusion and jump coefficients showing the asymptotic normality of the estimators under some explicit conditions relating the sampling step and the jump intensity of the process; such conditions on Δ_n are more restrictive as the intensity of jumps near zero is high. In the situation where the jump intensity is finite, the conditions of [87] reduces to $n\Delta_n^2 \to 0$.
In [38], the condition on the sampling step is relaxed to $n\Delta_n^3 \to 0$, when one estimates the drift parameter only. In [73] a jump-filtering technique similar to one used in [38] is employed in order to derive a nonparametric estimator for the drift which is robust to symmetric jumps of infinite variance and infinite variation, and which attains the same asymptotic variance as for a continuous diffusion process.
Also in the previous chapter only the estimation of the drift parameter is studied. In such a case, the sampling step $(t_i^n)_{i=0,\dots,n}$ can be irregular, no condition on the rate at which $\Delta_n \to 0$ is needed and the assumption that the jumps of the process are summable, present in [38], is suppressed.

In this chapter, we consider the joint estimation of the drift and the diffusion parameters with a jump intensity which is finite. Since for the applications it is important that assumptions on the rate at which Δ_n should tend to zero are less stringent as possible, we aim to weaken the conditions on the decay of the sampling step in a way comparable to Kessler's work [51], but in the framework of jump-diffusion processes. We therefore want to extend the previous chapter looking for the same results, but for the joint estimation of the drift and the diffusion parameters instead of focusing on the drift parameter only.
The joint estimation of the two parameters introduces some significant difficulties: since the drift and the volatility parameters aren't estimated at the same rate, we have to deal with asymptotic properties in two different regimes.
Compared to previous results in which the parameters are estimated jointly (see [87]), we show that it is possible to remove any condition on the rate at which Δ_n has to go to zero.
Moreover, we consider a discretization step which isn't uniform. This case, to our knowledge, has never been studied before for the joint estimation of the drift and the volatility of a diffusion with jumps.

A natural approach to estimate the unknown parameters would be to use a maximum likelihood estimation, but the likelihood function based on the discrete sample is not tractable in this setting, since it depends on the transition densities of X which are not explicitly known.
To overcome this difficulty several methods have been developed. For instance, in [2] and [62] closed form expansions of the transition density of jump-diffusions are studied while in [50] the asymptotic behaviour of estimating functions is considered in the high frequency observation framework. They give condition to ensure the rate optimality and the efficiency.

Considering again the case of high frequency observation, a widely-used method is to consider pseudo-likelihood function, for instance based on the high frequency approximation of the dynamic of the process by the dynamic of the Euler scheme. This leads to explicit contrast functions with Gaussian structures (see e.g. [88],[87],[68]).

In Kessler's paper the idea is to replace, in the Euler scheme contrast function, the contribution of the drift and the diffusion by two quantities m and m_2 (or their explicit approximations with arbitrarily high order when $\Delta_n \to 0$); with

$$m(\mu, \sigma, x) := \mathbb{E}[X^\theta_{t_{i+1}} | X^\theta_{t_i} = x] \quad \text{and} \tag{2.1}$$

$$m_2(\mu, \sigma, x) := \mathbb{E}[(X^\theta_{t_{i+1}} - m(\mu, \sigma, X^\theta_{t_i}))^2 | X^\theta_{t_i} = x].$$

In presence of jumps, the contrasts functions in [88] (see also [87], [38]) resort to a filtering procedure in order to suppress the contribution of jumps and recover the continuous part of the process.

The contrast function we introduce is based on both the ideas described here above. Indeed, we define it as

$$U_n(\mu, \sigma) := \sum_{i=0}^{n-1} [\frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{m_2(\mu, \sigma, X_{t_i})} + log(\frac{m_2(\mu, \sigma, X_{t_i})}{\Delta_{n,i}})] \varphi_{\Delta^\beta_{n,i}}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}, \tag{2.2}$$

where the function φ is a smooth version of the indicator function that vanishes when the increments of the data are too large compared to the typical increments of a continuous diffusion process, and thus can be used to filter the contribution of the

jumps. The idea is to use the size of $X_{t_{i+1}} - X_{t_i}$ in order to judge if a jump occurred or not in the interval $[t_i, t_{i+1})$. The increment of X with continuous transition could hardly exceed the threshold $\Delta_{n,i}^{\beta}$, therefore we can judge the existence of a jump in the interval if $|X_{t_{i+1}} - X_{t_i}| > \Delta_{n,i}^{\beta}$, for $\beta \in (\frac{1}{4}, \frac{1}{2})$.
The last indicator in (2.2) avoids the possibility that $|X_{t_i}|$ is too big, the constant k is positive and will be chosen later, related to the developments of m and m_2 that are the natural extension to the case with jumps of the quantities proposed in [51]. Indeed, we have defined them as

$$m(\mu, \sigma, x) := \frac{\mathbb{E}[X_{t_{i+1}}^{\theta} \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta}) | X_{t_i}^{\theta} = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta}) | X_{t_i}^{\theta} = x]} \quad \text{and}$$

$$m_2(\mu, \sigma, x) := \frac{\mathbb{E}[(X_{t_{i+1}}^{\theta} - m(\mu, \sigma, X_{t_i}))^2 \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta}) | X_{t_i}^{\theta} = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta}) | X_{t_i}^{\theta} = x]}.$$

The rates for the estimation of the two parameters is not the same, which implies we have to deal with two different scaling of the contrast function, which lead us to the study of the asymptotic properties of the contrast function in two different asymptotic schemes.

The main result of this chapter is that the estimator $\hat{\theta}_n := (\hat{\mu}_n, \hat{\sigma}_n)$ associated to the proposed contrast function converges with some explicit asymptotic variances. Comparing to earlier results ([88], [87], [38], [3]), the sampling step $(t_i^n)_{i=0,\dots,n}$ can be irregular, no condition is needed on the rate at which $\Delta_n \to 0$ and the parameters of drift and diffusion are jointly estimated.

Moreover, we provide explicit approximations of m_2 that allows us to circumvent the fact that the contrast function is non explicit (explicit approximations of m are given in the chapter before). We give an expansion of m_2 exact up to order Δ_n^2, which involves the jump intensity near zero, and is valid for any smooth truncation function φ. With the specific choices of φ being oscillating functions, in particular, we remove the contribution of the jumps and we are able to prove explicit developments of the function m_2 valid up to any order. Together with the approximation of the function m showed in Proposition 2 of the previous chapter, this allows us to approximate our contrast function, at arbitrary high order, by a completely explicit one, as it was in the paper by Kessler [51] in the continuous case.
This yields to a consistent and asymptotic normal estimator under the condition $n\Delta_n^k \to 0$, where k is related to the oscillating properties of the function φ. As k can be chosen arbitrarily high, up to a proper choice of φ, our method allows to estimate the drift and the diffusion parameters, under the assumption that the sampling step tends to zero at some polynomial rate.

Furthermore, we implement numerically our main results building two approximations of m and m_2 from which we deduce two approximations of the contrast that we minimize in order to get the joint estimator of the parameters. We compare the estimators we find with the estimator that would result from the choice of an Euler scheme approximation for m and m_2. From our simulations it appears that our joint estimator performs better than the Euler one, especially for the estimation of the parameter σ.

The outline of this chapter is the following. In Section 3.2 we introduce the model and we state the assumptions we need. The Section 3.3 contains the construction of the estimator and our main results while in Section 2.4 we explain how to use in practical the contrast function for the joint estimation of the drift and the diffusion parameters, dealing with its approximation. We provide numerical results in Section 3.4. In Section 2.6 we state useful propositions that we will use repeatedly in the following sections. Section 2.7 is devoted to the proof of our main results while in Section 2.8.1 of the Appendix we prove the propositions stated in the sixth section. We conclude giving the proofs of some technical results in the Sections 2.8.2–2.8.3 of the Appendix.

2.2 Model, assumptions

We want to estimate the unknown parameter $\theta = (\mu, \sigma)$ in the stochastic differential equation with jumps

$$X_t^\theta = X_0^\theta + \int_0^t b(\mu, X_s^\theta)ds + \int_0^t a(\sigma, X_s^\theta)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}^\theta)z\tilde{\mu}(ds, dz), \quad t \in \mathbb{R}_+, \tag{2.3}$$

where θ belongs to $\Theta := \Pi \times \Sigma$, a compact set of $\mathbb{R}^2$; $W = (W_t)_{t\geq 0}$ is a one dimensional Brownian motion and μ is a Poisson random measure associated to the Lévy process $L = (L_t)_{t\geq 0}$ such that $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$. The compensated measure $\tilde{\mu} = \mu - \bar{\mu}$ is defined on $[0, \infty) \times \mathbb{R}$, the compensator is $\bar{\mu}(dt, dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.
We denote by $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which W and μ are defined and we assume that the initial condition X_0^θ, W and L are independent.

2.2.1 Assumptions

We suppose that the functions $b : \Pi \times \mathbb{R} \to \mathbb{R}$, $a : \Sigma \times \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ satisfy the following assumptions:

A1: *The functions $\gamma(x)$, $b(\mu, x)$ for all $\mu \in \Pi$ and $a(\sigma, x)$ for all $\sigma \in \Sigma$ are globally Lipschitz. Moreover, the Lipschitz constants of b and a are uniformly bounded on Π and Σ, respectively.*

Under Assumption 1 the equation (4.3) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [7] (Theorems 6.2.9. and 6.4.6.). The next assumption was used in [67] to prove the irreducibility of the process X^θ.

A2: *For all $\theta \in \Theta$ there exists a constant $t > 0$ such that X_t^θ admits a density $p_t^\theta(x, y)$ with respect to the Lebesgue measure on $\mathbb{R}$; bounded in $y \in \mathbb{R}$ and in $x \in K$ for every compact $K \subset \mathbb{R}$. Moreover, for every $x \in \mathbb{R}$ and every open ball $U \in \mathbb{R}$, there exists a point $z = z(x, U) \in supp(F)$ such that $\gamma(x)z \in U$.*

Assumption 2 ensures, together with the Assumption 3 below, the existence of unique invariant distribution π^θ, as well as the ergodicity of the process X^θ.

A3 (Ergodicity): *(i) For all $q > 0$, $\int_{|z|>1} |z|^q F(z)dz < \infty$.*
(ii) For all $\mu \in \Pi$ there exists $C > 0$ such that $xb(\mu, x) \leq -C|x|^2$, if $|x| \to \infty$.
(iii) $|\gamma(x)|/|x| \to 0$ as $|x| \to \infty$.
(iv) For all $\sigma \in \Sigma$ we have $|a(\sigma, x)|/|x| \to 0$ as $|x| \to \infty$.

(v) $\forall\theta \in \Theta$, $\forall q > 0$ *we have* $\mathbb{E}|X_0^\theta|^q < \infty$.

A4 (Jumps): *1. The jump coefficient* γ *is bounded from below, that is* $\inf_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$.
2. The Lévy measure F *is absolutely continuous with respect to the Lebesgue measure and we denote* $F(z) = \frac{F(dz)}{dz}$.
3. F *is such that* $F(z) = \lambda F_0(z)$ *and* $\int_{\mathbb{R}} F_0(z)dz = 1$.

Assumption 4.1 is useful to compare size of jumps of X and L. The Assumption 5 ensures the existence of the contrast function we will define in next section.

A5 (Non-degeneracy): *There exists some* $c > 0$, *such that* $\inf_{x,\sigma} a^2(\sigma, x) \geq c > 0$.

From now on we denote the true parameter value by θ_0, an interior point of the parameter space Θ that we want to estimate. We shorten X for X^{θ_0}.

We will use some moment inequalities for jump diffusions, gathered in the following lemma that follows from Theorem 66 of [80] and Proposition 3.1 in [88].

Lemma 12. *Let* X *satisfies Assumptions 1-4. Let* $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$ *and let* $\mathcal{F}_s := \sigma\left\{(W_u)_{0<u\leq s}, (L_u)_{0<u\leq s}, X_0\right\}$.
Then, for all $t > s > 0$,
1) for all $p \geq 2$, $\mathbb{E}[|X_t - X_s|^p]^{\frac{1}{p}} \leq c|t - s|^{\frac{1}{p}}$,
2) for all $p \geq 2$, $p \in \mathbb{N}$, $\mathbb{E}[|X_t - X_s|^p|\mathcal{F}_s] \leq c|t - s|(1 + |X_s|^p)$,
3) for all $p \geq 2$, $p \in \mathbb{N}$, $\sup_{h\in[0,1]} \mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] \leq c(1 + |X_s|^p)$.

An important role is played by ergodic properties of solution of equation (4.3) The following Lemma states that Assumptions $1-4$ are sufficient for the existence of an invariant measure π^θ such that an ergodic theorem holds and moments of all order exist.

Lemma 13. *Under assumptions 1 to 4, for all* $\theta \in \Theta$, X^θ *admits a unique invariant distribution* π^θ *and the ergodic theorem holds:*
1)For every measurable function $g : \mathbb{R} \to \mathbb{R}$ *s. t.* $\pi^\theta(g) < \infty$, *we have a.s.* $\lim_{t\to\infty} \frac{1}{t}\int_0^t g(X_s^\theta)ds = \pi^\theta(g)$.
2)For all $q > 0$, $\pi^\theta(|x|^q) < \infty$.
3) For all $q > 0$, $\sup_{t\geq 0} \mathbb{E}[|X_t^\theta|^q] < \infty$.
4) Moreover, $\lim_{t\to\infty} \frac{1}{t}\int_0^t \mathbb{E}[|X_s^\theta|^q]ds = \pi^\theta(|x|^q)$.

A proof is in [38] (Section 8 of Supplement), it relies mainly on results of [67].

A6 (Identifiability): *For all* μ_1, μ_2 *in* Π, $\mu_1 = \mu_2$ *if and only if* $b(\mu_1, x) = b(\mu_2, x)$ *for almost all* x. *Moreover,* $\forall\sigma_1, \sigma_2$ *in* Σ, $\sigma_1 = \sigma_2$ *if and only if* $a(\sigma_1, x) = a(\sigma_2, x)$ *for almost all* x.

A7: *1. The derivatives* $\frac{\partial^{k_1+k_2}b}{\partial x^{k_1}\partial\theta^{k_2}}$, *with* $k_1 + k_2 \leq 4$ *and* $k_2 \leq 3$, *exist and they are bounded if* $k_1 \geq 1$. *If* $k_1 = 0$, *for each* $k_2 \leq 3$ *they have polynomial growth.*
2. The derivatives $\frac{\partial^{k_1+k_2}a}{\partial x^{k_1}\partial\theta^{k_2}}$, *with* $k_1 + k_2 \leq 4$ *and* $k_2 \leq 3$, *exist and they are bounded if* $k_1 \geq 1$. *If* $k_1 = 0$, *for each* $k_2 \leq 3$ *they have polynomial growth.*

3. The derivatives $\gamma^{(k)}(x)$ exist and they are bounded for each $1 \leq k \leq 4$.

A8: *Let B be* $\begin{pmatrix} -2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx) & 0 \\ 0 & 4\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx) \end{pmatrix}$, *then $det(B) \neq 0$.*

2.3 Construction of the estimator and main results

Now we present a contrast function for estimating parameters.

2.3.1 Construction of contrast function.

Suppose that we observe a finite sample $X_{t_0}, ..., X_{t_n}$ with $0 = t_0 \leq t_1 \leq ... \leq t_n =: T_n$, where X is the solution to (4.3) with $\theta = \theta_0$. Every observation time point depends also on n, but to simplify the notation we suppress this index. We will be working in a high-frequency setting, i.e. $\Delta_n := \sup_{i=0,...,n-1} \Delta_{n,i} \longrightarrow 0$ for $n \to \infty$, with $\Delta_{n,i} := (t_{i+1} - t_i)$. We assume that $\lim_{n\to\infty} T_n = \infty$.
In the sequel we will always suppose that the following assumption on the step discretization holds true.
$A_{\textbf{Step}}$: there exist two constants c_1, c_2 such that $c_2 < \frac{\Delta_n}{\Delta_{min}} < c_1$, where we have denoted Δ_{min} as $\min_{i=0,...,n-1} \Delta_{n,i}$.
We introduce a jump filtered version of the gaussian quasi-likelihood. In our setting, the observed data are discrete, hence we have to decide whether jumps occur or not in an interval from only the increment of our process, although that is a stochastic decision which may sometimes include some misjudgments. This criterion should be chosen depending on n, and increase the accuracy of judgements as n tends to infinity. This leads to the following contrast function:

Definition 13. *For $\beta \in (\frac{1}{4}, \frac{1}{2})$ we define the contrast function $U_n(\mu, \sigma)$ as*

$$U_n(\mu,\sigma) := \sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2}{m_2(\mu,\sigma,X_{t_i})} + log(\frac{m_2(\mu,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}, \tag{2.4}$$

with

$$m(\mu,\sigma,x) := \frac{\mathbb{E}[X_{t_{i+1}}^\theta \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}; \tag{2.5}$$

$$m_2(\mu,\sigma,x) := \frac{\mathbb{E}[(X_{t_{i+1}}^\theta - m(\mu,\sigma,X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|X_{t_i}^\theta = x]} \tag{2.6}$$

and

$$\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) = \varphi(\frac{X_{t_{i+1}} - X_{t_i}}{\Delta_{n,i}^\beta}).$$

The function φ is a smooth version of the indicator function, such that $\varphi(\zeta) = 0$ for each ζ, with $|\zeta| \geq 2$ and $\varphi(\zeta) = 1$ for each ζ, with $|\zeta| \leq 1$.
The last indicator aims to avoid the possibility that $|X_{t_i}|$ is big. The constant k is

positive and it will be chosen later, related to the development of both m and m_2. Moreover we remark that m and m_2 depend also on t_i and t_{i+1}. By the homogeneity of the equation they actually depend on the difference $t_{i+1} - t_i$ but we omit such a dependence in the notation of the two functions here above to make the reading easier.

We define an estimator $\hat{\theta}_n$ of θ_0 as

$$\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n) \in \arg \min_{(\mu,\sigma)\in\Theta} U_n(\mu, \sigma). \tag{2.7}$$

The idea is to use the size of $X_{t_{i+1}} - X_{t_i}$ in order to judge the existence of a jump in an interval $[t_i, t_{i+1})$. The increment of X with continuous transition could hardly exceed the threshold $\Delta_{n,i}^{\beta}$, therefore we can judge a jump occurred if $|X_{t_{i+1}} - X_{t_i}| > \Delta_{n,i}^{\beta}$.

The value β has to be chosen carefully. For instance if β is too large, and therefore $\Delta_{n,i}^{\beta}$ is too small, the probability of getting the increment $\Delta_{n,i}^{\beta}$ by the continuous diffusion can not be ignored, on the other hand, if β is too small, and therefore $\Delta_{n,i}^{\beta}$ is too large, we cannot ignore the probability of getting an increment less than $\Delta_{n,i}^{\beta}$ when a jump occurs in an interval. In the definition of the contrast function we have taken $\beta > \frac{1}{4}$ because, in Lemma 14 and Proposition 14 below (and so, as a consequence, in the majority of the theorems of this chapter), such a technical condition on β is required.

We observe that, in general, there is no closed expression for m and m_2, hence the contrast is not explicit. However, it is proved in Chapter 1 before an explicit development of m in the case where the intensity is finite and in this chapter we provide as well an explicit development of m_2 that lead us to an explicit version of our contrast function.

2.3.2 Main results

Before stating our main results, we recall some further notations and assumptions. For $\delta \geq 0$, we will denote $R(\theta, \Delta_{n,i}^{\delta}, x)$ for any function $R(\theta, \Delta_{n,i}^{\delta}, x) = R_{i,n}(\theta, x)$, where $R_{i,n} : \Theta \times \mathbb{R} \longrightarrow \mathbb{R}$, $(\theta, x) \mapsto R_{i,n}(\theta, x)$ is such that

$$\exists c > 0 \qquad |R_{i,n}(\theta, x)| \leq c(1 + |x|^c)\Delta_{n,i}^{\delta} \tag{2.8}$$

uniformly in θ and with c independent of i, n.

The functions R represent the term of rest and have the following useful property, consequence of the just given definition:

$$R(\theta, \Delta_{n,i}^{\delta}, x) = \Delta_{n,i}^{\delta} R(\theta, \Delta_{n,i}^{0}, x). \tag{2.9}$$

We point out that it does not involve the linearity of R, since the functions R on the left and on the right side are not necessarily the same but only two functions on which the control (3.23) holds with $\Delta_{n,i}^{\delta}$ and $\Delta_{n,i}^{0}$, respectively.

In the sequel, we will need a development for the function m_2. We will assume that such a development exists, as stated in the next assumption:

Ad: There exist three functions $r(\mu, \sigma, x)$, $r(x)$, $R(\theta, 1, x)$ and $\delta_1, \delta_2 > 0$ and $k_0 > 0$ such that, for $|x| \leq \Delta_{n,i}^{-k_0}$,

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma)(1 + \Delta_{n,i} r(\mu, \sigma, x)) + \Delta_{n,i}^{1+\delta_1} r(x) + \Delta_{n,i}^{2+\delta_2} R(\theta, 1, x), \tag{2.10}$$

where $r(\mu, \sigma, x)$ and $r(x)$ are particular functions $R(\theta, 1, x)$, that turns out from the development of m_2, and the function $r(x)$ does not depend on θ. Moreover, the order of such functions does not change by deriving them with respect to both the parameters, that is for $\vartheta = \mu$ and $\vartheta = \sigma$, $|\partial_\vartheta r(\mu, \sigma, x)| \leq c(1 + |x|^c)$ and $|\partial_\vartheta R(\theta, 1, x)| \leq c(1 + |x|^c)$.

Assumption Ad is not restrictive. Examples of frameworks in which Ad holds are introduced in Propositions 10 and 12, that will be stated in the next section and proven in the appendix. Let us stress that it is crucial for the proof of the consistency of the estimator that the second main term of the expansion (2.10), $\Delta_{n,i}^{1+\delta_1} r(x)$, does not depend on the parameter θ.

The following theorems give a general consistency result and the asymptotic normality of the estimator $\hat{\theta}_n$.

Theorem 8. *(Consistency) Suppose that Assumptions 1 to 7, A_{Step} and Ad hold. Then the estimator $\hat{\theta}_n$ is consistent in probability:*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0, \qquad n \to \infty.$$

Theorem 9. *(Asymptotic normality) Suppose that Assumptions 1 to 8, A_{Step} and Ad hold. Then*

$$(\sqrt{T_n}(\hat{\mu}_n - \mu_0), \sqrt{n}(\hat{\sigma}_n - \sigma_0)) \xrightarrow{\mathcal{L}} N(0, K) \quad \text{for } n \to \infty,$$

where $K = \begin{pmatrix} (\int_{\mathbb{R}} (\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} & 0 \\ 0 & 2(\int_{\mathbb{R}} (\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2 \pi(dx))^{-1} \end{pmatrix}.$

The proof of our main results will be presented in Section 2.7.

2.4 Practical implementation of the contrast method

In order to use in practice the contrast function (2.4), one need to know the values of the quantities $m(\mu, \sigma, X_{t_i})$ and $m_2(\mu, \sigma, X_{t_i})$. Even if in most cases, it seems impossible to find an explicit expression for them, explicit or numerical approximations of this functions seem available in many situations.

2.4.1 Approximation of the contrast function

Let us assume that one has at disposal an approximation of the functions $m(\mu, \sigma, x)$ and $m_2(\mu, \sigma, x)$, denoted by $\tilde{m}(\mu, \sigma, x)$ and $\tilde{m}_2(\mu, \sigma, x)$ which satisfy, for $|x| \leq h^{-k_0}$, the following assumptions.

$A\rho$ **:**

1. $|\tilde{m}(\mu, \sigma, x) - m(\mu, \sigma, x)| \leq R(\theta, h^{\rho_1}, x)$, $|\tilde{m}_2(\mu, \sigma, x) - m_2(\mu, \sigma, x)| \leq R(\theta, h^{\rho_2}, x)$, where the constants $\rho_1 > 1$ and $\rho_2 > 1$ assess the quality of the approximation.

2. $|\partial_\mu^i \tilde{m}(\mu, \sigma, x) - \partial_\mu^i m(\mu, \sigma, x)| + |\partial_\sigma^i \tilde{m}_2(\mu, \sigma, x) - \partial_\sigma^i m_2(\mu, \sigma, x)| \leq R(\theta, h^{1+\epsilon}, x)$, for $i = 1, 2$,for all $|x| \leq h^{-k_0}$ and where $\epsilon > 0$.

3. The bounds on the derivatives of m and m_2 gathered in Propositions 16, 17 and 18 hold true for $\tilde{m}$ and $\tilde{m}_2$ replacing m and m_2.

We have to act on the derivatives of the two approximated functions $\tilde{m}$ and $\tilde{m}_2$ as we do on m and m_2. That's the reason why we need to add the third technical assumption here above, which assure we can move from the derivatives of the real functions to the approximated ones committing an error which is negligible. Now, we consider $\widetilde{\theta}_n$ the estimator obtained from minimization of the contrast function (2.4) where one has replaced the functions $m(\mu, \sigma, X_{t_i})$ and $m_2(\mu, \sigma, X_{t_i})$ by their approximations $\tilde{m}(\mu, \sigma, x)$ and $\tilde{m}_2(\mu, \sigma, x)$. Then, the result of Theorem 9 can be extended as follows.

Proposition 9. *Suppose that Assumptions 1 to 8, Ad, A_{Step} and $A\rho$ hold, with $0 < k < k_0$, and that $\sqrt{n}\Delta_n^{\rho_1 - 1/2} \to 0$ and $\sqrt{n}\Delta_n^{\rho_2 - 1} \to 0$ as $n \to \infty$.*
Then, the estimator $\widetilde{\theta}_n := (\tilde{\mu}_n, \tilde{\sigma}_n)$ is asymptotically normal:

$$(\sqrt{T_n}(\tilde{\mu}_n - \mu_0), \sqrt{n}(\tilde{\sigma}_n - \sigma_0)) \xrightarrow{\mathcal{L}} N(0, K) \quad \text{for } n \to \infty,$$

where K is the matrix defined in Theorem 9.

Proposition 9 will be proven in Section 2.7.4.

We give below several examples of approximations of $m_2(\mu, \sigma, X_{t_i})$ which can be used, together with the approximations of $m(\mu, \sigma, X_{t_i})$ given in Proposition 2 of Chapter 1, to construct an explicit contrast function.

2.4.2 Development of $m_2(\mu, \sigma, x)$.

We provide two kinds of expansion for the function m_2. First, we prove high order expansions that involve only the continuous part of the generator of the process and necessitate the choice of oscillating functions φ. Second, we find an expansion up to order Δ_n^2 for any function φ, and, in particular, show the validity of the condition Ad in a general setting. For completeness, we recall also the expansions of the function m found in the previous chapter.

2.4.2.1 Arbitrarily high expansion with oscillating truncation functions.

We show we can write an explicit development for the function m_2, as we did for the function m in Proposition 2 of Chapter 1 before, taking a particular oscillating function φ. In this way, it is therefore possible to make the contrast explicit with approximation at any order. We define $A_{K_1}^{(k)}(x) := \bar{A}_c^k(h_1)(x)$ and $A_{K_2}^{(k)}(x) := \bar{A}_c^k(h_2)(x)$, where $\bar{A}_c(f) := \bar{b}f' + \frac{1}{2}a^2 f''$, with $\bar{b}(\mu, y) = b(\mu, y) - \int_{\mathbb{R}} \gamma(y) z F(z) dz$; K_1 and K_2 we have written here above stand for "Kessler", based on the fact that the development we find is the same obtained in [51] in the case without jumps by the iteration of the continuous generator $\bar{A}_c$. The functions who appear in the definition of $A_{K_1}^{(k)}$ and $A_{K_2}^{(k)}$ are the following: $h_1(y) := (y - x)$, $h_2(y) = y^2$.
To get Proposition 10 we need to add the following assumption:

Af: We assume that $x \mapsto a(x, \sigma)$, $x \mapsto b(x, \mu)$ and $x \mapsto \gamma(x)$ are $\mathcal{C}^\infty$ functions, they have at most uniform in μ and σ polynomial growth as well as their derivatives.

Proposition 10. *Assume that Assumptions 1-4 and Af hold and let φ be a $\mathcal{C}^\infty$ function that has compact support and such that $\varphi \equiv 1$ on $[-1,1]$ and $\forall k \in \{0, ..., M\}$, $\int_{\mathbb{R}} x^k \varphi(x)dx = 0$ for $M \geq 0$. Moreover we suppose that the Lévy density F is $\mathcal{C}^\infty$. Then, for $|x| \leq \Delta_{n,i}^{-k_0}$ with some $k_0 > 0$,*

$$m_2(\mu, \sigma, x) = \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_2}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} - (x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!})^2 + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x). \tag{2.11}$$

Moreover, for $\vartheta = \mu$ or $\vartheta = \sigma$

$$|\partial_\vartheta R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)| \leq R(\theta, \Delta_{n,i}^{\beta(M+2)}, x). \tag{2.12}$$

It is proved below Proposition 2 in the chapter before that a function φ which satisfies the assumptions here above exists: it is possible to build it through ψ, a function with compact support, $\mathcal{C}^\infty$ and such that $\psi|_{[-1,1]}(x) = \frac{x^M}{M!}$. It is enough to define $\varphi(x) := \frac{\partial^M}{\partial x^M} \psi(x)$ to get $\varphi \equiv 1$ on $[-1,1]$; φ is $\mathcal{C}^\infty$, with compact support and such that for each $l \in \{0, ...M\}$, using the integration by parts, $\int_{\mathbb{R}} x^l \varphi(x)dx = 0$.
It is thanks to such a choice of an oscillating function φ that the contribution of the discontinuous part of the generator disappears and we get the same development found in the continuous case, in Kessler [51], due only to the continuous generator. In the situation where (2.11) holds true with $\lfloor \beta(M+2) \rfloor > 2$, we get a development for m_2 as in *Ad* for $r(x)$ identically 0 and $r(\mu, \sigma, x)$ being an explicit function.

For completeness, let us recall that under the same assumptions as in Proposition 10, we have the following expansion for m (see Proposition 2 in Chapter 1):

$$m(\mu, \sigma, x) = x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x), \text{ for } |x| \leq \Delta_{n,i}^{-k_0}, \text{ with } k_0 > 0. \tag{2.13}$$

2.4.2.2 Second order expansion with general truncation functions.

Another situation in which *Ad* holds is gathered in the following proposition, that will be still proven in the appendix:

Proposition 11. *Suppose that Assumptions A1 -A5 and A7 hold, that $\beta \in (\frac{1}{4}, \frac{1}{2})$ and that the Lévy density F is $\mathcal{C}^1$. Then there exists $k_0 > 0$ such that, for $|x| \leq \Delta_{n,i}^{-k_0}$,*

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)} F(0) \int_{\mathbb{R}} v^2 \varphi(v)dv + \Delta_{n,i}^2 (3\bar{b}^2(x, \mu) + h_2(x, \theta)) + \Delta_{n,i}^{(1+4\beta)\wedge(2+\beta)\wedge(3-2\beta)} R(\theta, 1, x); \tag{2.14}$$

where $h_2 = \frac{1}{2} a^2 (a')^2 + \frac{1}{2} a^3 a'' + a^2 \bar{b}' + a a' \bar{b} + \bar{b}^2$. Moreover, for both $\vartheta = \mu$ and $\vartheta = \sigma$, $\partial_\vartheta R(\theta, 1, x)$ is still a $R(\theta, 1, x)$ function.

We observe that, defining $r(x) := F(0) \int_{\mathbb{R}} v^2 \varphi(v) dv$ and $r(\mu, \sigma, x) := \frac{3\bar{b}^2(x,\mu) + h_2(x,\theta)}{a^2(x,\sigma)}$, we get a development as in Ad for $\delta_1 := 3\beta$.
We observe that if $\int_{\mathbb{R}} v^2 \varphi(v) dv = 0$, we fall back in development of Proposition 10 up to order 2. We therefore see that the choice of an oscillating truncated function φ is necessary in order to remove the jump contribution.

It is worth noting here that biggest term after the main one is due to the jump part and do not depend on the parameters μ and σ. We'll see in the sequel that is necessary, in order to prove the consistency of $\hat{\mu}_n$, that this contribution does not depend on the drift parameter. Considering indeed the difference of the contrast computed for two different values of the drift parameter, its presence results irrelevant.

We remark that the term with order $1 + 4\beta$ is negligible compared to the order 2 terms since in our setting β is assumed to be bigger than $\frac{1}{4}$.

In Proposition 11 before F is required to be $\mathcal{C}^1$, such assumption is no longer needed in following more general proposition.

Proposition 12. *Suppose that Assumptions A1 -A5 and A7 hold. Then there exists $k_0 > 0$ such that, for $|x| \le \Delta_{n,i}^{-k_0}$,*

$$m_2(\mu, \sigma, x) = \Delta_{n,i} a^2(x, \sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)} \int_{\mathbb{R}} u^2 \varphi(u) F(u \frac{\Delta_{n,i}^{\beta}}{\gamma(x)}) du + \Delta_{n,i}^2 (3\bar{b}^2(x, \mu) + h_2(x, \theta)) +$$

$$+ \frac{\Delta_{n,i}^{2+\beta} a^2(x, \sigma)}{2\gamma(x)} \int_{\mathbb{R}} (u\varphi'(u) + u^2 \varphi''(u)) F(\frac{u \Delta_{n,i}^{\beta}}{\gamma(x)}) du + \Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)} R(\theta, 1, x), \quad (2.15)$$

where $h_2 = \frac{1}{2} a^2 (a')^2 + \frac{1}{2} a^3 a'' + a^2 \bar{b}' + a a' \bar{b} + \bar{b}^2$.
Moreover, for both $\vartheta = \mu$ and $\vartheta = \sigma$, $\partial_\vartheta R(\theta, 1, x)$ is still a $R(\theta, 1, x)$ function.

We see that the contributions of the jumps depend on the density F which argument in the integral depend on $\Delta_{n,i}$. If we choose a particular density function F which is null in the neighborhood of 0 the contribution of the jumps disappears and, in this case, we fall on the development for m_2 found by Kessler in the case without jumps ([51]), up to order $\Delta_{n,i}^2$.

The expansion (2.15) looks cumbersome, however all terms are necessary to get an expansion with a remainder term of explicit order strictly greater than 2, and valid for any finite intensity F. In the particular case where F is $\mathcal{C}^1$, the first three terms in the expansion give the the main terms of the expansions (2.14), while the last integral term, with order $\Delta_{n,i}^{2+\beta}$, is clearly a rest term. However, in the situation where F may be unbounded near 0, with $\int F(z) dz < \infty$, the last integral term is only seen to be negligible versus $\Delta_{n,i}^2$. Hence, this last integral term may be non negligible compared to the rest term and is needed in the expansion.

Finally, we recall the expansion of m valid under the Assumptions A1-A4 (see Theorem 2 in Chapter 1),

$$m(\mu, \sigma, x) = x + \Delta_{n,i} \bar{b}(x, \mu) + \frac{\Delta_{n,i}^{1+2\beta}}{\gamma(x)} \int_{\mathbb{R}} u \varphi(u) F(\frac{u \Delta_{n,i}^{\beta}}{\gamma(x)}) du + R(\theta, \Delta_{n,i}^{2-2\beta}, x). \quad (2.16)$$

Developments of m_2 given in Propositions 10 - 12, together with the developments of m given in (2.13), (2.16) will be useful for the applications, as illustrated in the following section.

2.5 Simulation study

Let us consider the model

$$X_t = X_0 + \int_0^t (\theta_1 X_s + \theta_2) ds + \sigma W_t + \gamma \int_0^t \int_{\mathbb{R}\setminus\{0\}} z\tilde{\mu}(ds, dz), \tag{2.17}$$

where the compensator of the jump measure is $\overline{\mu}(ds, dz) = \lambda F_0(z) ds dz$ for F_0 the probability density of the law $\mathcal{N}(\mu_J, \sigma_J^2)$ with $\mu_J \in \mathbb{R}$, $\sigma_J > 0$, $\sigma > 0$, $\theta_1 < 0$, $\theta_2 \in \mathbb{R}$, $\gamma \geq 0$, $\lambda \geq 0$.
We want to explore approximations for m and m_2 which make us able to find en explicit version of the contrast function. According to Proposition 10 (respectively Proposition 2 in Chapter 1), we know that using sufficiently oscillating truncation functions the expansion of m_2 (respectively m) is the same as Kessler's expansion for the continuous part of the SDE.
Since the Kessler's expansion approximates the first conditional moments of $\bar{X}_t = \bar{X}_0 + \int_0^t (\theta_1 \bar{X}_s + \theta_2 - \gamma\lambda\mu_J) ds + \sigma W_t$ (see (2.1)), which is the continuous part of (2.17) and which is explicit due to the linearity of the model, we decide to use directly the expression of the conditional moment and set

$$\widetilde{m}(\theta_1, \theta_2, x) = (x + \frac{\theta_2}{\theta_1} - \frac{\gamma\lambda\mu_J}{\theta_1}) e^{\theta_1 \Delta_{n,i}} + \frac{\gamma\lambda\mu_J - \theta_2}{\theta_1}; \tag{2.18}$$

while the approximation of $m_2(\mu, \sigma, x)$ is

$$\widetilde{m}_2(\theta_1, \sigma, x) = \frac{\sigma^2}{2\theta_1}(e^{2\theta_1 \Delta_{n,i}} - 1). \tag{2.19}$$

We want to compare the estimator $\tilde{\theta}_n$ we get by the minimization of the contrast function obtained by the Kessler exact correction of the bias in which we use the approximations (2.18) and (2.19) for m and m_2 with the estimator based on the Euler scheme approximation:

$$\widetilde{m}^E(\theta_1, \theta_2, x) = x + (\theta_1 x + \theta_2 - \lambda\gamma\mu_j)\Delta_{n,i}, \qquad \widetilde{m}_2^E(\sigma, x) = \sigma^2 \Delta_{n,i}. \tag{2.20}$$

According with Proposition 10, we build oscillating truncation functions. To do it, we choose $\psi : \mathbb{R} \to [0, 1]$ a $\mathcal{C}^\infty$ symmetric function with support on $[-2, 2]$ such that $\psi(x) = 1$ for $|x| \leq 1$. We let, for $d > 1$, $\varphi(x) := (d\psi(x) - \psi(x/d))/(d-1)$, which is a function equal to 1 on $[-1, 1]$, vanishing on $[-d, d]^c$, such that $\int_{\mathbb{R}} \varphi(x) dx = 0$ and that $\int_{\mathbb{R}} x\varphi(x) dx = 0$. In the contrast function we use the truncation function $x \mapsto \varphi_{c\Delta_{i,n}^\beta}(x)$, where $c > 0$ is some constant. In the theoretical parts of the paper we have chosen $c = 1$, as this constant does not matter asymptotically, however for the practical usage of the method on finite sample the choice of the constant c is an important issue.
For numerical simulations we choose $T = 100$, $n = 5000$, $\lambda = 1$, $\gamma = 1$, $\theta_1 = -1$, $\theta_2 = 2$, $\sigma = 0.5$, $c = 2$ and $d = 2$. We estimate the bias of the estimators using a Monte Carlo method based on 500 replications.
First, we consider β as big as possible, fixing it equal to 0.49; then we will take $\beta = 0.3$; in both cases the jumps size has common law $\mathcal{N}(4, 0.25)$.
In Tables 2.1 and 2.2 we illustrate, for $\beta = 0.49$ and $\beta = 0.3$ respectively, the behaviour of the two estimators $\tilde{\theta}_n^{\text{Euler}}$ and $\tilde{\theta}_n$. The first is obtained using Euler

	Mean for $\theta_1 = -1$	Mean for $\theta_2 = 2$	Mean for $\sigma = 0.5$
$\tilde{\theta}_n^{\text{Euler}}$	-0.98576674	2.03648076	0.48215837
$\widetilde{\theta}_n$	-1.00242758	1.99964172	0.49684659

Table 2.1 – $\beta = 0.49$

	Mean for $\theta_1 = -1$	Mean for $\theta_2 = 2$	Mean for $\sigma = 0.5$
$\tilde{\theta}_n^{\text{Euler}}$	-0.95638861	2.45094605	1.32435848
$\widetilde{\theta}_n$	-1.00262437	2.00372846	0.49855887

Table 2.2 – $\beta = 0.3$

scheme to approximate the two unknown functions m and m_2 through $\widetilde{m}^E(\theta_1, \theta_2, x)$ and $\widetilde{m}_2^E(\sigma, x)$, while we obtain $\widetilde{\theta}_n$ through the developments (2.18) and (2.19) for m and m_2, based on the Kessler exact correction of the bias.
We see from Table 2.1 that when β is big the noise, already small in the first line, is reduced for the joint estimation of all three parameters through Kessler exact approximation. We make β decrease in Table 2.2, now the estimations of the three parameters through Euler approximation doesn't work well anymore (it results evident, in particular, for the estimation of σ). Looking at the second line of Table 2.2, instead, we see that the noise is visibly reduced for the estimation of all the three parameters and so that the estimator $\widetilde{\theta}_n$ performs well for any β.
Regarding the value of β, it might not seem like a very natural choice to take it small, since in this way the threshold is big and so we can not ignore the probability to have a jump in the interval considered even if the increment is less than Δ_n^β. However, the choice of β can be traced back to the choice of c and, because of the well performance of our estimator for no matter which β considered, we can deduce $\tilde{\theta}_n$ is less sensitive than $\tilde{\theta}_n^{\text{Euler}}$ to the threshold issue.

In the previous case we have used Kessler approximation to remove the bias deriving from the continuous part at any order and, according with Proposition 10, we have taken an oscillating truncated function φ for which the initial contributions of the discontinuous part of the generator disappear. Now we still consider (2.17) in which F_0 is still the probability density of the law $\mathcal{N}(\mu_j, \sigma_J^2)$, but we use the low order expansion available for any truncation function.
According to Theorem 2 of Chapter 1 (see also (2.16)) we have, considering a threshold level which is $c\Delta_{n,i}^\beta$, the following development for m:

$$m(\theta_1, \theta_2, \sigma, x) = x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^\beta}{\gamma})du + R(\theta, \Delta_{n,i}^{2-2\beta}, x);$$

which leads us to the approximation

$$\widetilde{m}(\theta_1, \theta_2, x) = x + \Delta_{n,i}\bar{b}(x, \theta_1, \theta_2) + \frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^\beta}{\gamma})du,$$

for $\bar{b}(x,\theta_1,\theta_2) = (\theta_1 x + \theta_2) - \gamma\lambda\mu_j$. It follows $|m(\theta_1,\theta_2,\sigma,x) - \widetilde{m}(\theta_1,\theta_2,x)| \le R(\theta,\Delta_{n,i}^{2-2\beta},x)$.
Concerning the approximation of m_2, from its development gathered in Proposition 12 we define

$$\widetilde{m}_2(\sigma) := \Delta_{n,i}\sigma^2 + \frac{\Delta_{n,i}^{1+3\beta}c^3}{\gamma}\int_{\mathbb{R}} u^2\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du,$$

which is such that $|m_2(\theta_1,\theta_2,\sigma,x) - \widetilde{m}_2(\sigma)| \le R(\theta,\Delta_{n,i}^2,x)$.
We estimate jointly the parameter $\theta = (\theta_1,\theta_2,\sigma)$ by minimization of the contrast function

$$U_n(\theta) = \sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))^2}{\widetilde{m}_2(\sigma,X_{t_i})} + \log(\frac{\widetilde{m}_2(\theta_1,\sigma,X_{t_i})}{\Delta_{n,i}})]\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}), \tag{2.21}$$

where $c > 0$ will be specified later.
We compute the derivatives of the contrast function with respect to the three parameters:

$$\partial_{\theta_1}U_n(\theta) = \sum_{i=0}^{n-1}\frac{2(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))X_{t_i}}{\widetilde{m}_2(\sigma)}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) =$$

$$= \frac{2}{\widetilde{m}_2(\sigma)}\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)X_{t_i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}),$$

where we have denoted as J_1^i the term in the development of $\widetilde{m}$ turning up from the presence of jumps, which is $\frac{c^2\Delta_{n,i}^{1+2\beta}}{\gamma}\int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_{n,i}^{\beta}}{\gamma})du$.
We want $\partial_{\theta_1}U_n(\theta) = 0$, it leads us to the definition of the following estimator:

$$\tilde{\theta}_{1,n} := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)X_{t_i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}X_{t_i}^2\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}.$$

In the same way

$$\partial_{\theta_2}U_n(\theta) = \sum_{i=0}^{n-1}\frac{2(X_{t_{i+1}} - \widetilde{m}(\theta_1,\theta_2,X_{t_i}))}{\tilde{m}_2(\sigma)}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) =$$

$$= \frac{2}{\widetilde{m}_2(\sigma)}\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i} - \Delta_{n,i}\theta_2 + \Delta_{n,i}\gamma\lambda\mu_j - J_1^i)\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}).$$

Since we want $\partial_{\theta_2}U_n(\theta) = 0$, we define $\tilde{\theta}_{2,n}$ in the following way:

$$\tilde{\theta}_{2,n} := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - X_{t_i} - \Delta_{n,i}\theta_1 X_{t_i})\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} + \tag{2.22}$$

$$+\gamma\lambda\mu_j - \frac{\sum_{i=0}^{n-1}J_1^i\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})} =$$

$$= \tilde{\theta}_n^{\text{Euler}} - \frac{\sum_{i=0}^{n-1}J_1^i\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}\varphi_{c\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i})};$$

we can see $\tilde{\theta}_{2,n}$ as a corrected version of the estimator $\tilde{\theta}_n^{\text{Euler}}$ that would result considering the Euler scheme approximation for the function m, as in (2.20). We observe moreover that, considering a uniform discretization step, the last term in (2.22) becomes simply $\frac{J_1}{\Delta_n} := \frac{c^2\Delta_n^{2\beta}}{\gamma} \int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_n^\beta}{\gamma})du$.
Computing also the derivative of the contrast function with respect to σ^2, we have

$$\partial_{\sigma^2}U_n(\theta) = \sum_{i=0}^{n-1}[\frac{-(X_{t_{i+1}} - \widetilde{m}(\theta_1, \theta_2, X_{t_i}))^2\partial_{\sigma^2}\widetilde{m}_2(\sigma) + \widetilde{m}_2(\sigma)\partial_{\sigma^2}\widetilde{m}_2(\sigma)}{\widetilde{m}_2^2(\sigma)}]\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}),$$

which is equal to zero if and only if

$$\sum_{i=0}^{n-1}[-(X_{t_{i+1}} - \widetilde{m}(\theta_1, \theta_2, X_{t_i}))^2\Delta_{n,i} + \Delta_{n,i}(\Delta_{n,i}\sigma^2 + J_2^i)]\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) = 0,$$

for $J_2^i := \frac{\Delta_{n,i}^{1+3\beta}c^3}{\gamma} \int_{\mathbb{R}} u^2\varphi(u)F(\frac{uc\Delta_{n,i}^\beta}{\gamma})du$, which is the part deriving from the jumps in the development of $\widetilde{m}_2(\sigma)$. It drives us to the estimator

$$\tilde{\sigma}_n^2 := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - \tilde{m}(\theta_1, \theta_2, X_{t_i}))^2\Delta_{n,i}\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}^2\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})} +$$

$$-\frac{\sum_{i=0}^{n-1}\Delta_{n,i}J_2^i\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})}{\sum_{i=0}^{n-1}\Delta_{n,i}^2\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})}.$$

Considering an uniform discretization step it is

$$\tilde{\sigma}_n^2 := \frac{\sum_{i=0}^{n-1}(X_{t_{i+1}} - \tilde{m}(\theta_1, \theta_2, X_{t_i}))^2\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})}{\Delta_n\sum_{i=0}^{n-1}\varphi_{c\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})} - \frac{J_2}{\Delta_n}.$$

Again, it can be seen as a corrected version of $\tilde{\sigma}_n^{2,\text{Euler}}$, the estimator that would have resulted considering the approximation of the functions m and m_2 as defined in (2.20). In such a case, not only we wouldn't have seen the contribution of the jumps appearing in the second term here above, but also in the first term we should have replaced $\tilde{m}$ with its Euler approximation.
To illustrate the estimation method, we focus on the estimation of the parameters θ_2 and σ^2 only.
For numerical simulations we choose the parameters as we did above Table 2.1.
The real values of the parameters we want to estimate are $\theta_2 = 2$ and $\sigma = 0.5$; we estimate the bias of the estimators using a Monte Carlo method based on 500 replications.

We see that there is not a big difference in the estimation of θ_2 with and without the correction term while, regarding σ^2, the estimator found through our approximated function $\tilde{m}$ and $\tilde{m}_2$ performs better than the estimator we got through Euler scheme, especially for β smaller.
The two estimators of θ_2 differs in fact only for the contribution of the jumps $\frac{J_1}{\Delta_n} = \frac{c^2\Delta_n^{2\beta}}{\gamma} \int_{\mathbb{R}} u\varphi(u)F(\frac{uc\Delta_n^\beta}{\gamma})du$ that is in this case close to zero, because of the natural choice of taking a truncated function which is symmetric. Indeed, even if the density function F isn't symmetric, asymptotically the only contribution it gives

	Mean for $\theta_2 = 2$	Mean for $\sigma = 0.5$
$\tilde{\theta}_n^{\text{Euler}}$	2.04057	0.67069
$\widetilde{\theta}_n$	2.03926	0.53084

(a) $\beta = 0.3$

	Mean for $\theta_2 = 2$	Mean for $\sigma = 0.5$
$\tilde{\theta}_n^{\text{Euler}}$	2.02889	0.51503
$\widetilde{\theta}_n$	2.02686	0.50930

(b) $\beta = 0.49$

Table 2.3 – Monte Carlo estimates of θ_2 and σ^2 from 500 samples. We have here fixed $\beta = 0.3$ in the first table and $\beta = 0.49$ in the second one.

is due by its value in zero and, therefore, the symmetry of φ is enough to ensure the limit of the integral is zero.
Also in this case the estimator $\widetilde{\theta}_n$ performs well for any β, which means that even if we take a wrong threshold c for which the well detection of the jumps is not guaranteed, the performance of the estimator $\widetilde{\theta}_n$ remains high.

2.6 Preliminary results

Before proving the main statistical results of Section 3.3, we need to state several propositions which will be useful in the sequel. They will be proven in Section 2.8.1.

2.6.1 Limit theorems

The asymptotic properties of estimators are deduced from the asymptotic behavior of the contrast function. We therefore state some propositions useful to get the asymptotic behavior of U_n.

Proposition 13. *Suppose that Assumptions 1 to 4 and A_{Step} hold, $\Delta_n \to 0$ and $T_n \to \infty$ and f is a differentiable function $\mathbb{R}\times\Theta \to \mathbb{R}$ such that $|f(x,\theta)| \le c(1+|x|^c)$, $|\partial_x f(x,\theta)| \le c(1+|x|^c)$ and, for $\vartheta = \mu$ and $\vartheta = \sigma$, $|\partial_\vartheta f(x,\theta)| \le c(1+|x|^c)$. Then $x \mapsto f(x,\theta)$ is a π-integrable function for any $\theta \in \Theta$ and the following convergences hold as $n \to \infty$:*
1. $|\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0,$
2. $|\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i} f(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0,$
3. $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0,$
4. $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0.$

Statements $1-2$ and $3-4$ of the proposition here above, as well as the first and the second point of Proposition 14 below, turn out being similar if the sampling step $\Delta_{n,i} = \Delta_n$ considered is uniform. Otherwise, we need these two different convergences because, in order to estimate μ and σ jointly, we have to deal with different scaling of the contrast function.

Proposition 14. *Suppose that Assumptions 1 to 4 and A_{Step} hold, $\Delta_n \to 0$ and $T_n \to \infty$ and f: $\mathbb{R}\times\Theta \to \mathbb{R}$. Moreover we suppose that $\exists c$: $|f(x,\theta)| \le c(1+|x|^c)$ and that $\beta \in (\frac{1}{4}, \frac{1}{2})$.*

Then, $\forall \theta \in \Theta$,

1. $$\frac{1}{T_n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)\,(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\,\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\xrightarrow{\mathbb{P}}$$
$$\xrightarrow{\mathbb{P}}\int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx).$$

2. $$\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i},\theta)}{\Delta_{n,i}}\,(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\,\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\xrightarrow{\mathbb{P}}$$
$$\xrightarrow{\mathbb{P}}\int_{\mathbb{R}} f(x,\theta)a^2(x,\sigma_0)\pi(dx).$$

The proof relies on the following lemma. In the sequel we will denote $\mathbb{E}_i[.]$ for $\mathbb{E}[.|\mathcal{F}_{t_i}]$, where $(\mathcal{F}_s)_s$ is the filtration defined in Lemma 25.

Lemma 14. *Suppose that Assumptions 1 to 4 hold. Moreover we suppose that* $\beta\in(\frac{1}{4},\frac{1}{2})$. *Then*

1.

$$\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\,\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]=\Delta_{n,i}a^2(X_{t_i},\sigma_0)+R(\theta,\Delta_{n,i}^{1+\beta},X_{t_i}), \tag{2.23}$$

2.

$$\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^4\,\varphi^4_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]=3\Delta_{n,i}^2a^4(X_{t_i},\sigma_0)+R(\theta,\Delta_{n,i}^{\frac{7}{4}+\beta},X_{t_i}), \tag{2.24}$$

3.

$$\text{For } k\ge 1,\quad |\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))\,\varphi^k_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]|\le R(\theta,\Delta_{n,i},X_{t_i}), \tag{2.25}$$

4.

$$\text{For } k\ge 2,\ \forall k'>0,\quad \mathbb{E}_i[|X_{t_{i+1}}-m(\mu,\sigma,X_{t_i})|^k\,|\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|^{k'}]\le \tag{2.26}$$
$$\le R(\theta,\Delta_{n,i}^{\frac{k}{2}\wedge(1+\beta k)},X_{t_i}).$$

5.

$$\forall k'>0,\quad \mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^3|\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})|^{k'}]=R(\theta_0,\Delta_{n,i}^{\frac{4}{3}+\beta},X_{t_i}).$$

We observe that the first and the second points here above are particular cases of the the fourth one, in which we get some better estimation. In particular, we can identify in detail the main term.
Concerning the fifth point, instead, we remark that for $k=3$ we don't have the main contribution of the Brownian part anymore, which gave us the rest function of size $\Delta_{n,i}^{\frac{k}{2}}$ in (2.26). In this case the main term of the development is given by

the square of the Brownian integral times the jump part, which magnitude can be estimated by $\frac{4}{3} + \beta$.

In next lemma we consider the derivatives of φ, getting an improvement of the estimations here above. It relies on the fact that, from the definition we gave of such a function, we know its derivatives are different from zero only if the increments of our process are smaller than $2\Delta_{n,i}^{\beta}$ (as it was for φ) and bigger than $\Delta_{n,i}^{\beta}$ (extra bound that we did not get using φ). Having therefore no longer only an upper bound but also a lower bound for $X_{t_{i+1}} - X_{t_i}$, it is now possible to prove a better version of (2.26):

Lemma 15. *Suppose that Assumptions A1-A5 A7 and Ad hold. Then* $\forall p \geq 1$, $\forall k \geq 1$ *and* $\forall r > 0$,

$$\mathbb{E}_i[|X_{t_{i+1}}^{\theta} - m(\mu, \sigma, X_{t_i})|^p |\varphi_{\Delta_{n,i}^{\beta}}^{(k)}(X_{t_{i+1}}^{\theta} - X_{t_i}^{\theta})|^r] \leq R(\theta, h^{1+\beta p}, X_{t_i}).$$

Considering only the jump part, the following result holds:

Lemma 16. *Suppose that Assumptions A1-A4 holds. Then,* $\forall q \geq 1$ *we have*

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)|^q] = R(\theta_0, \Delta_{n,i}^{(1+\beta q)\wedge q}, X_{t_i}),$$

where $\Delta X_i^J := \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds, dz)$.

Other estimation about the expected value of the jump part in the presence of an indicator function which is 0 if the increments are bigger than $c\Delta_{n,i}^{\beta}$ are gathered in Lemma 4 of [5].

Using the lemmas stated here above, it is possible to prove the following proposition, that will be proved in Section 2.8.1 and which is useful to show the tightness of the contrast function.

Proposition 15. *Suppose that Assumptions 1 to 4 and* A_{Step} *hold,* $\Delta_n \to 0$ *and* $T_n \to \infty$ *and* $g_{i,n}$ *is a differentiable function* $\mathbb{R} \times \Theta \to \mathbb{R}$ *such that* $|g_{i,n}(x, \theta)| \leq c(1 + |x|^c)$ *and, for* $\vartheta = \mu$ *and* $\vartheta = \sigma$, $|\partial_{\vartheta} g_{i,n}(x, \theta)| \leq c(1 + |x|^c)$. *We define*

$$S_n(\theta) := \frac{1}{T_n} \sum_{i=0}^{n-1} (X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) g_{i,n}(X_{t_i}, \theta).$$

Then $S_n(\theta)$ *is tight in* $(C(\Theta), \|.\|_{\infty})$.

2.6.2 Derivatives of m and m_2

We now state some propositions which concern the derivatives of m and m_2 that will be useful in the sequel.

Proposition 16. *Suppose that Assumptions A1-A5 A7 and Ad hold. Then, for* $|x| \leq \Delta_{n,i}^{-k_0}$ *and* $\forall \epsilon > 0$, *we have*

$$1.\ \partial_{\mu} m(\mu, \sigma, X_{t_i}) = \Delta_{n,i} \partial_{\mu} b(X_{t_i}, \mu) + R(\theta, \Delta_{n,i}^{\frac{5}{2} - \beta - \epsilon}, X_{t_i}),$$

$$2.\ |\partial_\sigma m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}, X_{t_i}),$$

$$3.\ |\partial_\mu m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i}),$$

$$4.\ \partial_\sigma m_2(\mu, \sigma, X_{t_i}) = 2\Delta_{n,i}\partial_\sigma a(X_{t_i}, \sigma) a(X_{t_i}, \sigma) + R(\theta, \Delta_{n,i}^{1+\beta}, X_{t_i}).$$

Estimation on the second derivatives are gathered in the following proposition:

Proposition 17. *Suppose that Assumptions A1 - A5, A7 and Ad hold. Then*

$$|\partial^2_{\mu\sigma} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i}), \qquad |\partial^2_{\sigma} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}, X_{t_i}), \quad (2.27)$$

$$\partial^2_{\mu} m(\mu, \sigma, X_{t_i}) = \Delta_{n,i}\partial^2_\mu b(\mu, X_{t_i}) + R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i}), \quad (2.28)$$

$$|\partial^2_{\mu\sigma} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i}), \qquad |\partial^2_{\mu} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i}), \quad (2.29)$$

$$\partial^2_{\sigma} m_2(\mu, \sigma, X_{t_i}) = 2\Delta_{n,i}\partial_\sigma a(\sigma, X_{t_i}) a(\sigma, X_{t_i}) + R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i}). \quad (2.30)$$

Deriving once again, the orders do not get worse. Indeed, the following estimations hold.

Proposition 18. *Suppose that Assumptions A1 - A5, A7 and Ad hold. Then*

1. $|\partial^3_{\mu} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i})$; 2. $|\partial^3_{\sigma\mu\sigma} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i})$,

3. $|\partial^3_{\mu\mu\sigma} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^2, X_{t_i})$; 4. $|\partial^3_{\sigma} m_2(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}, X_{t_i})$,

5. $|\partial^3_{\mu} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}, X_{t_i})$; 6. $|\partial^3_{\sigma\mu\sigma} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i})$,

7. $|\partial^3_{\mu\mu\sigma} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}^{\frac{3}{2}}, X_{t_i})$; 8. $|\partial^3_{\sigma} m(\mu, \sigma, X_{t_i})| \le R(\theta, \Delta_{n,i}, X_{t_i})$.

Propositions 16, 17 and 18 will be proved in the appendix.

2.7 Proof of main results

We first of all study the asymptotic behaviour of the contrast, from which we find the consistency of our estimator.
We underline that, to get the consistency of the drift parameter, the normalization of the contrast function is different than the normalization we use to find the consistency of $\hat{\sigma}_n$. Even if it doesn't seem a natural choice, it works well on the basis of Proposition 14.

2.7.1 Contrast's convergence

To prove the contrast convergences, the development (2.10) of m_2 will be useful. We have shown in Chapter 1 (see (2.16) also) that under Assumptions $(A1)$-$(A4)$ the following development of $m(\mu, \sigma, x)$ holds :

$$m(\mu, \sigma, x) = x + \Delta_{n,i} b(x, \mu) + R^J(\Delta_{n,i}, x) + r_1(\mu, \sigma, x), \tag{2.31}$$

where $r_1(\mu, \sigma, x)$ is a particular $R(\theta, \Delta_{n,i}^{1+\delta}, X_{t_i})$ function (with $\delta > 0$) and $R^J(\Delta_{n,i}, x) = -\Delta_{n,i} \int_{\mathbb{R}} z\gamma(x)[1 - \varphi_{\Delta_{n,i}^\beta}(\gamma(x)z)]F(z)dz$; the J underlines that it turns out from a jump term. It has the same properties of the function R defined in Section 2.4.2 but it does not depend on θ.
Let us now prove the consistency of $\hat{\theta}_n$. The first step are the following lemmas:

Lemma 17. *Suppose that A1 - A5, A_{Step} and Ad hold. Moreover we suppose that $\beta \in (\frac{1}{4}, \frac{1}{2})$. Then*

$$\frac{1}{n} U_n(\mu, \sigma) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} [\frac{c(x, \sigma_0)}{c(x, \sigma)} + log(c(x, \sigma))]\pi(dx), \tag{2.32}$$

where $c(x, \sigma) = a^2(x, \sigma)$ and π is the invariant distribution defined in Lemma 13.

Lemma 17 is useful to prove the consistency of $\hat{\sigma}_n$, while we will use next lemma to show the consistency of $\hat{\mu}_n$

Lemma 18. *Suppose that A1 - A5, A_{Step} and Ad hold. Moreover we suppose that $\beta \in (\frac{1}{4}, \frac{1}{2})$ and that $2\delta_1 > 1$. Then*

$$\frac{1}{T_n}(U_n(\mu, \sigma) - U_n(\mu_0, \sigma)) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x, \mu_0) - b(x, \mu))^2}{c(x, \sigma)} \pi(dx) + \tag{2.33}$$

$$+ \int_{\mathbb{R}} [r(\mu, \sigma, x) - r(\mu_0, \sigma, x)](1 - \frac{c(x, \sigma_0)}{c(x, \sigma)})\pi(dx),$$

where $r(\mu, \sigma, x)$ is the particular $R(\theta, 1, x)$ function who turns out from the development (2.10) of m_2.

2.7.1.1 Proof of Lemma 17

Proof. We first of all observe that by the equation (2.10) we have, for $|X_{t_i}| \le \Delta_{n,i}^{-k}$,

$$\frac{1}{m_2(\mu, \sigma, X_{t_i})} = \frac{1}{\Delta_{n,i} c(X_{t_i}, \sigma)(1 + \Delta_{n,i}^{\delta_1} \frac{r(X_{t_i})}{c(X_{t_i}, \sigma)} + \Delta_{n,i} r(\mu, \sigma, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\delta_2}, X_{t_i}))} =$$

$$= \frac{1}{\Delta_{n,i} c(X_{t_i}, \sigma)}(1 - \Delta_{n,i}^{\delta_1} \frac{r(X_{t_i})}{c(X_{t_i}, \sigma)} - \Delta_{n,i} r(\mu, \sigma, X_{t_i}) + R(\theta, \Delta_{n,i}^{2\delta_1 \wedge (1+\delta_2) \wedge 2}, X_{t_i})). \tag{2.34}$$

In the sequel we will just write $\bar{r}$ for $2\delta_1 \wedge (1 + \delta_2) \wedge 2$. We observe that, as a consequence of the Assumption Ad, we have for both $\vartheta = \mu$ and $\vartheta = \sigma$, $|\partial_\vartheta R(\theta, \Delta_{n,i}^{\bar{r}}, X_{t_i})| \le R(\theta, \Delta_{n,i}^{\bar{r}}, X_{t_i})$, where the two rest functions are not necessarily the same.
Similarly,

$$log(\frac{m_2(\mu, \sigma, X_{t_i})}{\Delta_{n,i}}) =$$

$$log(c(X_{t_i},\sigma)) + log(1+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)} + \Delta_{n,i}r(\mu,\sigma,X_{t_i}) + R(\theta,\Delta_{n,i}^{1+\delta_2},X_{t_i})) =$$

$$= log(c(X_{t_i},\sigma)) + \Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)} + \Delta_{n,i}r(\mu,\sigma,X_{t_i}) + R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}). \tag{2.35}$$

Using both (2.34) and (2.35) in the definition of $U_n(\mu,\sigma)$, we have to show that

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}(1-\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+$$

$$-\Delta_{n,i}r(\mu,\sigma,X_{t_i})+R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}))+\frac{1}{n}\sum_{i=0}^{n-1}(log(c(X_{t_i},\sigma))+\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}+\Delta_{n,i}r(\mu,\sigma,X_{t_i})+$$

$$+R(\theta,\Delta_{n,i}^{\bar{r}},X_{t_i}))\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} =: \sum_{j=1}^{8}I_j^n$$

converges to the right hand side of (2.32). We know that $I_1^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}}(\frac{c(x,\sigma_0)}{c(x,\sigma)})\pi(dx)$ because of Proposition 14. Using the third point of Proposition 13,

$$I_5^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} log(c(x,\sigma))\pi(dx).$$

All the other terms converge to zero in norm 1 and so in probability. Indeed, passing through the conditional expectation and using the first point of Lemma 14 we have

$$\mathbb{E}[|I_2^n|] \le$$

$$\le \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|\frac{\Delta_{n,i}^{\delta_1}r(X_{t_i})}{\Delta_{n,i}c^2(X_{t_i},\sigma)}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}}-X_{t_i})]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}|] \le$$

$$\le \frac{\Delta_n^{\delta_1}}{n}\sum_{i=0}^{n-1}\mathbb{E}[|R(\theta,1,X_{t_i})|] \le c\Delta_n^{\delta_1},$$

reminding that $r(X_{t_i})$ is a function $R(\theta,1,X_{t_i})$ by its definition and having used the property (3.24) on R, its polynomial growth and the third point of Lemma 13. In the same way we obtain

$$\mathbb{E}[|I_3^n|] \le c\Delta_n \qquad \text{and} \qquad \mathbb{E}[|I_4^n|] \le c\Delta_n^{\bar{r}},$$

that goes to zero since $\bar{r} = 2\delta_1 \wedge (1+\delta_2) \wedge 2$ is always positive.
Concerning I_6^n, as a consequence of the definition of $r(x)$ and the fact that $\Delta_{n,i} \le \Delta_n$ we have again

$$\mathbb{E}[|I_6^n|] \le \frac{\Delta_n^{\delta_1}}{n}\sum_{i=0}^{n-1}\mathbb{E}[|R(\theta,1,X_{t_i})|] \le c\Delta_n^{\delta_1},$$

which converges to zero for $n \to \infty$. Again, acting in the same way we have

$$\mathbb{E}[|I_7^n|] \le c\Delta_n \qquad \text{and} \qquad \mathbb{E}[|I_8^n|] \le c\Delta_n^{\bar{r}}.$$

Convergence (2.32) follows. □

2.7.1.2 Proof of Lemma 18

Proof. Using again (2.34) and (2.35) we have that

$$\frac{1}{T_n}(U_n(\mu,\sigma)-U_n(\mu_0,\sigma)) =$$

$$= \frac{1}{T_n}\sum_{i=0}^{n-1}[\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)} - \frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}\frac{\Delta_{n,i}^{\delta_1}r(X_{t_i})}{c(X_{t_i},\sigma)}[\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2-(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})[1-\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$-\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu_0,\sigma,X_{t_i})[1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}R((\mu,\sigma),\Delta_{n,i}^{\bar r},X_{t_i})[1+\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{T_n}\sum_{i=0}^{n-1}R((\mu_0,\sigma),\Delta_{n,i}^{\bar r},X_{t_i})[1+\frac{(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} =$$

$$=: \sum_{j=1}^{6} I_j^n,$$

where we have introduced the notation $\Delta_i X := X_{t_{i+1}} - X_{t_i}$ and we recall that $\bar r = 2\delta_1 \wedge (1+\delta_2) \wedge 2$.

We have already proved in Lemma 4 of [3] that $I_1^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\mu_0)-b(x,\mu))^2}{c(x,\sigma)}\pi(dx)$. We observe that I_2^n differs from I_1^n only from the presence of $\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma)}$ and so, since δ_1 is positive, acting exactly as we did in order to prove the convergence of I_1^n it is possible to show that the added $\Delta_{n,i}^{\delta_1}$ make I_2^n converge to zero in probability.

Concerning I_3^n,

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} r(\mu,\sigma,x)\pi(dx) \tag{2.36}$$

as a consequence of the second point of Proposition 13. Moreover, using the third point of Proposition 14, we have that

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\Delta_{n,i}r(\mu,\sigma,X_{t_i})\frac{(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2}{\Delta_{n,i}c(X_{t_i},\sigma)}\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} \tag{2.37}$$

$$\int_{\mathbb{R}} r(\mu,\sigma,x)\frac{c(x,\sigma_0)}{c(x,\sigma)}\pi(dx).$$

From (2.36) and (2.37) it follows

$$I_3^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} r(\mu,\sigma,x)[1-\frac{c(x,\sigma_0)}{c(x,\sigma)}]\pi(dx).$$

Acting on I_4^n exactly like we did on I_3^n we get

$$I_4^n \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} r(\mu_0, \sigma, x)[1 - \frac{c(x, \sigma_0)}{c(x, \sigma)}]\pi(dx).$$

Concerning I_5^n, it is

$$\frac{1}{T_n} \sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}^{\bar{r}}, X_{t_i})\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} \le \Delta_n^{\bar{r}-1} \frac{1}{n} \sum_{i=0}^{n-1} R(\theta, 1, X_{t_i})\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X),$$

which converges to zero in norm 1 and so in probability because of the boundedness of φ, the polynomial growth of R, the fact that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$ and that $r-1$ is always positive since we have assumed $2\delta_1 > 1$. Moreover, passing through the conditional expectation and using the first point of Lemma 14 we have that

$$\frac{1}{T_n} \sum_{i=0}^{n-1} \mathbb{E}[R(\theta, \Delta_{n,i}^{\bar{r}}, X_{t_i})\mathbb{E}_i[\frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{\Delta_{n,i} c(X_{t_i}, \sigma)}\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)] 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}] \le$$

$$\le \Delta_n^{\bar{r}-1} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[R(\theta, 1, X_{t_i})] \le c\Delta_n^{\bar{r}-1}.$$

We have therefore proved that the second part of I_5^n converges to 0 in norm 1 and therefore in probability. It follows $I_5^n \xrightarrow{\mathbb{P}} 0$ and, acting exactly in the same way, we have also $I_6^n \xrightarrow{\mathbb{P}} 0$. It yields (2.33). □

2.7.2 Consistency of the estimator.

In order to prove the consistency of $\hat{\theta}_n$, we need that the convergences (2.32) and (2.33) take place in probability uniformly in both the parameters, we want therefore to show the uniformity of the convergence in θ.
We regard $\frac{U_n(\mu,\sigma)}{n}$ and $S_n(\mu, \sigma) := \frac{1}{T_n}(U_n(\mu, \sigma) - U_n(\mu_0, \sigma))$ as random elements taking values in $(C(\Theta), \|.\|_\infty)$. It suffices to prove the tightness of these sequences; to do it we need some estimations for the derivatives of m and m_2 with respect to both the parameters, which are stated in Proposition 16, that will be proved in the Appendix. Such a proposition will be also useful to study the asymptotic behavior of the derivatives of the contrast function. We observe that, as a consequence of (2.31) and Proposition 16, for $\vartheta = \mu$ or $\vartheta = \sigma$ it is $|\partial_\vartheta r_1(\mu, \sigma, x)| \le R(\theta, \Delta_{n,i}, X_{t_i})$.

Lemma 19. *Suppose that Assumption A1-A5, Ad, A_{Step} and A7 are satisfied. Then, the sequence $\frac{U_n(\mu,\sigma)}{n}$ is tight in $(C(\Theta), \|.\|_\infty)$.*

Proof. The tightness is implied by $\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma} |\partial_\vartheta U_n(\mu, \sigma)|] < \infty$ (see Corollary B.1 in [86]), for $\vartheta = \mu$ and $\vartheta = \sigma$. It is

$$\partial_\vartheta U_n(\mu, \sigma) = \sum_{i=0}^{n-1} [\frac{-2\partial_\vartheta m(\mu, \sigma, X_{t_i})(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))}{m_2(\mu, \sigma, X_{t_i})} + \tag{2.38}$$

$$-\frac{\partial_\vartheta m_2(\mu, \sigma, X_{t_i})(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{m_2^2(\mu, \sigma, X_{t_i})} + \frac{\partial_\vartheta m_2(\mu, \sigma, X_{t_i})}{m_2(\mu, \sigma, X_{t_i})}]\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}.$$

Using the first and the third point of Proposition 16 and the development (2.10) of m_2 it follows

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu U_n(\mu,\sigma)|] \leq \sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|1_{i,n}]+$$

$$+\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|1_{i,n}]+ \tag{2.39}$$

$$+\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i},X_{t_i})|1_{i,n}],$$

where we have used $1_{i,n}$ instead of $1_{\left\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\right\}}$ to shorten the notation.
We observe that

$$\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|1_{i,n}] \leq$$

$$\leq \mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma}|(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|)1_{i,n}] \leq$$

$$\leq \mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|1_{i,n})(|(X_{t_{i+1}}-m(\mu_0,\sigma,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|)]+$$

$$+c\mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma}|m(\mu,\sigma,X_{t_i})-m(\mu_0,\sigma,X_{t_i}))|)1_{i,n}]. \tag{2.40}$$

We can now use Cauchy-Schwartz inequality and (2.23) in Lemma 14 on the first, while on the second we use the development (2.31) of m getting that (2.40) is upper bounded by

$$c\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})]^{\frac{1}{2}}+c\mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|)(\sup_{\mu,\sigma}|\Delta_{n,i}(b(X_{t_i},\mu)-b(X_{t_i},\mu_0))+$$

$$+r_1(\mu,\sigma,X_{t_i})-r_1(\mu_0,\sigma,X_{t_i})|)] \leq c\Delta_n^{\frac{1}{2}}+c\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i},X_{t_i})|] \leq c\Delta_n^{\frac{1}{2}}+c\Delta_n \leq c\Delta_n^{\frac{1}{2}}, \tag{2.41}$$

where we have also used the boundedness of φ, the fact that R has polynomial growth uniformly in θ and the third point of Lemma 13 to say that our process has finite moments.
In the same way,

$$\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|1_{i,n}] \leq \tag{2.42}$$

$$\leq \mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i},X_{t_i})|]+\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i}^2,X_{t_i})|] \leq c\Delta_n.$$

Replacing (2.41) and (2.42) in (2.39) it follows

$$\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu U_n(\mu,\sigma)|] \leq c\Delta_n^{\frac{1}{2}} \leq c < \infty.$$

We can act in the same way on $\partial_\sigma U_n(\mu,\sigma)$. Considering this time the second and the fourth point of Proposition 16 and still using the development (2.10) of m_2 and (2.23) in Lemma 14 it follows

$$\sup_n \frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma U_n(\mu,\sigma)|] \leq \sup_n \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})+R(\theta,1,X_{t_i})|] \leq c < \infty.$$

The tightness is therefore proved. □

Lemma 20. *Suppose that Assumption A1-A5, A7, A_{Step} and Ad are satisfied. We suppose moreover that δ_1 in the Assumption Ad is such that $2\delta_1 > 1$. Then, the sequence $S_n(\mu, \sigma) = \frac{1}{T_n}(U_n(\mu, \sigma) - U_n(\mu_0, \sigma))$ is tight in $(C(\Theta), \|.\|_\infty)$.*

Proof. Let's take again the notation used in the proof of Lemma 18, for which $S_n(\mu, \sigma) = \sum_{j=1}^6 I_j^n$. Since the sum of tight sequences is still tight, we will proceed showing that they are all tight. We start with I_3^n; acting as we did in Lemma 19, we prove that $\sup_n \mathbb{E}[\sup_{\mu,\sigma} |\partial_\vartheta I_3^n|] < \infty$. We observe that, for $\vartheta = \mu$ and $\vartheta = \sigma$,

$$\partial_\vartheta I_3^n = \frac{1}{T_n} \sum_{i=0}^{n-1} \Delta_{n,i}[\partial_\vartheta r(\mu, \sigma, X_{t_i})(1 - \frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{\Delta_{n,i} c(X_{t_i}, \sigma)})+$$

$$-r(\mu, \sigma, X_{t_i})\partial_\vartheta(\frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{\Delta_{n,i} c(X_{t_i}, \sigma)})]\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}} =: I_{3,1}^n + I_{3,2}^n.$$

On $I_{3,1}^n$ we use the first point of Lemma 14 and that $|\partial_\vartheta r(\mu, \sigma, X_{t_i})| \le R(\theta, 1, X_{t_i})$ as stated in Assumption Ad to get

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma} |I_{3,1}^n|] \le$$

$$\le c + \sup_n \frac{1}{n\Delta_n} \sum_{i=0}^{n-1} \mathbb{E}[\mathbb{E}_i[\sup_{\mu,\sigma} |R(\theta, 1, X_{t_i})(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|1_{i,n}]] \le c \tag{2.43}$$

where we have used the polynomial growth of R, the third point of Lemma 13 and (2.42) and the notation $1_{i,n}$ instead of $1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}$.
Concerning $I_{3,2}^n$, the derivatives of $\frac{(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2}{\Delta_{n,i} c(X_{t_i}, \sigma)}$ with respect to μ and σ are different but in both cases they are upper bounded, using the first and the second point of Proposition 16, by $|R(\theta, 1, X_{t_i})(X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(\Delta_i X)|$. We can therefore use (2.40) and (2.41), getting

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma} |I_{3,2}^n|] \le c \sup_n \Delta_n^{\frac{1}{2}} \le c < \infty. \tag{2.44}$$

From (2.43) and (2.44) it follows the tightness of I_3^n. Acting exactly in the same way on I_4^n it is clear it is tight too. Concerning I_5^n and I_6^n, recalling that the function $R(\theta, \Delta_{n,i}^{\bar{r}}, X_{t_i})$ turns out from (2.34) and it is such that its derivatives with respect to both the parameters remains of the same order, we observe it is possible to act like we did on I_3^n getting

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma} |I_5^n|] \le c\Delta_n^{\bar{r}-1} + c\Delta_n^{\bar{r}-\frac{1}{2}} < \infty,$$

since we have chosen $2\delta_1 > 1$ and so $\bar{r} - 1$ is positive. Clearly the same estimation hold for I_6^n.
We now prove that I_1^n is tight. To do it we observe that, using the development (2.31) and the dynamic (4.3) of the process X we have

$$X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}) =$$

$$= \int_{t_i}^{t_{i+1}} b(X_s, \mu_0)ds + \int_{t_i}^{t_{i+1}} a(\sigma_0, X_s)dW_s + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds, dz)+ \tag{2.45}$$

$$-R^J(\Delta_{n,i}, X_{t_i}) - \Delta_{n,i} b(X_{t_i}, \mu) - r_1(\mu, \sigma, X_{t_i}).$$

It is worth noting that only the last two terms here above depend on μ and so replacing (2.45) in I_1^n some terms are deleted by compensation. Therefore we can define

$$I_{1,1}^n : \frac{1}{T_n} \sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{\Delta_{n,i} c(\sigma, X_{t_i})} [\Delta_{n,i}^2 (b^2(X_{t_i}, \mu) - b^2(X_{t_i}, \mu_0)) +$$

$$+ r_1^2(\mu, \sigma, X_{t_i}) - r_1^2(\mu_0, \sigma, X_{t_i}) + 2\Delta_{n,i} b(X_{t_i}, \mu) r_1(\mu, \sigma, X_{t_i}) +$$

$$-2\Delta_{n,i} b(X_{t_i}, \mu_0) r_1(\mu_0, \sigma, X_{t_i}) + 2[\int_{t_i}^{t_{i+1}} b(X_s, \mu_0) ds + \Delta X_i^J +$$

$$-R^J(\Delta_{n,i}, X_{t_i})][\Delta_{n,i}(b(X_{t_i}, \mu) - b(X_{t_i}, \mu_0)) + r_1(\mu, \sigma, X_{t_i}) - r_1(\mu_0, \sigma, X_{t_i})],$$

where we have denoted by ΔX_i^J the jump part in ΔX_i, that is $\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-}) z \tilde{\mu}(ds, dz)$.
Moreover we define

$$I_{1,2}^n := \frac{1}{T_n} \sum_{i=0}^{n-1} \frac{2(b(X_{t_i}, \mu) - b(X_{t_i}, \mu_0)) \int_{t_i}^{t_{i+1}} a(\sigma_0, X_s) dW_s}{c(\sigma, X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}},$$

$$I_{1,3}^n := \frac{1}{T_n} \sum_{i=0}^{n-1} \frac{2(r_1(\mu, \sigma, X_{t_i}) - r_1(\mu_0, \sigma, X_{t_i})) \int_{t_i}^{t_{i+1}} a(\sigma_0, X_s) dW_s}{\Delta_{n,i} c(\sigma, X_{t_i})} \varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}.$$

It is $I_1^n = I_{1,1}^n + I_{1,2}^n + I_{1,3}^n$. We are going to prove that $I_{1,1}^n$ is tight showing that the expected value of the derivatives is bounded, like we have already done. On $I_{1,2}^n$ and $I_{1,3}^n$ we will use instead the Kolmogorov criterion for which, if for some positive constant H independent of n and for $m \ge r > 2$, S_n is a sequence such that

$$\mathbb{E}[(S_n(\theta))^m] \le H \qquad \forall \theta \in \Theta, \tag{2.46}$$

$$\mathbb{E}[(S_n(\theta_1) - S_n(\theta_2))^m] \le H|\mu_1 - \mu_2|^r + H|\sigma_1 - \sigma_2|^r \qquad \forall \theta_1, \theta_2 \in \Theta, \tag{2.47}$$

then S_n is tight.
Let us start considering $I_{1,1}^n$: we want to show that $\sup_n \mathbb{E}[\sup_{\mu,\sigma} |\partial_\vartheta I_{1,1}^n|] < \infty$. We observe it is

$$\partial_\mu I_{1,1}^n = \frac{1}{T_n} \sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{\Delta_{n,i} c(\sigma, X_{t_i})} [\Delta_{n,i}^2 (2b\, \partial_\mu b)(X_{t_i}, \mu) + (2r_1 \partial_\mu r_1)(\mu, \sigma, X_{t_i}) +$$

$$+2\Delta_{n,i}((\partial_\mu b)(X_{t_i}, \mu) r_1(\mu, \sigma, X_{t_i}) + b(X_{t_i}, \mu)(\partial_\mu r_1)(\mu, \sigma, X_{t_i})) + 2(\int_{t_i}^{t_{i+1}} b(X_s, \mu_0) ds +$$

$$+\Delta X_i^J - R^J(\Delta_{n,i}, X_{t_i}))(\Delta_{n,i} \partial_\mu b(X_{t_i}, \mu) + \partial_\mu r_1(\mu, \sigma, X_{t_i}))];$$

$$\partial_\sigma I_{1,1}^n = \frac{1}{T_n} \sum_{i=0}^{n-1} \frac{\varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{\Delta_{n,i} c(\sigma, X_{t_i})} [2r_1 \partial_\sigma r_1(\mu, \sigma, X_{t_i}) - 2r_1 \partial_\sigma r_1(\mu_0, \sigma, X_{t_i}) +$$

$$+2\Delta_{n,i}(b(X_{t_i}, \mu) \partial_\sigma r_1(\mu, \sigma, X_{t_i}) - b(X_{t_i}, \mu_0) \partial_\sigma r_1(\mu_0, \sigma, X_{t_i})) + 2(\int_{t_i}^{t_{i+1}} b(X_s, \mu_0) ds +$$

$$+\Delta X_i^J - R^J(\Delta_{n,i}, X_{t_i}))(\partial_\sigma r_1(\mu_0, \sigma, X_{t_i}) - \partial_\sigma r_1(\mu, \sigma, X_{t_i}))] +$$

$$-\frac{\partial_\sigma c(X_{t_i}, \sigma) \varphi_{\Delta_{n,i}^\beta}(\Delta_i X) 1_{\{|X_{t_i}| \le \Delta_{n,i}^{-k}\}}}{\Delta_{n,i} c^2(\sigma, X_{t_i})} [\Delta_{n,i}^2 (b^2(X_{t_i}, \mu) - b^2(X_{t_i}, \mu_0)) +$$

$$+r_1^2(\mu,\sigma,X_{t_i})-r_1^2(\mu_0,\sigma,X_{t_i})+2\Delta_{n,i}b(X_{t_i},\mu)r_1(\mu,\sigma,X_{t_i})+$$
$$-2\Delta_{n,i}b(X_{t_i},\mu_0)r_1(\mu_0,\sigma,X_{t_i})+2(\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds+\Delta X_i^J+$$
$$-R^J(\Delta_{n,i},X_{t_i}))(\Delta_{n,i}(b(X_{t_i},\mu)-b(X_{t_i},\mu_0))+r_1(\mu,\sigma,X_{t_i})-r_1(\mu_0,\sigma,X_{t_i}))].$$

Using the polynomial growth of b and recalling that r_1 is the particular $R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i})$ function that turns out from the development (2.31) of m and it is such that $|\partial_\vartheta r_1(\mu,\sigma,X_{t_i})|\le R(\theta,\Delta_{n,i},X_{t_i})$ as a consequence of the first two points of Proposition 16, we get

$$|\partial_\vartheta I_{1,1}^n|\le\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}|\varphi_{\Delta_{n,i}^\beta}(\Delta_iX)|[R(\theta,\Delta_{n,i},X_{t_i})+R(\theta,\Delta_{n,i}^{1+\delta},X_{t_i})+$$
$$+2(|\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds|+|\Delta X_i^J|+R^J(\Delta_{n,i},X_{t_i}))(R(\theta,1,X_{t_i})+R(\theta,\Delta_{n,i}^\delta,X_{t_i}))].$$

Using Lemma 16, the boundedness of φ, the fact that $\frac{1}{T_n}=O(\frac{1}{n\Delta_n})$ and that $2\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds|]$ is a $R(\theta_0,\Delta_{n,i},X_{t_i})$, it follows

$$\sup_n\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I_{1,1}^n|]\le\sup_n(\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta,\Delta_{n,i},X_{t_i})+(R(\theta_0,\Delta_{n,i},X_{t_i})+$$
$$+R^J(\Delta_{n,i},X_{t_i}))R(\theta,1,X_{t_i})|]+$$
$$+\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[(\sup_{\mu,\sigma}|R(\theta,1,X_{t_i})|)\mathbb{E}_i[|\Delta X_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta_iX)|]])\le c,$$

where in the last inequality we have used Lemma 16 here above, the polynomial growth of R uniform in θ and the third point of Lemma 13. $I_{1,1}^n$ is therefore tight. We now show that (2.46) and (2.47) hold on $I_{1,2}^n$. Indeed, using Burkholder and Jensen inequalities, we get

$$\mathbb{E}[|I_{1,2}^n(\theta_1)-I_{1,2}^n(\theta_2)|^m]\le\frac{c}{n^m\Delta_n^m}n^{\frac{m}{2}-1}\sum_{i=0}^{n-1}\mathbb{E}[|\frac{b(X_{t_i},\mu_1)-b(X_{t_i},\mu_0)}{c(\sigma_1,X_{t_i})}+$$
$$-\frac{b(X_{t_i},\mu_2)-b(X_{t_i},\mu_0)}{c(\sigma_2,X_{t_i})}|^m|\int_{t_i}^{t_{i+1}}a(\sigma_0,X_s)dW_s|^m|\varphi_{\Delta_{n,i}^\beta}(\Delta_iX)|^m]. \tag{2.48}$$

We observe that, as a consequence of the finite-increments theorem, we have

$$|\frac{b(X_{t_i},\mu_1)-b(X_{t_i},\mu_0)}{c(\sigma_1,X_{t_i})}-\frac{b(X_{t_i},\mu_2)-b(X_{t_i},\mu_0)}{c(\sigma_2,X_{t_i})}|^m\le|\frac{\partial_\mu b(X_{t_i},\tilde\mu)}{c(X_{t_i},\tilde\sigma)}(\mu_1-\mu_2)+ \tag{2.49}$$
$$-\frac{(b(X_{t_i},\tilde\mu)-b(X_{t_i},\mu_0))\partial_\sigma c(X_{t_i},\tilde\sigma)}{c(X_{t_i},\tilde\sigma)}(\sigma_1-\sigma_2)|^m\le$$
$$\le R(\theta,1,X_{t_i})|\mu_1-\mu_2|^m+R(\theta,1,X_{t_i})|\sigma_1-\sigma_2|^m,$$

where actually the functions R are calculated in a point $\tilde\theta:=(\tilde\mu,\tilde\sigma)$, with $\tilde\mu\in(\mu_1,\mu_2)$ and $\tilde\sigma\in(\sigma_1,\sigma_2)$ but, since the property (3.23) of R is uniform in θ, we have chosen to write it simply as $R(\theta,1,X_{t_i})$. Using also the boundedness of φ, we get that (2.48) is upper bounded by

$$\frac{cn^{\frac{m}{2}-1}}{n^m\Delta_n^m}\sum_{i=0}^{n-1}\mathbb{E}[|\int_{t_i}^{t_{i+1}}a(\sigma_0,X_s)dW_s|^mR(\theta,1,X_{t_i})]|\mu_1-\mu_2|^m+$$

$$+\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(\sigma_0, X_s)dW_s|^m R(\theta,1,X_{t_i})]|\sigma_1-\sigma_2|^m.$$

Using Burkholder-Davis-Gundy inequality we have, $\forall p \geq 2$,

$$\mathbb{E}[(\int_{t_i}^{t_{i+1}} a(\sigma, X_s)dW_s)^p] \leq \mathbb{E}[(\int_{t_i}^{t_{i+1}} a^2(\sigma, X_s)ds)^{\frac{p}{2}}] \leq \mathbb{E}[R(\theta, \Delta_{n,i}, X_{t_i})^{\frac{p}{2}}] = c\Delta_{n,i}^{\frac{p}{2}}, \tag{2.50}$$

where in the last inequality we have used the polynomial growth of a and the third point of Lemma 13.

From Holder inequality and (2.50) it therefore follows

$$\mathbb{E}[|I_{1,2}^n(\theta_1) - I_{1,2}^n(\theta_2)|^m] \leq$$

$$\leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}}|\mu_1-\mu_2|^m + \frac{c}{(n\Delta_n)^{\frac{m}{2}}}|\sigma_1-\sigma_2|^m \leq c|\mu_1-\mu_2|^m + c|\sigma_1-\sigma_2|^m,$$

where we have also used that $n\Delta_n \to \infty$ for $n \to \infty$. For $r := m$ (2.47) is proved.

Concerning (2.46), acting in the same way we get

$$\mathbb{E}[|I_{1,2}^n(\theta)|^m] \leq \frac{cn^{\frac{m}{2}-1}}{n^m\Delta_n^m}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta_1,1,X_{t_i})^m|\int_{t_i}^{t_{i+1}} a(\sigma_0,X_s)dW_s|^m|\varphi_{\Delta_{n,i}^{\beta}}(\Delta_i X)|^m] \leq$$

$$\leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}} \leq c.$$

$I_{1,2}^n$ is hence tight. The tightness of $I_{1,3}^n$ is obtained acting exactly in the same way, remarking that

$$|\frac{r_1(\mu_0,\sigma_1,X_{t_i}) - r_1(\mu_1,\sigma_1,X_{t_i})}{\Delta_{n,i}c(\sigma_1,X_{t_i})} - \frac{r_1(\mu_0,\sigma_2,X_{t_i}) - r_1(\mu_2,\sigma_2,X_{t_i})}{\Delta_{n,i}c(\sigma_2,X_{t_i})}|^m \leq$$

$$\leq |\frac{\partial_\mu r_1(\tilde{\mu},\tilde{\sigma},X_{t_i})}{\Delta_{n,i}c(X_{t_i},\tilde{\sigma})}(\mu_1-\mu_2) + [\frac{\partial_\sigma r_1(\mu_0,\tilde{\sigma},X_{t_i}) - \partial_\sigma r_1(\tilde{\mu},\tilde{\sigma},X_{t_i})}{\Delta_{n,i}c(\tilde{\sigma},X_{t_i})} +$$

$$-\frac{\partial_\sigma c(\tilde{\sigma},X_{t_i})(r_1(\mu_0,\tilde{\sigma},X_{t_i}) - r_1(\tilde{\mu},\tilde{\sigma},X_{t_i}))}{\Delta_{n,i}c(\tilde{\sigma},X_{t_i})}](\sigma_1-\sigma_2)|^m \leq \tag{2.51}$$

$$\leq R(\theta,1,X_{t_i})|\mu_1-\mu_2|^m + R(\theta,1,X_{t_i})|\sigma_1-\sigma_2|^m,$$

as a consequence of the fact that $(r_1(\mu,\sigma,X_{t_i}))^m$ and $(\partial_\vartheta r_1(\mu,\sigma,X_{t_i}))^m$ are respectively upper bounded by $R(\theta,\Delta_{n,i}^{m(1+\delta)},X_{t_i})$ and $R(\theta,\Delta_{n,i}^m,X_{t_i})$.

Concerning I_2^n, we act like we did on I_1^n. We still use (2.45) getting $I_{2,1}^n$, $I_{2,2}^n$ and $I_{2,3}^n$.

We observe that, if we define s_j^n as $I_{1,j}^n =: \sum_{i=0}^{n-1} s_j^n$, then $I_{2,j}^n = \sum_{i=0}^{n-1}\Delta_{n,i}^{\delta}\frac{r(X_{t_i})}{c(\sigma,X_{t_i})}s_j^n$.

By the computation of $\partial_\mu I_{2,1}^n$ and $\partial_\sigma I_{2,1}^n$ it follows that

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta I_{2,1}^n|] \leq \sup_n(c\Delta_n^{\delta} + c\Delta_n^{\delta+\beta}) \leq c.$$

In order to prove that also $I_{2,2}^n$ and $I_{2,3}^n$ are tight we still use Kolmogorov criterion. From (2.49) and (2.50) it follows

$$\mathbb{E}[|I_{2,2}^n(\theta_1) - I_{2,2}^n(\theta_2)|^m] \leq c\frac{\Delta_n^{\delta m}}{(n\Delta_n)^{\frac{m}{2}}}|\mu_1-\mu_2|^m + c\frac{\Delta_n^{\delta m}}{(n\Delta_n)^{\frac{m}{2}}}|\sigma_1-\sigma_2|^m \leq$$

$$\leq c|\mu_1-\mu_2|^m + c|\sigma_1-\sigma_2|^m$$

and $\mathbb{E}[(I_{2,2}^n(\theta))^m] \leq c$.

The tightness of $I_{2,3}^n$ is obtained in the same way, through Kolmogorov criterion and (2.51).

The sequence S_n is therefore tight in $(C(\Theta), \|.\|_\infty)$, as we wanted. □

2.7.2.1 Proof of Theorem 8.

Proof. Let us begin with the consistency of $\hat{\sigma_n}$. An application of lemmas 17 and 19 yields

$$\frac{1}{n}U_n(\mu,\sigma) \xrightarrow{\mathbb{P}} U(\sigma,\sigma_0) := \int_{\mathbb{R}} [\frac{c(x,\sigma_0)}{c(x,\sigma)} + \log(c(x,\sigma))]\pi(dx) \tag{2.52}$$

uniformly in θ.
In order to prove that the uniform convergence here above implies the consistency of $\hat{\sigma_n}$, since the convergence in probability is equivalent to the existence, for any subsequence, of a subsequence converging almost surely, we will consider that the convergence in (2.52) is almost sure and prove that it implies that $\hat{\sigma}_n \to \sigma_0$ almost surely. For a fixed ω, thanks to the compactness of Θ, there exists a subsequence n_k such that $(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})$ tends to a limit $\theta_\infty := (\mu_\infty, \sigma_\infty)$. Hence, (2.52) together with the continuity of $\sigma \mapsto U(\sigma, \sigma_0)$, implies

$$\frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})(\omega) \to U(\sigma_\infty, \sigma_0).$$

But, by the definition of our estimator $\hat{\theta}_n$,

$$\frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) \leq \frac{1}{n_k}U_{n_k}(\hat{\mu}_{n_k}, \sigma_0).$$

So, using again the convergence (2.52), we get $U(\sigma_\infty, \sigma_0) \leq U(\sigma_0, \sigma_0)$. On the other hand, since for all $y > 0$, $y_0 > 0$ it is $\frac{y_0}{y} + \log(y) \geq 1 + \log(y_0)$ we deduce, using also the identifiability stated in Assumption A6 and Proposition 8.1 in Supplemental materials of [38], that $\sigma_\infty = \sigma_0$. We have proved that any convergent subsequence of $\hat{\sigma}_n$ tends to σ_0, hence $\hat{\sigma}_n \xrightarrow{\mathbb{P}} \sigma_0$ and we are done.
Concerning the consistency of $\hat{\mu}_n$, we have from Lemmas 18 and 20 that the convergence (2.33) holds uniformly in θ. In order to deduce the consistency of $\hat{\mu}_n$ the method is similar to the previous one. We know now that $(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k})$ tends to (μ_∞, σ_0), hence

$$\frac{1}{T_{n_k}}(U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) - U_{n_k}(\mu_0, \hat{\sigma}_{n_k})) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} \frac{(b(x,\mu_0) - b(x,\mu_\infty))^2}{c(x,\sigma_0)}\pi(dx) \geq 0.$$

But $U_{n_k}(\hat{\mu}_{n_k}, \hat{\sigma}_{n_k}) - U_{n_k}(\mu_0, \hat{\sigma}_{n_k}) \leq 0$ and so we conclude by A6, getting $\mu_\infty = \mu_0$ and therefore the consistency of $\hat{\mu}_n$. □

2.7.3 Asymptotic normality of the estimator.

The proof of the asymptotic normality goes along a classical route (see for instance Section 5a of [36]). We define the following notations:

$$M_n := \begin{pmatrix} \frac{1}{\sqrt{T_n}} & 0 \\ 0 & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

Let

$$S_n := \begin{pmatrix} \sqrt{T_n}(\hat{\mu}_n - \mu_0) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) \end{pmatrix}, \qquad L_n(\theta_0) := \begin{pmatrix} -\frac{1}{\sqrt{T_n}}\partial_\mu U_n(\mu_0,\sigma_0) \\ -\frac{1}{\sqrt{n}}\partial_\sigma U_n(\mu_0,\sigma_0) \end{pmatrix}$$

and

$$C_n(\theta) = \begin{pmatrix} \frac{1}{T_n}\frac{\partial^2}{\partial\mu^2}U_n(\mu,\sigma) & \frac{1}{\sqrt{nT_n}}\frac{\partial^2}{\partial\mu\sigma}U_n(\mu,\sigma) \\ \frac{1}{\sqrt{nT_n}}\frac{\partial^2}{\partial\mu\sigma}U_n(\mu,\sigma) & \frac{1}{n}\frac{\partial^2}{\partial\sigma^2}U_n(\mu,\sigma) \end{pmatrix}.$$

Then

$$M_n\nabla^2_\theta U_n(\mu,\sigma)M_n = C_n(\theta). \tag{2.53}$$

Now, by Taylor's formula,

$$\int_0^1 \nabla^2_\theta U_n(\theta_0 + u(\hat{\theta}_n - \theta_0))du \begin{pmatrix} \hat{\mu}_n - \mu_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} = -\nabla_\theta U_n(\theta_0),$$

since $\nabla_\theta U_n(\hat{\theta}_n) = 0$. Then, using (2.53), we have

$$\int_0^1 C_n(\theta_0 + u(\hat{\theta}_n - \theta_0))duS_n = L_n(\theta_0). \tag{2.54}$$

We deduce from this equality that, in order to prove the asymptotic normality of $\hat{\theta}_n$ and hence to end the proof of Theorem 9, it is enough to prove the following lemmas:

Lemma 21. *Suppose that Assumptions A1-A8 and Ad hold. Then, as $n \to \infty$,*

$$L_n(\theta_0) \xrightarrow{d} L \sim N(0, K'),$$

where $K' = \begin{pmatrix} 4\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx) & 0 \\ 0 & 8\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx) \end{pmatrix}.$

Lemma 22. *Suppose that Assumptions A1-A8 and Ad hold. Then the following statements hold:*

1. $C_n(\theta_0) \xrightarrow{\mathbb{P}} B = \begin{pmatrix} -2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx) & 0 \\ 0 & 4\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx) \end{pmatrix},$

2. $\sup_{\{|\tilde{\theta}|\le\epsilon_n\}} |C_n(\theta_0 + \tilde{\theta}) - C_n(\theta_0)| \xrightarrow{\mathbb{P}} 0,$ *where* $\epsilon_n \to 0$.

2.7.3.1 Proof of Lemma 21.

Proof. As a consequence of a combination of Theorem 3.2 and Theorem 3.4 in [42] (c.f. also Section A.2 in the Appendix of [86]) we get the result if we prove that $L_n(\theta_0)$ is a triangular array of martingale increments such that, for a constant $r > 0$, the following convergences hold.
We define ζ_i and $\tilde{\zeta}_i$ such that $\partial_\mu U_n(\mu_0, \sigma_0) =: \sum_{i=0}^{n-1}\zeta_i(\theta_0)$ and $\partial_\sigma U_n(\mu_0,\sigma_0) =: \sum_{i=0}^{n-1}\tilde{\zeta}_i(\theta_0)$. Then it must be

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 4\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx) \qquad \frac{1}{(\sqrt{T_n})^{2+r}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{2.55}$$

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^2(\theta_0)] \xrightarrow{\mathbb{P}} 8\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx) \qquad \frac{1}{(\sqrt{n})^{2+r}}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^{2+r}(\theta_0)] \xrightarrow{\mathbb{P}} 0, \tag{2.56}$$

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i[\zeta_i(\theta_0)\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0. \tag{2.57}$$

First of all we observe that $L_n(\theta_0)$ is a triangular array of martingale increments as a consequence of the definitions of m and m_2. Indeed, using (2.38), we clearly have

$$\mathbb{E}_i[\zeta_i(\theta_0)] = \frac{-2\partial_\mu m(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}+$$

$$+\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(1-\frac{(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} = 0.$$

In the same way, computing the derivative with respect to σ we clearly have $\mathbb{E}_i[\tilde{\zeta}_i(\theta_0)] = 0$.
Concerning $\partial_\mu U_n$, using (2.34) we can see $\zeta_i(\theta_0)$ as

$$\frac{-2\partial_\mu m(\mu_0,\sigma_0,X_{t_i})}{\Delta_{n,i}c(X_{t_i},\sigma_0)}(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} + R_{i,n}(\theta_0) =$$

$$=: \hat{\zeta}_i(\theta_0) + R_{i,n}(\theta_0),$$

we have already proved in Lemma 6 of Chapter 1 the asymptotic normality of $\frac{1}{\sqrt{T_n}}\sum_{i=0}^{n-1}\hat{\zeta}_i(\theta_0)$ and, in particular, that convergences (2.55) hold with $\hat{\zeta}_i(\theta_0)$ instead of $\zeta_i(\theta_0)$.
In order to conclude the proof of (2.55), it is enough to have $\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}_i[R^2_{i,n}(\theta_0)] \xrightarrow{\mathbb{P}} 0$ and $(\frac{1}{\sqrt{T_n}})^{2+r}\sum_{i=0}^{n-1}\mathbb{E}_i[R^{2+r}_{i,n}(\theta_0)] \xrightarrow{\mathbb{P}} 0$. It is

$$\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}_i[R^2_{i,n}(\theta_0)] \le$$

$$\le \frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\frac{\partial_\mu m(\mu_0,\sigma_0,X_{t_i})R(\theta,\Delta_{n,i}^{\delta_1\wedge 1}X_{t_i})}{\Delta_{n,i}c(X_{t_i},\sigma_0)})^2\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)]+$$

$$+\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2 1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}+$$

$$+\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\frac{\partial_\mu m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\frac{\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0,\sigma_0,X_{t_i}))^4\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}}{m_2^2(\mu_0,\sigma_0,X_{t_i})}$$

As a consequence of the first and the third point of Proposition 16 and using first and second point of Lemma 14 it is upper bounded by

$$\frac{c}{n}\sum_{i=0}^{n-1}R(\theta_0,\Delta_{n,i}^{2\delta_1\wedge 2},X_{t_i}) + \frac{c}{n\Delta_n}\sum_{i=0}^{n-1}R(\theta_0,\Delta_{n,i}^2,X_{t_i}),$$

which converges to zero in norm 1 and so in probability.
Acting in the same way, using this time the fourth point of Lemma 14 twice, for $k = (2+r)$ and $k = 2(2+r)$, it follows that also $(\frac{1}{\sqrt{T_n}})^{2+r}\sum_{i=0}^{n-1}\mathbb{E}_i[R^{2+r}_{i,n}(\theta_0)]$ goes to zero in probability.
Concerning the derivative of the contrast with respect to σ, it is

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^2(\theta_0)] =$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}(\frac{-2\partial_\sigma m(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}+$$

$$+\frac{1}{n}\sum_{i=0}^{n-1}(\frac{\partial_\sigma m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\mathbb{E}_i[(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}. \tag{2.58}$$

The first term here above goes to zero for $n \to \infty$. Indeed, by the development (2.10) of m_2, the second point of Proposition 16 and equation (2.23) in Lemma 14 it is $\frac{1}{n}\sum_{i=0}^{n-1} R(\theta_0,\Delta_{n,i},X_{t_i})$, that goes to zero in norm 1 because of the property (3.24) of R, its polynomial growth and the third point of Lemma 13. The convergence to zero in probability follows.
The second term of (2.58) instead, using the fourth point of Proposition 16, the development (2.10) of m_2 and (2.34), is

$$\frac{1}{n}\sum_{i=0}^{n-1}[(\frac{2\partial_\sigma a(X_{t_i},\sigma_0)a(X_{t_i},\sigma_0)}{c(X_{t_i},\sigma_0)})^2+ \tag{2.59}$$

$$+R(\theta_0,\Delta_{n,i}^{\beta\wedge\delta_1},X_{t_i})]\mathbb{E}_i[(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)]1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}}.$$

We now need the following lemma:

Lemma 23. *Suppose that Assumptions A1-A4 hold. Then, $\forall q\ge 1$,*

$$\mathbb{E}_i[|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^q] = 1 + R(\theta,\Delta_{n,i},X_{t_i}).$$

Proof. Lemma 23.
We can see $\mathbb{E}_i[|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^q]$ as $1+\mathbb{E}_i[|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^q-1]$.
Because of the definition of φ,the expected value here above is different from zero only if $|\Delta_i X|\ge \Delta_{n,i}^\beta$. Hence, using (2.65) it is

$$\mathbb{E}_i[|\varphi_{\Delta^\beta_{n,i}}(\Delta_i X)|^q-1] \le c\mathbb{E}_i[1_{\{|\Delta_i X|\ge \Delta^\beta_{n,i}\}}] \le R(\theta,\Delta_{n,i},X_{t_i}).$$

□

Using the lemma here above, still the development (2.10) of m_2 and (2.23) and (2.24) in Lemma 14 we have

$$\mathbb{E}_i[(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)] = 1 + R(\theta_0,\Delta_{n,i}^{1\wedge\delta_1},X_{t_i})+$$

$$+3\frac{a^4(X_{t_i},\sigma_0)}{c^2(X_{t_i},\sigma_0)} - 2\frac{a^2(X_{t_i},\sigma_0)}{c(X_{t_i},\sigma_0)} + R(\theta_0,\Delta_{n,i}^{\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}) = 2 + R(\theta_0,\Delta_{n,i}^{\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}),$$

we remind that $c(x,\sigma)=a^2(x,\sigma)$. Replacing the last equation in (2.59) we get

$$\frac{1}{n}\sum_{i=0}^{n-1} 2\cdot\frac{4(\partial_\sigma a(X_{t_i},\sigma_0))^2}{c(X_{t_i},\sigma_0)}1_{\{|X_{t_i}|\le \Delta_{n,i}^{-k}\}} + \frac{1}{n}\sum_{i=0}^{n-1} R(\theta_0,\Delta_{n,i}^{\beta\wedge\delta_1\wedge(\beta-\frac{1}{4})},X_{t_i}).$$

The first term here above converges to $8\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx)$ as a consequence of the third point of Proposition 13 while the second one clearly goes to zero in norm 1 and so in probability thanks to the polynomial growth of R, its property (3.24) and the assumption we made $\beta > \frac{1}{4}$. It follows the first convergence of (2.56). To obtain the second one we observe it is

$$\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}\mathbb{E}_i[\tilde{\zeta}_i^{2+r}(\theta_0)] \leq$$

$$\leq \frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}(\frac{-2\partial_\sigma m(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^{2+r}1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^{2+r}\varphi^{2+r}_{\Delta^\beta_{n,i}}(\Delta_i X)]+$$

$$+\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}(\frac{\partial_\sigma m_2(\mu_0,\sigma_0,X_{t_i})}{m_2(\mu_0,\sigma_0,X_{t_i})})^{2+r}1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}(c+$$

$$+\frac{c}{m_2^{2+r}(\mu_0,\sigma_0,X_{t_i})}\mathbb{E}_i[(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^{2(2+r)}\varphi^{2(2+r)}_{\Delta^\beta_{n,i}}(\Delta_i X)]) \leq$$

$$\leq \Delta_n^{(1+\frac{r}{2})\wedge(1+\beta(2+r))}\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})+$$

$$+\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})+\Delta_n^{0\wedge(1+2\beta(2+r)-(2+r))}\frac{1}{n^{1+\frac{r}{2}}}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i}), \tag{2.60}$$

where we have acted like before using the development (2.10) on m_2 and the second and the fourth point of Proposition 16. Besides, we have used the fourth point of Lemma 14 with $k = 2 + r$ and $k = 2(2 + r)$, respectively. It is now clear that the first two terms of (2.60) go to zero in norm 1 and so in probability for $n \to \infty$. Concerning the third one, if the minimum between 0 and $1 + 2\beta(2 + r) - (2 + r)$ is 0 it is exactly like the second one and so we know it goes to zero in probability, otherwise it can be seen as $\frac{1}{(n\Delta_n)^{\frac{r}{2}}}\Delta_n^{1+2\beta(2+r)-(2+r)+\frac{r}{2}}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta_0,1,X_{t_i})$, that goes to zero since $n\Delta_n \to \infty$ for $n \to \infty$ and because of the fact that the exponent on Δ_n is always positive. Indeed $1 + 2\beta(2 + r) - (2 + r) + \frac{r}{2} > 0$ iff $2\beta(2 + r) > 1 + \frac{r}{2}$, that is $\beta > \frac{1+\frac{r}{2}}{2(2+r)} = \frac{1}{4}$.

To conclude, we prove the convergence (2.57). We have

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i[\zeta_i(\theta_0)\tilde{\zeta}_i(\theta_0)]| =$$

$$= \frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}|\mathbb{E}_i\Big[[(\frac{4\partial_\mu m\,\partial_\sigma m}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2+$$

$$-2(\frac{\partial_\mu m\,\partial_\sigma m_2+\partial_\mu m_2\,\partial_\sigma m}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))\times \tag{2.61}$$

$$\times(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})+$$

$$+\frac{\partial_\mu m_2\,\partial_\sigma m_2}{m_2^2})(\mu_0,\sigma_0,X_{t_i})(1-\frac{(X_{t_{i+1}}-m(\mu_0,\sigma_0,X_{t_i}))^2}{m_2(\mu_0,\sigma_0,X_{t_i})})^2]\varphi^2_{\Delta^\beta_{n,i}}(\Delta_i X)\Big]|1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}.$$

Now using the four points of Proposition 16, the first, second, third and fifth points of Lemma 14 and the lemma here above we get that (2.61) is upper bounded by

$$\frac{1}{n\sqrt{\Delta_n}}\sum_{i=0}^{n-1}R(\theta_0,\Delta_{n,i},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{1}{3}+\beta},X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{4}{3}+\beta},X_{t_i}) \leq$$

$$\leq \Delta_n^{\beta-\frac{1}{6}} \frac{1}{n} \sum_{i=0}^{n-1} R(\theta_0, 1, X_{t_i}),$$

which converges to zero in norm 1 and so in probability since we have chosen $\beta > \frac{1}{4} > \frac{1}{6}$. □

2.7.3.2 Proof of Lemma 22.

Proof. Point 1.
We start showing the convergence of $C_n(\theta_0)$ to B. We observe that

$$\partial^2_{\mu\sigma} U_n(\mu_0, \sigma_0) = \sum_{i=0}^{n-1} [\frac{-2(X_{t_{i+1}} - m)\partial^2_{\mu\sigma} m}{m_2} - 2\frac{\partial_\mu m \partial_\sigma m}{m_2} + \frac{2(X_{t_{i+1}} - m)\partial_\mu m \partial_\sigma m_2}{m_2^2} + \tag{2.62}$$

$$+ \frac{2(X_{t_{i+1}} - m)\partial_\mu m_2 \partial_\sigma m}{m_2^2} + \frac{2(X_{t_{i+1}} - m)^2 \partial_\mu m_2 \partial_\sigma m_2}{m_2^3} + \frac{\partial^2_{\mu\sigma} m_2}{m_2}(1 - \frac{(X_{t_{i+1}} - m)^2}{m_2}) +$$

$$- \frac{\partial_\mu m_2 \partial_\sigma m_2}{m_2^2}] \varphi_{\Delta^\beta_{n,i}}(\Delta X_i) 1_{\left\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\right\}} =: \sum_{j=1}^{7} \sum_{i=0}^{n-1} I^n_{i,j}$$

where, for shortness, we omit that m, m_2 and their derivatives are calculated in $(\mu_0, \sigma_0, X_{t_i})$.
In order to show that $\frac{1}{\sqrt{nT_n}} \partial^2_{\mu\sigma} U_n(\mu_0, \sigma_0) \xrightarrow{\mathbb{P}} 0$ we will use repeatedly Lemma 9 in [36] and the estimation of the derivatives of m and m_2 gathered in Propositions 16 and 17.
$\sum_{i=0}^{n-1} I^n_{i,1}$ goes to zero in probability for $n \to \infty$ because $I^n_{i,1}$ is centered and, using also the first point of Lemma 14

$$\frac{1}{T_n} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}_i[(I^n_{i,1})^2] \leq \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}, X_{t_i}) R(\theta, \Delta_{n,i}, X_{t_i}) \leq$$

$$\leq \frac{1}{n\Delta_n} \frac{1}{n} \Delta_n^2 \sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}),$$

which converges to zero in norm 1 and so in probability as a consequence of the polynomial growth of R and the third point of Lemma 13. $\sum_{i=0}^{n-1} I^n_{i,2}$ goes to zero in norm 1 and so in probability. Indeed, using the development (2.10) of m_2 and the first two point of Proposition 16, it is

$$\mathbb{E}[\frac{1}{\sqrt{nT_n}} | \sum_{i=0}^{n-1} I^n_{i,2}|] \leq \Delta_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[|R(\theta, \Delta_n, X_{t_i})|] \leq c\Delta_n^{\frac{1}{2}},$$

which goes to zero.
Concerning $I^n_{i,3}$, it is still centered and from the first and fourth points of Proposition 16 and the first point of Lemma 14 it follows

$$\frac{1}{T_n} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}_i[(I^n_{i,3})^2] \leq \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}) R(\theta, \Delta_{n,i}, X_{t_i}) \leq \frac{1}{n\Delta_n} \frac{1}{n} \Delta_n \sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}),$$

which converges to 0 in norm 1 and so in probability. Hence, $\frac{1}{\sqrt{nT_n}} \sum_{i=0}^{n-1} I^n_{i,3} \xrightarrow{\mathbb{P}} 0$ from Lemma 9 in [36].

The same applies to $I^n_{i,4}$, which squared is upper bounded by

$$\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}^2, X_{t_i})R(\theta, \Delta_{n,i}, X_{t_i}) \le \frac{1}{n\Delta_n}\frac{1}{n}\Delta_n^3\sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}).$$

On $\sum_{i=0}^{n-1} I^n_{i,5}$ we prove the convergence in norm 1 and so we have the convergence in probability: from the first point of Lemma 14, the development (2.10) of m_2 and the third and the fourth points of Proposition 16, it is

$$\frac{1}{\sqrt{nT_n}}\sum_{i=0}^{n-1}\mathbb{E}[|I^n_{i,5}|] \le \frac{1}{n\sqrt{\Delta_n}}\sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}, X_{t_i})R(\theta, 1, X_{t_i}) \le \Delta_n^{\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}),$$

which goes to zero.
We observe that, as a consequence of the definition of m_2, $I^n_{i,6}$ is centered. In order to apply Lemma 9 in [36] we evaluate its squared value, that is

$$\frac{1}{T_n}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[(I^n_{i,6})^2] \le$$

$$\le \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}, X_{t_i})R(\theta, 1, X_{t_i}) \le \Delta_n\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}),$$

where we have also used the estimation of the mixed second derivative of m_2 contained in Proposition 17.
To conclude the proof about the mixed derivative of the contrast function, we observe that $\sum_{i=0}^{n-1} I^n_{i,7}$ converges to 0 in L^1 from the third and the fourth point of Proposition 16 and the boundedness of φ. We obtain

$$\frac{1}{\sqrt{nT_n}}\mathbb{E}[|\sum_{i=0}^{n-1} I^n_{i,7}|] \le \Delta_n^{-\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta, \Delta_{n,i}, X_{t_i})] \le c\Delta_n^{\frac{1}{2}},$$

that goes to 0 as we wanted.
The next step is to prove that $\frac{1}{T_n}\partial^2_\mu U_n(\mu_0, \sigma_0) \xrightarrow{\mathbb{P}} -2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx)$. In order to do it we compute $\partial^2_\mu U_n(\mu_0, \sigma_0)$ and we observe it is exactly like (2.62) but all the derivatives are with respect to μ. For such a reason we keep referring to (2.62) and we write $\partial^2_\mu U_n(\mu_0, \sigma_0) =: \sum_{j=1}^7\sum_{i=0}^{n-1}\tilde{I}^n_{i,j}$. We are going to show, in particular, that $\frac{1}{T_n}\sum_{i=0}^{n-1}\tilde{I}^n_{i,2}$ converges to the wanted integral, while $\frac{1}{T_n}(\sum_{i=0}^{n-1}\tilde{I}^n_{i,1}+\sum_{j=3}^7\sum_{i=0}^{n-1}\tilde{I}^n_{i,j}) \xrightarrow{\mathbb{P}} 0$. Indeed, we observe that $\tilde{I}^n_{i,1}$, $\tilde{I}^n_{i,3}$, $\tilde{I}^n_{i,4}$, $\tilde{I}^n_{i,6}$ are still centered and, using Lemma 14 and Propositions 16 and 17 it is easy to show that their squared values are upper bounded in the following way:

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}^n_{i,1})^2] \le \frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}, X_{t_i})R(\theta, 1, X_{t_i}) \le \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}),$$

that goes to zero in norm 1 and so in probability since $n\Delta_n \to \infty$ for $n \to \infty$.

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}^n_{i,3})^2] \le \frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, \Delta_{n,i}, X_{t_i})R(\theta, \Delta_{n,i}^2, X_{t_i}) \le$$

$$\le \Delta_n^2\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} R(\theta, 1, X_{t_i}) \xrightarrow{\mathbb{P}} 0.$$

Now $\tilde{I}^n_{i,4}$ and $\tilde{I}^n_{i,3}$ are exactly the same quantity and so the estimation here above clearly holds also for $\tilde{I}^n_{i,4}$ instead of $\tilde{I}^n_{i,3}$. Concerning $\tilde{I}^n_{i,6}$, we have

$$\frac{1}{T_n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[(\tilde{I}^n_{i,6})^2] \le \frac{1}{n\Delta_n}\Delta_n^{-1}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,\Delta_{n,i}^2,X_{t_i})R(\theta,1,X_{t_i}) \le$$

$$\le \Delta_n\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}R(\theta,1,X_{t_i}) \xrightarrow{\mathbb{P}} 0.$$

The application of Lemma 9 in [36] gives us $\frac{1}{T_n}\sum_{i=0}^{n-1}(\tilde{I}^n_{i,1}+\tilde{I}^n_{i,3}+\tilde{I}^n_{i,4}+\tilde{I}^n_{i,6}) \xrightarrow{\mathbb{P}} 0$. We now prove the convergence to 0 in norm 1 of $\frac{1}{T_n}\sum_{i=0}^{n-1}(\tilde{I}^n_{i,5}+\tilde{I}^n_{i,7})$. Indeed, using again the first point of Lemma 14, the development (2.10) of m_2 and the last two points of Proposition 17 it is $\frac{1}{T_n}\sum_{i=0}^{n-1}\mathbb{E}[|\tilde{I}^n_{i,5}+\tilde{I}^n_{i,7}|] \le$

$$\le \frac{1}{n\Delta_n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})R(\theta,\Delta_{n,i},X_{t_i})+R(\theta,\Delta_{n,i}^2,X_{t_i})] \le$$

$$\le \Delta_n\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,1,X_{t_i})] \le c\Delta_n,$$

which clearly goes to 0.
Concerning the principal term $\frac{1}{T_n}\sum_{i=0}^{n-1}\tilde{I}^n_{i,2}$, we observe that using (2.34) and the first point of Proposition 16, it is

$$\frac{(\partial_\mu m)^2}{m_2}(\mu_0,\sigma_0,X_{t_i}) =$$

$$= \frac{(\Delta_{n,i}\partial_\mu b(\mu_0,X_{t_i})+R(\theta_0,\Delta_{n,i}^{\frac{5}{2}-\beta-\epsilon},X_{t_i}))^2}{\Delta_{n,i}c(\sigma_0,X_{t_i})}(1-\Delta_{n,i}^{\delta_1}\frac{r(X_{t_i})}{c(X_{t_i},\sigma_0)}-\Delta_{n,i}r(\mu_0,\sigma_0,X_{t_i})+$$

$$+R(\theta_0,\Delta_{n,i}^{\bar{r}},X_{t_i})) = \frac{\Delta_{n,i}(\partial_\mu b(\mu_0,X_{t_i}))^2}{c(\sigma_0,X_{t_i})}+R(\theta_0,\Delta_{n,i}^{(\frac{5}{2}-\beta-\epsilon)\wedge(1+\delta_1)},X_{t_i}),$$

with $\bar{r} = 2\wedge(1+\delta_2)\wedge 2\delta_1$, as defined below (2.34). In the last equality we have used that the other terms are negligible. Now we have that

$$\frac{1}{T_n}\sum_{i=0}^{n-1}-2\frac{\Delta_{n,i}(\partial_\mu b(\mu_0,X_{t_i}))^2}{c(\sigma_0,X_{t_i})}\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}} \xrightarrow{\mathbb{P}} -2\int_{\mathbb{R}}(\frac{\partial_\mu b(x,\mu_0)}{a(x,\sigma_0)})^2\pi(dx),$$

as a consequence of the second point of Proposition 13 while
$\frac{1}{n\Delta_n}\sum_{i=0}^{n-1}-2R(\theta_0,\Delta_{n,i}^{(\frac{5}{2}-\beta-\epsilon)\wedge(1+\delta_1)},X_{t_i})\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)1_{\{|X_{t_i}|\le\Delta_{n,i}^{-k}\}}$ is upper bounded in norm 1 by

$$\Delta_n^{(\frac{3}{2}-\beta-\epsilon)\wedge\delta_1}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta_0,1,X_{t_i})] \le c\Delta_n^{(\frac{3}{2}-\beta-\epsilon)\wedge\delta_1},$$

that converges to 0 since the exponent on Δ_n is always positive.
To prove the first point of Lemma 22 we are left to show the convergence of $\frac{1}{n}\partial^2_\sigma U_n(\mu_0,\sigma_0)$.
Again, we still refer to (2.62) observing that that the only difference is that all the derivatives are with respect to σ. We write $\partial^2_\sigma U_n(\mu_0,\sigma_0) =: \sum_{j=1}^7\sum_{i=0}^{n-1}\hat{I}^n_{i,j}$.

We keep using Lemma 9 in [36] joint with the development (2.10) and Propositions 16 and 17 to show that the centered terms go to zero in probability, that is

$$\frac{1}{n}\sum_{i=0}^{n-1}(\hat{I}^n_{i,1}+\hat{I}^n_{i,3}+\hat{I}^n_{i,4}+\hat{I}^n_{i,6})\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}\xrightarrow{\mathbb{P}}0.$$

Moreover,

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|\hat{I}^n_{i,2}|]\le\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[R(\theta,\Delta_{n,i},X_{t_i})]\le c\Delta_n\to 0.$$

We are left to deal with the principal terms $\hat{I}^n_{i,5}$ and $\hat{I}^n_{i,7}$ and so we study the convergence of $\frac{1}{n}\sum_{i=0}^{n-1}\frac{(\partial_\sigma m_2)^2}{m_2^2}(\frac{2(X_{t_{i+1}}-m)^2}{m_2}-1)\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}$. From the development (2.10) of m_2 and the development of $\partial_\sigma m_2$ stated in the fourth point of Proposition 16 it follows that the conditional expected value of the quantity here above is $\frac{1}{n}\sum_{i=0}^{n-1}\frac{(2\Delta_{n,i}\partial_\sigma a(X_{t_i},\sigma_0)a(X_{t_i},\sigma_0))^2}{(\Delta_{n,i}a^2(X_{t_i},\sigma_0))^2}(\frac{2(X_{t_{i+1}}-m)^2}{\Delta_{n,i}a^2(X_{t_i},\sigma_0)}-1)1_{\left\{|X_{t_i}|\le\Delta_{n,i}^{-k}\right\}}$ plus a negligible term that comes from the developments (2.10) and (2.34) and it converges to zero in norm 1 and so in probability.
The principal term is therefore such that, using the second point of Proposition 14 and the third of Proposition 13, we get

$$\frac{1}{n}\sum_{i=0}^{n-1}(\hat{I}^n_{i,5}+\hat{I}^n_{i,7})\xrightarrow{\mathbb{P}}4\int_{\mathbb{R}}(\frac{\partial_\sigma a(x,\sigma_0)}{a(x,\sigma_0)})^2\pi(dx).$$

Point 2.
We start proving that $\frac{1}{\sqrt{nT_n}}\sup_{|\tilde{\theta}|\le\epsilon_n}|\partial^2_{\mu\sigma}U_n(\theta_0+\tilde{\theta})-\partial^2_{\mu\sigma}U_n(\theta_0)|$ goes to 0 in probability for ϵ_n that goes to 0.
In order to do that, it is enough to show that the sequence $\frac{1}{\sqrt{nT_n}}\partial^2_{\mu\sigma}U_n(\theta)$ is tight, which is implied by $\sup_n\frac{1}{\sqrt{nT_n}}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_{\mu\sigma}U_n(\mu,\sigma))|]<\infty$ (see Corollary B.1 in [86]), for $\vartheta=\mu$ or $\vartheta=\sigma$.
We observe it is

$$\partial^3_{\mu\sigma\vartheta}U_n(\mu,\sigma):=\sum_{i=0}^{n-1}[\frac{2\partial_\vartheta m\partial^2_{\mu\sigma}m-2(X_{t_{i+1}}-m)\partial^3_{\mu\sigma\vartheta}m}{m_2}+\frac{2(X_{t_{i+1}}-m)\partial^2_{\mu\sigma}m\partial_\vartheta m_2}{m_2^2}+$$

$$-2\frac{(\partial^2_{\mu\vartheta}m\partial_\sigma m+\partial_\mu m\partial^2_{\sigma\vartheta}m)}{m_2}+\frac{2\partial_\mu m\partial_\sigma m\partial_\vartheta m_2}{m_2^2}-\frac{2\partial_\vartheta m\partial_\mu m\partial_\sigma m_2}{m_2^2}+$$

$$+\frac{2(X_{t_{i+1}}-m)(\partial^2_{\mu\vartheta}m\partial_\sigma m_2+\partial_\mu m\partial^2_{\sigma\vartheta}m_2)}{m_2^2}-\frac{4(X_{t_{i+1}}-m)\partial_\mu m\partial_\sigma m_2\partial_\vartheta m_2}{m_2^3}+$$

$$-\frac{2\partial_\vartheta m\partial_\mu m_2\partial_\sigma m}{m_2^2}+\frac{2(X_{t_{i+1}}-m)(\partial^2_{\mu\vartheta}m_2\partial_\sigma m+\partial_\mu m_2\partial^2_{\sigma\vartheta}m)}{m_2^2}-\frac{4(X_{t_{i+1}}-m)\partial_\mu m_2\partial_\sigma m\partial_\vartheta m_2}{m_2^3}+$$

$$-\frac{4(X_{t_{i+1}}-m)\partial_\vartheta m\partial_\mu m_2\partial_\sigma m_2}{m_2^3}+\frac{2(X_{t_{i+1}}-m)^2(\partial^2_{\mu\vartheta}m_2\partial_\sigma m_2+\partial_\mu m_2\partial^2_{\sigma\vartheta}m_2)}{m_2^3}+$$

$$-\frac{6(X_{t_{i+1}}-m)^2\partial_\mu m_2\partial_\sigma m_2\partial_\vartheta m_2}{m_2^4}+(\frac{\partial^3_{\mu\sigma\vartheta}m_2}{m_2}-\frac{\partial^2_{\mu\sigma}m_2\partial_\vartheta m_2}{m_2^2})(1-\frac{(X_{t_{i+1}}-m)^2}{m_2})+$$

$$+\frac{\partial^2_{\mu\sigma}m_2}{m_2}(\frac{2(X_{t_{i+1}}-m)\partial_\vartheta m}{m_2}+\frac{(X_{t_{i+1}}-m)^2\partial_\vartheta m_2}{m_2^2})+$$

$$-\frac{\partial^2_{\mu\vartheta}m_2\partial_\sigma m_2+\partial_\mu m_2\partial^2_{\sigma\vartheta}m_2}{m_2^2}+\frac{2\partial_\mu m_2\partial_\sigma m_2\partial_\vartheta m_2}{m_2^3}]\varphi_{\Delta^\beta_{n,i}}(\Delta X_i)1_{\{|X_{t_i}|\le\Delta^{-k}_{n,i}\}}$$

$$=:\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\vartheta).$$

Now using Assumption Ad, the estimation on the derivatives of m and m_2 gathered in Propositions 16, 17 and 18 and the inequalities (2.40), (2.41) and (2.42) it follows

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu(\partial^2_{\mu\sigma}U_n(\mu,\sigma))|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\mu)|]\le c\,n\Delta_n^{\frac{1}{2}}.$$

In the same way we get

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_{\mu\sigma}U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\sigma)|]\le c\,n\Delta_{n,i}^{\frac{1}{2}}.$$

Hence, for both $\vartheta=\mu$ and $\vartheta=\sigma$ we can say it is

$$\sup_n\frac{1}{\sqrt{nT_n}}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_{\mu\sigma}U_n(\mu,\sigma)|]\le c<\infty;$$

that implies the tightness of our sequence and so that $\frac{1}{\sqrt{nT_n}}\sup_{|\tilde\theta|\le\epsilon_n}|\partial^2_{\mu\sigma}U_n(\theta_0+\tilde\theta)-\partial^2_{\mu\sigma}U_n(\theta_0)|$ goes to 0 in probability for ϵ_n that goes to 0.
To prove the convergence to 0 in probability of $\frac{1}{n}\sup_{|\tilde\theta|\le\epsilon_n}|\partial^2_\sigma U_n(\theta_0+\tilde\theta)-\partial^2_\sigma U_n(\theta_0)|$ for ϵ_n that goes to 0 we act in the same way: we show that the sequence $\frac{1}{n}\partial^2_\sigma U_n(\theta)$ is tight through the criterion $\sup_n\frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma))|]<\infty$.
We observe that, computing the derivative with respect to ϑ of $\partial^2_\sigma U_n(\mu,\sigma)$, we obtain 17 terms analogous to the case just studied, with the only difference that also the derivatives that were with respect to μ are now with respect to σ. In particular, it is $\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma))=\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\vartheta)$.
We still use Assumption Ad, the estimation on the derivatives of m and m_2 gathered in Propositions 16, 17 and 18 and the inequalities (2.40), (2.41) and (2.42) to prove that $\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_\sigma U_n(\mu,\sigma)|]<\infty$. Indeed, it is

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\sigma(\partial^2_\sigma U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\sigma)|]\le c.$$

Moreover, since the order in which we compute the derivatives of the contrast function commute, we have

$$\mathbb{E}[\sup_{\mu,\sigma}|\partial_\mu(\partial^2_\sigma U_n(\mu,\sigma)|]=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\sigma,\sigma,\mu)|]=$$

$$=\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\sum_{j=1}^{17}I_{i,j}(\mu,\sigma,\sigma)|]\le\sum_{i=0}^{n-1}c\Delta_{n,i}^{\frac{1}{2}}.$$

We can therefore say that, for both $\vartheta=\mu$ and $\vartheta=\sigma$, it is

$$\sup_n\frac{1}{n}\mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta(\partial^2_\sigma U_n(\mu,\sigma)|]\le c<\infty;$$

that implies the tightness.
We are now left to show that $\frac{1}{T_n}\sup_{|\tilde{\theta}|\le\epsilon_n}|\partial^2_\mu U_n(\theta_0+\tilde{\theta})-\partial^2_\mu U_n(\theta_0)|\xrightarrow{\mathbb{P}} 0$ for $\epsilon_n\to 0$.
We still consider the notation introduced in the first point for which $\partial^2_\mu U_n(\theta) =: \sum_{j=1}^7\sum_{i=0}^{n-1}\tilde{I}^n_{i,j}(\theta)$ with $\partial^2_\mu U_n(\theta)$ that is as in (2.62) but both the derivatives are calculated with respect to μ.
From Proposition 15 we already know that $\frac{1}{T_n}\sum_{i=0}^{n-1}\tilde{I}^n_{i,j}(\theta)$ are tight sequences for $j\in\{1,3,4\}$; having taken as $g_{i,n}(\theta,X_{t_i})$ respectively $\frac{-2\partial^2_\mu m(\mu,\sigma,X_{t_i})}{m_2(\mu,\sigma,X_{t_i})}$ and $\frac{2\partial_\mu m(\mu,\sigma,X_{t_i})\partial_\mu m_2(\mu,\sigma,X_{t_i})}{m_2^2(\mu,\sigma,X_{t_i})}$, twice. We see that the assumptions required on $g_{i,n}$ hold as a consequence of the estimation on the derivatives of m and m_2 gathered in Propositions 16 and 17 and the development Ad of m_2.
We also show the tightness of the other terms proving that, for both $\vartheta=\mu$ and $\vartheta=\sigma$,
$\sup_n\frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\partial_\vartheta(\tilde{I}^n_{i,2}+\tilde{I}^n_{i,5}+\tilde{I}^n_{i,6}+\tilde{I}^n_{i,7})|]\le c$. Indeed, using the estimation on the first derivatives of m and m_2 gathered in Proposition 16 and the development Ad of m_2 it is

$$\sup_n\frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\partial_\mu\tilde{I}^n_{i,2}|]\le\sup_n\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\Delta_{n,i}+\Delta^2_{n,i})\le c,$$

$$\sup_n\frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\partial_\sigma\tilde{I}^n_{i,2}|]\le\sup_n\frac{c}{n\Delta_n}\sum_{i=0}^{n-1}(\Delta_{n,i}+\Delta_{n,i})\le c.$$

In the same way, using Assumption Ad, the estimation on the derivatives of m and m_2 gathered in Propositions 16, 17 and 18 and the inequalities (2.40), (2.41) and (2.42) it follows

$$\sup_n\frac{1}{T_n}\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\partial_\mu(\tilde{I}^n_{i,5}+\tilde{I}^n_{i,6}+\tilde{I}^n_{i,7})|]\le\sup_n\frac{c}{n\Delta_n}(\Delta_n^{\frac{5}{2}\wedge 2\wedge 3}+\Delta_n^{1\wedge 2\wedge\frac{3}{2}\wedge 2}+\Delta_n^{2\wedge 3})\le c;$$

$$\sup_n\frac{1}{n\Delta_n}\mathbb{E}[\sup_{\mu,\sigma}|\sum_{i=0}^{n-1}\partial_\sigma(\tilde{I}^n_{i,5}+\tilde{I}^n_{i,6}+\tilde{I}^n_{i,7})|]\le\sup_n\frac{c}{n\Delta_n}(\Delta_n^{\frac{5}{2}\wedge 2\wedge 2}+\Delta_n^{1\wedge 1\wedge\frac{3}{2}\wedge 2}+\Delta_n^{1\wedge 2})\le c.$$

We have therefore proved that the sequence $\frac{1}{T_n}\partial^2_\mu U_n(\theta)$ is tight, which implies the convergence to zero in probability of $\frac{1}{T_n}\sup_{|\tilde{\theta}|\le\epsilon_n}|\partial^2_\mu U_n(\theta_0+\tilde{\theta})-\partial^2_\mu U_n(\theta_0)|$. □

2.7.3.3 Proof of Theorem 9.

Proof. By (2.54) we get

$$(\int_0^1[C_n(\theta_0+u(\hat{\theta}_n-\theta_0))-C_n(\theta_0)]du+C_n(\theta_0))S_n=L_n(\theta_0).$$

We find that the matrix

$$\int_0^1[C_n(\theta_0+u(\hat{\theta}_n-\theta_0))-C_n(\theta_0)]du+C_n(\theta_0) \qquad (2.63)$$

converges in probability to the nonsingular matrix B. Hence, taking the limit on both sides after multiplying by the inverse of (2.63), we see by the continuous mapping theorem that $S_n\xrightarrow{d}B^{-1}L\sim N(0,K^{-1})$.
The asymptotic normality of S_n is therefore proved. □

2.7.4 Proof of Proposition 9

The proof of the proposition is essentially similar to the proof of the asymptotic normality of the estimator $\hat{\theta}_n$ given in Sections 2.7.1−2.7.3 and we skip it. The main difference comes from the fact that, the first derivative of the contrast function is no longer centered and, in order to prove Lemma 21, we need also that $\frac{1}{T_n}\sum_{i=0}^{n-1}|\mathbb{E}[\zeta_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0$ and $\frac{1}{n}\sum_{i=0}^{n-1}|\mathbb{E}[\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0$, where we have kept the notation of Lemma 21.
The first convergence here above holds true for $\tilde{m}$ and $\tilde{m}_2$ replacing m and m_2 if $\sqrt{n}\Delta_n^{\rho_1-\frac{1}{2}} \to 0$, $\sqrt{n}\Delta_n^{\rho_2-1} \to 0$ and $\sqrt{n}\Delta_n^{2\rho_1-\frac{1}{2}} \to 0$; which happens under our hypothesis.
Moreover we have that $\frac{1}{n}\sum_{i=0}^{n-1}|\mathbb{E}[\tilde{\zeta}_i(\theta_0)]| \xrightarrow{\mathbb{P}} 0$ with $\tilde{m}$ and $\tilde{m}_2$ replacing m and m_2 for $\rho_1 > \frac{1}{2}$ and $\rho_2 > 1$, which is always true in our setting.

2.8 Appendix

In this section we prove all the technical results we have introduced, starting from the preliminary results stated in Section 2.6.

2.8.1 Proof of limit theorems

We first show Proposition 13, observing that its last two points are the discretized version of the first point of Lemma 13.

2.8.1.1 Proof of Proposition 13

Proof. The first two points have already been proved in Proposition 3 of Chapter 1.
We want to show that $\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)$ converges in L^2 to $\int_{\mathbb{R}} f(x,\theta)\pi(dx)$.
Since $Var(\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)) \le \frac{1}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} Cov(f(X_{t_i},\theta), f(X_{t_j},\theta))$, we need to estimate the covariance.
We know that, under our assumptions, the process X is β- mixing with exponential decay (see [67]) that is $\exists \gamma > 0$ such that $\beta_X(k) = O(e^{-\gamma k})$; with $\beta_X(k)$ as defined in Section 1.3.2 of [27]. If a process is β- mixing, then it is also α-mixing and so the following estimation holds (see Theorem 3 in Section 1.2.2 of [27])

$$|Cov(X_{t_i}, X_{t_j})| \le c\,\|X_{t_i}\|_p \left\|X_{t_j}\right\|_q \alpha^{\frac{1}{r}}(X_{t_i}, X_{t_j})$$

with p, q and r such that $\frac{1}{p}+\frac{1}{q}+\frac{1}{r} = 1$. Using that $\alpha(X_{t_i}, X_{t_j}) \le \beta_X(|t_i - t_j|) = O(e^{-\gamma|t_i-t_j|})$, in our case the inequality here above becomes

$$|Cov(f(X_{t_i},\theta), f(X_{t_j},\theta))| \le c e^{-\frac{1}{r}\gamma|t_i-t_j|},$$

where we have also used the polynomial growth of f and the third point of Lemma 13 to include the two norms in the constant c.
We introduce a partition of $(0, T_n]$ based on the sets $A_k := (k\frac{T_n}{n}, (k+1)\frac{T_n}{n}]$, for which $(0,T_n] = \cup_{k=0}^{n-1} A_k$. Now each point t_i in $(0, T_n]$ can be seen as $t_{k,h}$, where k identifies the particular set A_k to which the point belongs while, defining M_k as $|A_k|$, h is a number in $\{1, ..., M_k\}$ which enumerates the points in each set. It follows

$$\frac{c}{n^2}\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} e^{-\frac{1}{r}\gamma|t_i-t_j|} \le \frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}} e^{-\frac{1}{r}\gamma|t_{k_1,h_1}-t_{k_2,h_2}|} \le$$

$$\leq \frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1}\sum_{h_1=1}^{M_{k_1}}\sum_{h_2=1}^{M_{k_2}} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}},$$

where the last inequality is a consequence of the following estimation: for each $k_1, k_2 \in \{0, ..., n-1\}$ it is $|t_{k_1,h_1} - t_{k_2,h_2}| \geq |k_1 - k_2|\frac{T_n}{n} - \frac{T_n}{n}$.
Now we observe that the exponent does not depend on h anymore, hence the last term here above can be upper bounded by $\frac{ce^{\frac{1}{r}\frac{T_n}{n}}}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} M_{k_1}M_{k_2}e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}}$.
Moreover, remarking that the length of each interval A_k is $\frac{T_n}{n}$, it is easy to show that we can always upper bound M_k with $\frac{T_n}{n}\frac{1}{\Delta_{min}}$, with $T_n = \sum_{i=0}^{n-1}\Delta_{n,i} \leq n\Delta_n$ and so $M_k \leq \frac{\Delta_n}{\Delta_{min}}$, that we have assumed bounded by a constant c_1.
Furthermore, still using that $T_n \leq n\Delta_n$, we have $e^{\frac{1}{r}\frac{T_n}{n}} \leq e^{\frac{1}{r}\Delta_n} \leq c$ because, by our hypothesis, Δ_{max} goes to 0 for $n \to \infty$. To conclude, we have to show that $\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}} \to 0$ for $n \to \infty$. We define $j := k_1 - k_2$ and we apply a change of variable, getting

$$\frac{c}{n^2}\sum_{k_1=0}^{n-1}\sum_{k_2=0}^{n-1} e^{-\frac{1}{r}\gamma|k_1-k_2|\frac{T_n}{n}} \leq \frac{c}{n^2}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\frac{T_n}{n}}|n-j| \leq$$

$$\leq \frac{c}{n}\sum_{j=-(n-1)}^{n-1} e^{-\frac{1}{r}\gamma|j|\Delta_{min}} \leq \frac{c}{n(1-e^{-\frac{1}{r}\gamma\Delta_{min}})} \leq \frac{c}{T_n},$$

that goes to 0 for n that goes to ∞. We therefore get

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \xrightarrow{\mathbb{P}} 0.$$

In order to show the third point we observe that

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} - \int_{\mathbb{R}} f(x,\theta)\pi(dx)| \leq |\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}} +$$

$$-\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)| + |\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta) - \int_{\mathbb{R}} f(x,\theta)\pi(dx)|.$$

We have already proved that the second goes to 0 in probability, while the first is $|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\{|X_{t_i}|>\Delta_{n,i}^{-k}\}}|$, that converges to 0 in L^1 as a consequence of the polynomial growth of f and the third point of Lemma 13 and so in probability.
We act in the same way in order to show the fourth point, observing that, by the definition of φ, it is

$$|\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i},\theta)1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}(\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i) - 1)| \leq \tag{2.64}$$

$$\leq \frac{c}{n}\sum_{i=0}^{n-1}|f(X_{t_i},\theta)|1_{\{|X_{t_i}|\leq\Delta_{n,i}^{-k}\}}1_{\{|X_{t_i}|\geq\Delta_{n,i}^{\beta}\}}.$$

We observe that, since $\Delta X_i^c = \Delta X_i - \Delta X_i^J$, if $|\Delta X_i| \geq \Delta_{n,i}^{\beta}$ and $|\Delta X_i^J| < \frac{\Delta_{n,i}^{\beta}}{2}$, then $|\Delta X_i^c|$ must be more than $\frac{\Delta_{n,i}^{\beta}}{2}$. Hence

$$\mathbb{E}[1_{\{|\Delta X_i|\geq\Delta_{n,i}^{\beta}\}}] = \mathbb{E}[1_{\left\{|\Delta X_i|\geq\Delta_{n,i}^{\beta}, |\Delta X_i^J|<\frac{\Delta_{n,i}^{\beta}}{2}\right\}}] + \mathbb{E}[1_{\left\{|\Delta X_i|\geq\Delta_{n,i}^{\beta}, |\Delta X_i^J|\geq\frac{\Delta_{n,i}^{\beta}}{2}\right\}}] \leq$$

$$\leq \mathbb{P}(|\Delta X_i^c| \geq \frac{\Delta_{n,i}^{\beta}}{2}) + \mathbb{P}(|\Delta X_i^J| \geq \frac{\Delta_{n,i}^{\beta}}{2}) \leq c\frac{\mathbb{E}[|\Delta X_i^c|^r]}{\Delta_{n,i}^{\beta r}} + R(\theta, \Delta_{n,i}, X_{t_i}) \qquad (2.65)$$

$$\leq R(\theta, \Delta_{n,i}^{(\frac{1}{2}-\beta)r\wedge 1}, X_{t_i}) = R(\theta, \Delta_{n,i}, X_{t_i}).$$

On the first probability here above we have used Tchebychev inequality and the fourth point of Lemma 25, for $|\Delta X_i^J| \geq \frac{\Delta_{n,i}^{\beta}}{2}$ the fact that the intensity of jumps is finite and therefore the probability to have at least one jump bigger than $\frac{\Delta_{n,i}^{\beta}}{2}$ can be computed and it is of order Δ_n. Moreover, by the arbitrariness of $r > 1$, we get that $R(\theta, \Delta_{n,i}^{(\frac{1}{2}-\beta)r}, X_{t_i})$ is negligible compared to $R(\theta, \Delta_{n,i}, X_{t_i})$.
From Holder inequality, the polynomial growth of f, the third point of Lemma 13 and (2.65) it follows that the right hand side of (2.64) goes to 0 in norm 1 and so in probability. The proposition is therefore proved. □

We now prove Proposition 14, that is a consequence of Lemma 14.

2.8.1.2 Proof of Proposition 14

Proof. In order to show that the first convergence holds, we define
$s_i^n := \frac{1}{T_n} f(X_{t_i}, \theta)\, (X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2\, \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}$. From Lemma 9 in [36], if we show that

$$\sum_{i=0}^{n-1} \mathbb{E}_i[s_i^n] \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} f(x, \theta) a^2(x, \sigma_0) \pi(dx) \quad \text{and} \quad \sum_{i=0}^{n-1} \mathbb{E}_i[(s_i^n)^2] \to 0,$$

then the proposition is proved. We observe that (2.23) yields

$$\sum_{i=0}^{n-1} \mathbb{E}_i[s_i^n] =$$

$$= \frac{1}{T_n} \sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i}, \theta) a^2(X_{t_i}, \sigma_0) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}} + \frac{1}{T_n} \sum_{i=0}^{n-1} \Delta_{n,i} f(X_{t_i}, \theta) R(\theta_0, \Delta_{n,i}^{\beta}, X_{t_i}).$$

The first term here above converges in probability to $\int_{\mathbb{R}} f(x, \theta) a^2(x, \sigma_0) \pi(dx)$ as a consequence of the first point of Proposition 13, while the second is upper bounded by
$\Delta_n^{\beta} \frac{1}{n} \sum_{i=0}^{n-1} f(X_{t_i}, \theta) R(\theta_0, 1, X_{t_i})$, which converges to zero in norm 1 (and so in probability) by the polynomial growth of both R and f and the third point of Lemma 13. Moreover, using (2.24) and the fact that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$, we have

$$\sum_{i=0}^{n-1} |\mathbb{E}_i[(s_i^n)^2]| \leq \frac{1}{(n\Delta_n)^2} \sum_{i=0}^{n-1} (f(X_{t_i}, \theta))^2 R(\theta_0, \Delta_{n,i}^2, X_{t_i}) \leq \frac{1}{n^2} \sum_{i=0}^{n-1} f^2(X_{t_i}, \theta) R(\theta_0, 1, X_{t_i}),$$

which goes to zero in norm 1 and so in probability for $n \to \infty$ as a consequence of the polynomial growth of both R and f and the third point of Lemma 13. The first point is therefore proved. In order to show the second point of Proposition 14 is enough to act on the sequence $\tilde{s}_i^n := \frac{1}{n} \frac{f(X_{t_i}, \theta)}{\Delta_{n,i}} (X_{t_{i+1}} - m(\mu, \sigma, X_{t_i}))^2\, \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) 1_{\{|X_{t_i}| \leq \Delta_{n,i}^{-k}\}}$ exactly like we have just did here above, with the only difference that the third point of Proposition 13 has to be applied instead of the first one. □

2.8.1.3 Proof of Lemma 14

Proof. Replacing $m(\mu,\sigma,X_{t_i})$ with its development (2.31) and using the dynamic (4.3) of X we have

$$X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}) =$$

$$= \int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds + \int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s + \int_{t_i}^{t_{i+1}}\int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds,dz) + R(\theta,\Delta_{n,i},X_{t_i}) =$$

$$=: \int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s + B_{i,n}. \tag{2.66}$$

In order to prove (2.23) we start considering

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu,\sigma,X_{t_i}))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})] =$$

$$= \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})] +$$

$$+\mathbb{E}_i[(B_{i,n})^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})] + 2\mathbb{E}_i[B_{i,n}(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})].$$

Now the first term on the right hand side here above is

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - 1)] =$$

$$= \Delta_{n,i}a^2(X_{t_i},\sigma_0) + \mathbb{E}_i[\int_{t_i}^{t_{i+1}}[a^2(X_s,\sigma_0) - a^2(X_{t_i},\sigma_0)]ds] + \tag{2.67}$$

$$+\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - 1)].$$

Moreover,

$$|\mathbb{E}_i[\int_{t_i}^{t_{i+1}}[a^2(X_s,\sigma_0) - a^2(X_{t_i},\sigma_0)]ds] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2(\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) - 1)]| \le$$

$$\le \int_{t_i}^{t_{i+1}} \mathbb{E}_i[|2a\partial_x a(X_u,\sigma_0)||X_s - X_{t_i}|]ds + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^{2p}]^{\frac{1}{p}}\mathbb{E}_i[1_{\left\{|\Delta X_i| \ge \Delta_{n,i}^\beta\right\}}]^{\frac{1}{q}}, \tag{2.68}$$

where $X_u \in (X_s, X_{t_i})$ and we have used Holder inequality and the definition of φ, that is equal to 1 for $|\Delta X_i| < \Delta_{n,i}^\beta$.
Using Cauchy-Schwartz inequality and the second point of Lemma 25 on the first term of (2.68) and Burkholder-Davis-Gundy inequality and (2.65) on the second we have that the right hand side is upper bounded by

$$\int_{t_i}^{t_{i+1}} R(\theta_0,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})ds + R(\theta_0,\Delta_{n,i},X_{t_i})R(\theta_0,\Delta_{n,i}^{\frac{1}{q}},X_{t_i}) \le$$

$$\le R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{2-\epsilon},X_{t_i}),$$

where we have taken q next to 1.
Replacing in (2.67) we get

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})] = \Delta_{n,i}a^2(X_{t_i},\sigma_0) + R(\theta_0,\Delta_{n,i}^{\frac{3}{2}},X_{t_i}). \tag{2.69}$$

Now we evaluate the contribution of $(B_{i,n})^2$:

$$\mathbb{E}_i[(B_{i,n})^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] \le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} b(X_s,\mu_0)ds)^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]+$$

$$+c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}\gamma(X_{s^-})z\tilde{\mu}(ds,dz))^2\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]+R(\theta,\Delta_{n,i}^2,X_{t_i}) \le$$

$$\le c\Delta_{n,i}\int_{t_i}^{t_{i+1}}\mathbb{E}_i[b^2(X_s,\mu)]ds + +R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}) =$$

$$= R(\theta_0,\Delta_{n,i}^2,X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i}), \tag{2.70}$$

where we have used Jensen inequality, Lemma 16 and the fact that $R(\theta_0,\Delta_{n,i}^2,X_{t_i})$ is always negligible compared to $R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i})$ since $2>1+2\beta$.
We observe that (2.69) still holds with 1 instead of φ (see (2.67)). Using it, Cauchy-Schwartz inequality and (2.70) it follows

$$\mathbb{E}_i[B_{i,n}(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})] \le$$

$$\le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^2]^{\frac{1}{2}}\mathbb{E}_i[(B_{i,n})^2\varphi^2_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i})]^{\frac{1}{2}} \le$$

$$\le R(\theta_0,\Delta_{n,i},X_{t_i})^{\frac{1}{2}}R(\theta_0,\Delta_{n,i}^{1+2\beta},X_{t_i})^{\frac{1}{2}} = R(\theta_0,\Delta_{n,i}^{1+\beta},X_{t_i}). \tag{2.71}$$

From (2.66), (2.69) - (2.71) it follows (2.23).

Concerning (2.24), we have

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)] = \tag{2.72}$$

$$= \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4(\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-1)].$$

Using Holder inequality and the definition of φ we have that the second term here above is upper bounded by

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^{4p}]^{\frac{1}{p}}\mathbb{P}_i(|\Delta X_i| \ge \Delta_{n,i}^\beta)^{\frac{1}{q}} \le \tag{2.73}$$

$$R(\theta_0,\Delta_{n,i}^2,X_{t_i})R(\theta_0,\Delta_{n,i}^{\frac{1}{q}},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{3-\epsilon},X_{t_i}),$$

where we have used BDG inequality, (2.65) and we have taken q next to 1. Moreover,

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4] = \tag{2.74}$$

$$= \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_{t_i},\sigma_0)dW_s)^4] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4]+$$

$$+\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_{t_i},\sigma_0)dW_s)^j(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^{4-j}].$$

Since the expected value of the fourth moment of the gaussian law is known we have

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_{t_i},\sigma_0)dW_s)^4] = 3\Delta_{n,i}^2a^4(X_{t_i},\sigma_0). \tag{2.75}$$

On the second term of the right hand side of (2.74) we use again BDG inequality to get

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4]\le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]^2ds)^2]\le$$

$$\le c\Delta_{n,i}\int_{t_i}^{t_{i+1}}\|\partial_x a\|_\infty^4\,\mathbb{E}_i[|X_s-X_{t_i}|^4]ds\le \tag{2.76}$$

$$c\Delta_{n,i}\int_{t_i}^{t_{i+1}}|s-t_i|(1+|X_{t_i}|^4)ds\le R(\theta_0,\Delta_{n,i}^3,X_{t_i}),$$

where we have also used Jensen inequality and the second point of Lemma 25.
Concerning the last term in the right hand side of (2.74), from Holder inequality it is upper bounded by

$$\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^{jp_1}]^{\frac{1}{p_1}}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^{(4-j)p_2}]^{\frac{1}{p_2}}.$$

Now we take $p_1=\frac{4}{j}$ and so $p_2=\frac{4}{4-j}$. Therefore, using also (2.75) and (2.76), the expression here above is

$$\sum_{j=1}^3\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_{t_i},\sigma_0)dW_s)^4]^{\frac{j}{4}}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}[a(X_s,\sigma_0)-a(X_{t_i},\sigma_0)]dW_s)^4]^{\frac{4-j}{4}}\le$$

$$\le\sum_{j=1}^3\binom{4}{j}R(\theta_0,\Delta_{n,i}^2,X_{t_i})^{\frac{j}{4}}R(\theta_0,\Delta_{n,i}^3,X_{t_i})^{\frac{4-j}{4}}\le$$

$$\le\sum_{j=1}^3\binom{4}{j}R(\theta_0,\Delta_{n,i}^{3-\frac{j}{4}},X_{t_i})=R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i}), \tag{2.77}$$

since when $j=1$ and $j=2$ we get terms that are negligible if compared to $R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i})$.
Replacing (2.73), (2.75) - (2.77) in (2.72) it follows

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}a(X_s,\sigma_0)dW_s)^4\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)]=3\Delta_{n,i}^2a^4(X_{t_i},\sigma_0)+R(\theta_0,\Delta_{n,i}^{\frac{9}{4}},X_{t_i}). \tag{2.78}$$

We now study the contribution of $B_{i,n}$. First,

$$\mathbb{E}_i[(B_{i,n})^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|]\le c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}b(X_s,\mu_0)ds)^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|]+$$

$$+c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}\gamma(X_{s^-})z\tilde{\mu}(ds,dz))^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|]+R(\theta,\Delta_{n,i}^4,X_{t_i})\le$$

$$\le c\Delta_{n,i}^3\int_{t_i}^{t_{i+1}}\mathbb{E}_i[b^4(X_s,\mu_0)]ds+R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})=$$

$$=R(\theta_0,\Delta_{n,i}^4,X_{t_i})+R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})=R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i}), \tag{2.79}$$

where we have used Jensen inequality, Lemma 16 and the fact that $R(\theta_0,\Delta_{n,i}^4,X_{t_i})$ is always negligible compared to $R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})$ since $4>1+4\beta$.
Using (2.66) we have that

$$\mathbb{E}_i[(X_{t_{i+1}}-m(\mu,\sigma,X_{t_i}))^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|]= \tag{2.80}$$

$$= \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|] + \mathbb{E}_i[(B_{i,n})^4|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|]+$$

$$+\sum_{j=1}^{3}\binom{4}{j}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)^j(B_{i,n})^{4-j}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|].$$

On the last term here above we act like we did in (2.71), using holder inequality and taking $p_1 = \frac{4}{j}$. It follows, using also (2.74), (2.75), (2.76), (2.77) and (2.79), that it is upper bounded by

$$\sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^2,X_{t_i})^{\frac{j}{4}}R(\theta_0,\Delta_{n,i}^{1+4\beta},X_{t_i})^{\frac{4-j}{4}} = \sum_{j=1}^{3}\binom{4}{j}R(\theta_0,\Delta_{n,i}^{1+4\beta+\frac{j}{4}(1-4\beta)},X_{t_i}). \tag{2.81}$$

Since we have chosen $\beta > \frac{1}{4}$, the terms in which $j = 1, 2$ are negligible compared to the one in which $j = 3$ and so we get $R(\theta_0,\Delta_{n,i}^{\frac{7}{4}+\beta},X_{t_i})$. From (2.78)- (2.81) it follows (2.24).

In order to show (2.25) we start considering $B_{i,n}$:

$$|\mathbb{E}_i[B_{i,n}\varphi^k_{\Delta_{n,i}^\beta}(\Delta X_i)]| \le R(\theta_0,\Delta_{n,i},X_{t_i}) + c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} b(X_s,\mu)ds|]+$$

$$+c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds,dz)|] \le R(\theta_0,\Delta_{n,i},X_{t_i})+$$

$$+c\mathbb{E}_i[\int_{t_i}^{t_{i+1}}|b(X_s,\mu)|ds] + c\mathbb{E}_i[\int_{t_i}^{t_{i+1}}(\int_{\mathbb{R}}|z|F(z)dz)|\gamma(X_{s^-})|ds] \le R(\theta_0,\Delta_{n,i},X_{t_i}), \tag{2.82}$$

having used the definition of $B_{i,n}$ given in (2.66), the boundedness of φ^k, the polynomial growth of both b and γ and the third point of Lemma 25.
Moreover,

$$|\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)\varphi^k_{\Delta_{n,i}^\beta}(\Delta X_i)]| =$$

$$= |\mathbb{E}_i[\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s)(\varphi^k_{\Delta_{n,i}^\beta}(\Delta X_i)-1)]| \le$$

$$\le R(\theta_0,\Delta_{n,i}^{\frac{1}{2}},X_{t_i})\mathbb{E}_i[1_{\{|\Delta X_i|\ge\Delta_{n,i}^\beta\}}]^{\frac{1}{q}} \le R(\theta_0,\Delta_{n,i}^{\frac{3}{2}-\epsilon},X_{t_i}), \tag{2.83}$$

where we have used (2.65) and taken q next to 1. From the inequality here above and (2.82) it follows (2.25).

Concerning (2.26); we have

$$\mathbb{E}_i[|X_{t_{i+1}} - m(\mu,\sigma,X_{t_i})|^k|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^{k'}] \le$$

$$\le c\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} a(X_s,\sigma_0)dW_s|^k|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^{k'}] + c\mathbb{E}_i[|B_{i,n}|^k|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^{k'}] \le$$

$$\le R(\theta_0,\Delta_{n,i}^{\frac{k}{2}},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{k},X_{t_i}) + R(\theta_0,\Delta_{n,i}^{1+k\beta},X_{t_i}) = R(\theta_0,\Delta_{n,i}^{\frac{k}{2}\wedge(1+k\beta)},X_{t_i}),$$

where we have used on the first term here above the fact that φ is bounded and BDG inequality while on the second we have acted like we did in (2.70) or (2.79), with q that this time is equal to k.

We now want to show the fifth and last point of the lemma. Using (2.66) we have

$$(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 = \sum_{j=0}^{3} \binom{3}{j} (\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j}.$$

Therefore

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 \varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i)] =$$

$$= \sum_{j=0}^{3} \binom{3}{j} \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j} \varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i)].$$

We observe that, for $j = 3$, it is

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3 \varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i)] =$$

$$= \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3] + \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3 (\varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i) - 1)] \leq$$

$$\leq c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_{t_i}, \sigma_0) dW_s)^3] + c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} [a(X_s, \sigma_0) - a(X_{t_i}, \sigma_0)] dW_s)^3] +$$

$$+ R(\theta_0, \Delta_{n,i}^{\frac{5}{2} - \epsilon}, X_{t_i}).$$

We remark that the first term here above is centered while on the second we can act like we did on (2.76) (still with an exponent that is 3 instead of 4), obtaining

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} [a(X_s, \sigma_0) - a(X_{t_i}, \sigma_0)] dW_s)^3] \leq R(\theta_0, \Delta_{n,i}^{\frac{5}{2}}, X_{t_i}).$$

For $j = 0$, instead, we get a term on which we act like we did in (2.70) or (2.79), with q that this time is equal to 3 and so we can upper bound it with $R(\theta_0, \Delta_{n,i}^{(1+3\beta) \wedge 3}, X_{t_i})$. For $j = 1$ and $j = 2$ we use Holder inequality, getting

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^j B_{i,n}^{3-j} \varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i)] \leq$$

$$\leq \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^{jp}]^{\frac{1}{p}} \mathbb{E}_i[B_{i,n}^{(3-j)q} \varphi^{k'q}_{\Delta^{\beta}_{n,i}}(\Delta X_i)]^{\frac{1}{q}} \leq$$

$$\leq \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s, \sigma_0) dW_s)^3]^{\frac{j}{3}} \mathbb{E}[B_{i,n}^3 \varphi^{k' \frac{3}{3-j}}_{\Delta^{\beta}_{n,i}}(\Delta X_i)]^{1 - \frac{j}{3}} =$$

$$= R(\theta_0, \Delta_{n,i}^{\frac{3}{2}\frac{j}{3}}, X_{t_i}) R(\theta_0, \Delta_{n,i}^{(1+3\beta)(1-\frac{j}{3})}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{1+3\beta+j(\frac{1}{6}-\beta)}, X_{t_i}).$$

Now, since $\beta > \frac{1}{4} > \frac{1}{6}$, the term we get for $j = 1$ is negligible compared to the one we get for $j = 2$, which is $R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i})$. In conclusion we have

$$\mathbb{E}_i[(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))^3 \varphi^{k'}_{\Delta^{\beta}_{n,i}}(\Delta X_i)] = R(\theta_0, \Delta_{n,i}^{\frac{5}{2}-\epsilon}, X_{t_i}) + R(\theta_0, \Delta_{n,i}^{(1+3\beta) \wedge 3}, X_{t_i}) +$$

$$+ R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i}) = R(\theta_0, \Delta_{n,i}^{\frac{4}{3}+\beta}, X_{t_i}),$$

since we can always find an $\epsilon > 0$ for which $\frac{5}{2} - \epsilon > 1 + 3\beta > \frac{4}{3} + \beta$. The result follows. □

2.8.1.4 Proof of Lemma 15

Proof. We first of all observe that, for all $k \geq 1$, $|\varphi'_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|^k$ is different from 0 only if $|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]$. Recalling that from its definition (2.66) $B_{i,n} = \int_{t_i}^{t_{i+1}} b(X_s, \mu)ds + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} z\gamma(X_{s^-})\tilde{\mu}(ds, dz) + R(\theta, \Delta_{n,i}, X_{t_i})$, it follows

$$\mathbb{E}[|X_{t_{i+1}}^\theta - m(\mu, \sigma, X_{t_i})|^p |\varphi'_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta)|^k] \leq \tag{2.84}$$

$$c\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s^\theta, \sigma)dW_s|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}] + c\mathbb{E}[|B_{i,n}|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}].$$

On the first term here above we use Holder inequality, (2.50) and (2.65), remarking that $\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\} \subset \left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \geq \Delta_{n,i}^\beta\right\}$. We get it is upper bounded by

$$\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s^\theta, \sigma)dW_s|^{pp_1}]^{\frac{1}{p_1}} \mathbb{E}[1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}]^{\frac{1}{p_2}} \leq$$

$$\leq R(\theta, \Delta_{n,i}^{\frac{p}{2}}, X_{t_i}) R(\theta, \Delta_{n,i}, X_{t_i})^{\frac{1}{p_2}} = R(\theta, \Delta_{n,i}^{\frac{p}{2}+1-\epsilon}, X_{t_i}),$$

for all $\epsilon > 0$. In the last inequality we have taken p_1 big and p_2 next to 1.
We now study the second term of (2.84). From the definition of $B_{i,n}$ given here above, Holder inequality, the polynomial growth of b and still (2.65) we get that the second term of (2.84) is upper bounded by

$$R(\theta, \Delta_{n,i}^{p+1-\epsilon}, X_{t_i}) + \mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}].$$

We now recall that $\Delta X_i^c = (X_{t_{i+1}}^\theta - X_{t_i}^\theta) - \Delta X_i^J$ and so when $|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \leq 2\Delta_{n,i}^\beta$ and $|\Delta X_i^J| \geq 4\Delta_{n,i}^\beta$, then $|\Delta X_i^c|$ must be more than $2\Delta_{n,i}^\beta$. Hence

$$\mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta]\right\}}] \leq \mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta], |\Delta X_i^J| \geq 4\Delta_{n,i}^\beta\right\}}] +$$

$$+ \mathbb{E}[|\Delta X_i^J|^p 1_{\left\{|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta], |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}] \leq$$

$$\leq c\mathbb{E}[|\Delta X_i^J|^{pp_1}]^{\frac{1}{p_1}} \mathbb{P}(|\Delta X_i^c| \geq 2\Delta_{n,i}^\beta)^{\frac{1}{p_2}} +$$

$$+ c\Delta_{n,i}^{\beta p} \mathbb{P}(|X_{t_{i+1}}^\theta - X_{t_i}^\theta| \in [\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta], |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta) \leq \tag{2.85}$$

$$\leq R(\theta, \Delta_{n,i}^{\frac{1}{p_1}}, X_{t_i}) R(\theta, \Delta_{n,i}^{\frac{(\frac{1}{2}-\beta)r}{p_2}}, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}) =$$

$$= R(\theta, \Delta_{n,i}^{(\frac{1}{2}-\beta)r-\epsilon}, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}) = R(\theta, \Delta_{n,i}^{1+\beta p}, X_{t_i}),$$

where we have used Kunita inequality (for $pp_1 \geq 2$, that holds since we take p_1 big and p_2 next to 1), Tchebyschev inequality as we did in (2.65) on the first term and still (2.65) on the second. Moreover we have used that, by the arbitrariness of $r > 0$, we can always find r and ϵ such that $(\frac{1}{2} - \beta)r - \epsilon > 1 + \beta p$. The result follows. □

2.8.1.5 Proof of Lemma 16

Proof. The case $q \geq 2$ has already been proved in Lemma 10 of Chapter 1 so, we are going to focus on the case $q \in [1, 2)$.
For all $n \in \mathbb{N}$ and $i \in \mathbb{N}$ we define the set on which all the jumps of L on the interval $(t_i, t_{i+1}]$ are small:

$$N_n^i := \left\{ |\Delta L_s| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}; \quad \forall s \in (t_i, t_{i+1}] \right\}.$$

We split the jumps on $N_{i,n}$ and its complementary, getting

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{N_n^i}] + \mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|^q 1_{(N_n^i)^c}]. \tag{2.86}$$

We now observe that, by the definition of N_n^i, the first term here above is upper bounded by

$$\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} z\, \gamma(X_{s^-})\tilde{\mu}(ds, dz)|^q + |\int_{t_i}^{t_{i+1}} \int_{|z| \geq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|\, |\gamma(X_{s^-})|\bar{\mu}(ds, dz)|^q].$$

From our assumptions on the jump density the second term here above is upper bounded by a $R(\theta, \Delta_{n,i}^q, X_{t_i})$ function while on the first one we use Lemma 2.1.5 of [46]. We can therefore upper bound it with

$$c\mathbb{E}_i[\int_{t_i}^{t_{i+1}} \int_{|z| \leq \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}} |z|^q |\gamma(X_{s^-})|^q \bar{\mu}(ds, dz)] \leq R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}),$$

having used again that $\bar{\mu}(ds, dz) = F(dz)ds$ and Assumption 4 on F. It follows

$$\mathbb{E}_i[|\Delta X_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^q 1_{N_n^i}] \leq R(\theta, \Delta_{n,i}^q, X_{t_i}) + R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}) = R(\theta, \Delta_{n,i}^q, X_{t_i}).$$

Regarding the second term of (2.86), we have that $|\Delta X_i^J| \leq |\Delta X_i| + |\Delta X_i^c|$ and, as we have already remarked several times, by the definition of φ it is different from zero only if $|\Delta X_i|^q \leq c\Delta_{n,i}^{\beta q}$. It follows

$$\mathbb{E}_i[|\Delta X_i|^q |\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^q 1_{(N_n^i)^c}] \leq c\Delta_{n,i}^{\beta q} \mathbb{P}_i((N_n^i)^c) \leq R(\theta, \Delta_{n,i}^{1+\beta q}, X_{t_i}),$$

where the last inequality is a consequence of the following:

$$\mathbb{P}_i((N_n^i)^c) = \mathbb{P}_i(\exists s \in (t_i, t_{i+1}] \, : \, |\Delta L_s| > \frac{4\Delta_{n,i}^\beta}{\gamma_{min}}) \leq c \int_{t_i}^{t_{i+1}} \int_{\frac{4\Delta_{n,i}^\beta}{\gamma_{min}}}^\infty F(z) dz ds \leq c\Delta_{n,i}.$$

In the same way

$$\mathbb{E}_i[|\Delta X_i^c|^q |\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^q 1_{(N_n^i)^c}] \leq c\Delta_{n,i}^{\frac{1}{2}q - \epsilon} \Delta_{n,i}^{1-\epsilon}(1 + |X_{t_i}|^c) = R(\theta, \Delta_{n,i}^{1+\frac{1}{2}q}, X_{t_i}).$$

Putting all pieces together we have

$$\mathbb{E}_i[|\Delta X_i^J|^q |\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)|^q] \leq R(\theta, \Delta_{n,i}^q, X_{t_i}),$$

that is the result we wanted remarking that, for $q \in [1, 2)$, $1 + \beta q > q$ and so $\Delta_{n,i}^q = \Delta_{n,i}^{q \wedge (1+\beta q)}$. □

2.8.1.6 Proof of Proposition 15

Proof. We want to prove the tightness of the sequence $S_n(\theta)$. Since the sum of tight sequences is still tight, we show the tightness of the sequence $S_{n1}(\theta)$ and $S_{n2}(\theta)$ which are such that $S_n(\theta) = S_{n1}(\theta) + S_{n2}(\theta)$:

$$S_{n1}(\theta) := \frac{1}{T_n}\sum_{i=0}^{n-1}(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})g_{i,n}(\theta, X_{t_i}) \quad \text{and}$$

$$S_{n2}(\theta) := \frac{1}{T_n}\sum_{i=0}^{n-1}(m(\mu_0, \sigma_0, X_{t_i}) - m(\mu, \sigma, X_{t_i}))\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})g_{i,n}(\theta, X_{t_i}).$$

We prove that $S_{n1}(\theta)$ is tight using Kolmogorov criterion, we therefore want to show that inequalities analogous to (2.46) and (2.47) hold. Starting with the proof of (2.47) we have that, using Burkholder and Jensen inequality,

$$\mathbb{E}[|S_{n1}(\theta_1) - S_{n1}(\theta_2)|^m] \leq \tag{2.87}$$

$$\leq \frac{cn^{\frac{m}{2}-1}}{(n\Delta_n)^m}\sum_{i=0}^{n-1}\mathbb{E}[|(X_{t_{i+1}} - m(\mu_0, \sigma_0, X_{t_i}))|^m|\varphi^m_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i})|g_{i,n}(\theta_1, X_{t_i}) - g_{i,n}(\theta_2, X_{t_i})|^m].$$

We observe that, using finite-increments theorem and the assumption on the derivatives of $g_{i,n}$ with respect to the parameters, it is

$$|g_{i,n}(\theta_1, X_{t_i}) - g_{i,n}(\theta_2, X_{t_i})|^m \leq |R(\theta, 1, X_{t_i})|^m|\mu_1 - \mu_2|^m + |R(\theta, 1, X_{t_i})|^m|\sigma_1 - \sigma_2|^m \tag{2.88}$$

where actually the function R is computed in a point $\tilde{\theta} := (\tilde{\mu}, \tilde{\sigma})$, with $\tilde{\mu} \in (\mu_1, \mu_2)$ and $\tilde{\sigma} \in (\sigma_1, \sigma_2)$ but, since the property (3.23) of R is uniform in θ, we have chosen to write it simply as $R(\theta, 1, X_{t_i})$. Replacing (2.88) in (2.87) and using the fourth point of Lemma 14 it follows

$$\mathbb{E}[|S_{n1}(\theta_1) - S_{n1}(\theta_2)|^m] \leq \frac{cn^{\frac{m}{2}-1}}{(n\Delta_n)^m}n\Delta_n^{\frac{m}{2}\wedge(1+m\beta)}(|\mu_1 - \mu_2|^m + |\sigma_1 - \sigma_2|^m) \leq$$

$$\leq \frac{c}{(n\Delta_n)^{\frac{m}{2}}}\Delta_n^{0\wedge(1+m\beta-\frac{m}{2})}(|\mu_1 - \mu_2|^m + |\sigma_1 - \sigma_2|^m),$$

with $\frac{1}{(n\Delta_n)^{\frac{m}{2}}}\Delta_n^{0\wedge(1+m\beta-\frac{m}{2})} < c$ because $n\Delta_n$ is lower bounded by a constant and we can always find an $m \geq 2$ for which $1 + m\beta - \frac{m}{2} > 0$ since $\beta \in (\frac{1}{4}, \frac{1}{2})$. Hence, (2.47) is proved.

Acting exactly in the same way but using this time the control on $g_{i,n}$ instead of on its derivatives we have also an estimation for S_{n1} analogous to (2.46); the tightness of S_{n1} follows.

Concerning S_{n2} we observe that, for $\vartheta = \mu$ and $\vartheta = \sigma$, it is

$$|\partial_\vartheta S_{n2}(\theta)| \leq$$

$$\leq c|\partial_\vartheta m(\mu, \sigma, X_{t_i})||g_{i,n}(\theta, X_{t_i})| + c|m(\mu_0, \sigma_0, X_{t_i}) - m(\mu, \sigma, X_{t_i})||\partial_\vartheta g_{i,n}(\theta, X_{t_i})|.$$

From the controls we have assumed on $g_{i,n}$ and its derivatives, the finite-increments theorem and the first and the second point of Proposition 16 we have $|\partial_\vartheta S_{n2}(\theta)| \leq R(\theta, \Delta_{n,i}, X_{t_i})$. Therefore for both $\vartheta = \mu$ and $\vartheta = \sigma$, using also that $\frac{1}{T_n} = O(\frac{1}{n\Delta_n})$, we get

$$\sup_n \mathbb{E}[\sup_{\mu,\sigma}|\partial_\vartheta S_{n2}(\theta)|] \leq \sup_n \frac{c}{T_n}\sum_{i=0}^{n-1}\mathbb{E}[\sup_{\mu,\sigma}|R(\theta, \Delta_{n,i}, X_{t_i})] \leq c.$$

The tightness of S_{n2} (and therefore of S_n) follows.

□

2.8.2 Proof of derivatives of m and m_2

In order to prove the developments and the bounds on the derivatives of m and m_2, the following lemmas will be useful. We point out that X_t^θ is $X_t^{\theta,x}$ and so the process starts in 0: $X_0^{\theta,x} = x$.

Lemma 24. *Suppose that Assumptions from 1 to 4 and A7 hold. Then,* $\forall p \geq 2$ $\exists c > 0$*:* $\forall h \leq \Delta_n$ $\forall x$ *we have*

$$\mathbb{E}[|\frac{\partial_\mu X_h^{\theta,x}}{h}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial_\mu^2 X_h^{\theta,x}}{h}|^p] \leq c(1+|x|^c), \tag{2.89}$$

$$\mathbb{E}[|\frac{\partial_\sigma X_h^{\theta,x}}{h^{\frac{1}{2}}}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial_{\sigma\mu}^2 X_h^{\theta,x}}{h^{\frac{3}{2}}}|^p] \leq c(1+|x|^c), \tag{2.90}$$

$$\mathbb{E}[|\frac{\partial_\sigma^2 X_h^{\theta,x}}{h^{\frac{1}{2}}}|^p] \leq c(1+|x|^c), \tag{2.91}$$

$$\mathbb{E}[|\frac{\partial_\sigma^3 X_h^{\theta,x}}{h^{\frac{1}{2}}}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial_\mu^3 X_h^{\theta,x}}{h}|^p] \leq c(1+|x|^c) \tag{2.92}$$

$$\mathbb{E}[|\frac{\partial_{\sigma\mu\sigma}^3 X_h^{\theta,x}}{h^{\frac{3}{2}}}|^p] \leq c(1+|x|^c), \qquad \mathbb{E}[|\frac{\partial_{\mu\mu\sigma}^3 X_h^{\theta,x}}{h^{\frac{3}{2}}}|^p] \leq c(1+|x|^c), \tag{2.93}$$

Proof. Lemma 24.
Inequalities (2.89) have already been proved in Lemma 9 of Chapter 1. To show the first inequality of (2.90), we observe that the dynamic of the process $\partial_\sigma X^{\theta,x}$ is known (cf. [14], section 5):

$$\partial_\sigma X_h^{\theta,x} = \int_0^h \partial_x b(\mu, X_s^{\theta,x})\partial_\sigma X_s^{\theta,x} ds + \int_0^h (\partial_x a(\sigma, X_s^{\theta,x})\partial_\sigma X_s^{\theta,x} + \partial_\sigma a(\sigma, X_s^{\theta,x}))dW_s + \tag{2.94}$$
$$+ \int_0^h \int_{\mathbb{R}} \partial_x \gamma(X_{s^-}^{\theta,x})\partial_\sigma X_s^{\theta,x} z\tilde{\mu}(dz, ds).$$

From now on, we will drop the dependence of the starting point in order to make the notation easier. Taking the L^p norm of (2.94) we get it is upper bounded by the sum of three terms. On the first one we use Jensen inequality and the fact that the derivatives of b with rapport to x are supposed bounded to obtain

$$\mathbb{E}[|\int_0^h (\partial_x b(\mu, X_s^\theta)\partial_\sigma X_s^\theta ds|^p] \leq h^{p-1}\int_0^h \mathbb{E}[|\partial_x b(\mu, X_s^\theta)|^p|\partial_\sigma X_s^\theta|^p]ds \leq \tag{2.95}$$
$$\leq ch^{p-1}\int_0^h \mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds.$$

Let us now consider the stochastic integral. Using Burkholder-Davis-Gundy inequality we have

$$\mathbb{E}[|\int_0^h (\partial_x a(\sigma, X_s^\theta)\partial_\sigma X_s^\theta + \partial_\sigma a(\sigma, X_s^\theta))dW_s|^p] \leq c\mathbb{E}[|\int_0^h (\partial_x a(\sigma, X_s^\theta))^2(\partial_\sigma X_s^\theta)^2 ds|^{\frac{p}{2}}] +$$

$$+c\mathbb{E}[|\int_0^h (\partial_\sigma a(\sigma, X_s^\theta))^2 ds|^{\frac{p}{2}}] \leq ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds + ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[(1+|X_s^\theta|^c)]ds \leq \tag{2.96}$$

$$\leq ch^{\frac{p}{2}-1}\int_0^h \mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds + ch^{\frac{p}{2}}(1+|x|^c),$$

where we have used Jensen inequality, the fact that, by A7, the derivatives of a with rapport to x are supposed bounded and those with rapport to σ have polynomial growth and the second point of Lemma 25.
We now consider the third term on the right hand side of (2.94), it can be estimated using Kunita inequality (cf. the Appendix of [46]):

$$\mathbb{E}[|\int_0^h\int_{\mathbb{R}}\partial_x\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta z\tilde{\mu}(dz,ds)|^p] \leq c\mathbb{E}[\int_0^h\int_{\mathbb{R}}|\partial_x\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta|^p|z|^p\bar{\mu}(dz,ds)]+$$

$$+c\mathbb{E}[|\int_0^h\int_{\mathbb{R}}(\partial_x\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta)^2 z^2\bar{\mu}(dz,ds)|^{\frac{p}{2}}] \leq$$

$$\leq c\int_0^h \mathbb{E}[|\partial_x\gamma(X_{s^-}^\theta)|^p|\partial_\sigma X_s^\theta|^p](\int_{\mathbb{R}}|z|^pF(z)dz)ds+$$

$$+c\mathbb{E}[|\int_0^h(\partial_x\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta)^2(\int_{\mathbb{R}}z^2F(z)dz)ds|^{\frac{p}{2}}] \leq$$

$$\leq c\int_0^h\mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds + c\mathbb{E}[|\int_0^h(\partial_\sigma X_s^\theta)^2ds|^{\frac{p}{2}}],$$

where in the last two inequalities we have just used the definition of the compensated measure $\bar{\mu}$,the third point of Assumption 4 and the fact that the derivatives of γ are supposed bounded. By the Jensen inequality we get

$$\mathbb{E}[|\int_0^h\int_{\mathbb{R}}\partial_x\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta z\tilde{\mu}(dz,ds)|^p] \leq c(1+h^{\frac{p}{2}-1})\int_0^h\mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds. \tag{2.97}$$

From (2.95), (2.96) and (2.97), it follows

$$\mathbb{E}[|\partial_\sigma X_h^\theta|^p] \leq c(h^{p-1}+h^{\frac{p}{2}-1}+1)\int_0^h\mathbb{E}[|\partial_\sigma X_s^\theta|^p]ds + ch^{\frac{p}{2}}(1+|x|^c).$$

Gronwall Lemma gives us

$$\mathbb{E}[|\partial_\sigma X_h^\theta|^p] \leq ch^{\frac{p}{2}}(1+|x|^c)e^{c(h^{p-1}+h^{\frac{p}{2}-1}+1)},$$

we therefore obtain the first inequality in (2.90). Concerning the second, we observe we can deduce the dynamic of the process $\partial_{\mu\sigma}X^\theta$ from (2.94). It is $\partial^2_{\mu\sigma}X_h^\theta =$

$$= \int_0^h(\partial_x^2 b(\mu,X_s^\theta)\partial_\sigma X_s^\theta\partial_\mu X_s^\theta + \partial^2_{\mu x}b(\mu,X_s^\theta)\partial_\sigma X_s^\theta + \partial_x b(\mu,X_s^\theta)\partial^2_{\sigma\mu}X_s^\theta)ds+$$

$$+\int_0^h(\partial_x^2 a(\sigma,X_s^\theta)\partial_\sigma X_s^\theta\partial_\mu X_s^\theta + \partial^2_{\sigma x}a(\sigma,X_s^\theta)\partial_\mu X_s^\theta + \partial_x a(\sigma,X_s^\theta)\partial^2_{\sigma\mu}X_s^\theta)dW_s+$$

$$+\int_0^h\int_{\mathbb{R}}(\partial_x^2\gamma(X_{s^-}^\theta)\partial_\sigma X_s^\theta\partial_\mu X_s^\theta + \partial_x\gamma(X_{s^-}^\theta)\partial^2_{\sigma\mu}X_s^\theta)z\tilde{\mu}(ds,dz)$$

On the p-norm of the first integral we use Jensen inequality, the fact that the derivatives with respect to x are bounded and the estimation we have already proved on the L^p norm of the derivatives of our process with respect to μ and σ. We get it is upper bounded by

$$ch^{p-1}\int_0^h(\mathbb{E}[|\partial_\sigma X_s^\theta\partial_\mu X_s^\theta|^p] + \mathbb{E}[|\partial_\sigma X_s^\theta|^p] + \mathbb{E}[|\partial^2_{\sigma\mu}X_s^\theta|^p])ds \leq$$

$$\leq c(h^{\frac{5}{2}p} + h^{\frac{3}{2}p})(1 + |x|^c) + ch^{p-1} \int_0^h \mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_s|^p] ds,$$

having also used Holder inequality. Acting in the same way on the stochastic integral we get it is upper bounded by

$$ch^{\frac{p}{2}}(h^{\frac{3}{2}p} + h^p)(1 + |x|^c) + ch^{\frac{p}{2}-1} \int_0^h \mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_s|^p] ds,$$

while we upper bound the third term in the dynamic of $\partial^2_{\sigma\mu} X^\theta$, acting as we did in order to show (2.97), with

$$ch^{\frac{3}{2}p+1}(1 + |x|^c) + c \int_0^h \mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_s|^p] ds + ch^{2p}(1 + |x|^c) + ch^{\frac{p}{2}-1} \int_0^h \mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_s|^p] ds.$$

In total we have, not considering the negligible terms,

$$\mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_h|^p] \leq ch^{\frac{3}{2}p}(1 + |x|^c) + c(h^{p-1} + h^{\frac{p}{2}-1} + 1) \int_0^h \mathbb{E}[|\partial^2_{\sigma\mu} X^\theta_s|^p] ds.$$

From Gronwall Lemma it follows the second inequality of (2.90), as we wanted. We are left to show (2.91). Again, the dynamic of $\partial^2_\sigma X^\theta$ is known:

$$\partial^2_\sigma X^\theta_h = \int_0^h (\partial^2_x b(\mu, X^\theta_s)(\partial_\sigma X^\theta_s)^2 + \partial_x b(\mu, X^\theta_s)\partial^2_\sigma X^\theta_s) ds +$$

$$+ \int_0^h (\partial^2_x a(\sigma, X^\theta_s)(\partial_\sigma X^\theta_s)^2 + 2\partial^2_{\sigma x} a(\sigma, X^\theta_s)\partial_\sigma X^\theta_s + \partial_x a(\sigma, X^\theta_s)\partial^2_\sigma X^\theta_s + \partial^2_\sigma a(\sigma, X^\theta_s)) dW_s +$$

$$+ \int_0^h \int_{\mathbb{R}} (\partial^2_x \gamma(X^\theta_{s^-})(\partial_\sigma X^\theta_s)^2 + \partial_x \gamma(X^\theta_{s^-})\partial^2_\sigma X^\theta_s) z \tilde{\mu}(ds, dz). \quad (2.98)$$

Acting exactly like we did for the estimation of the p-moments of the processes $\partial_\sigma X^\theta$ and $\partial^2_{\sigma\mu} X^\theta$ we get

$$\mathbb{E}[|\partial^2_\sigma X^\theta_h|^p] \leq c(1+|x|^c)(h^{2p}+h^{\frac{3}{2}p}+h^p+h^{\frac{p}{2}}+h^{p+1})+c(h^{p-1}+h^{\frac{p}{2}-1}+1) \int_0^h \mathbb{E}[|\partial^2_\sigma X^\theta_s|^p] ds. \quad (2.99)$$

Using Gronwall Lemma and remarking that the other terms are negligible compared to $ch^{\frac{p}{2}}(1 + |x|^c)$, we obtain the result wanted.
Concerning the third derivatives, it is easy to see that, writing the dynamics of $\partial^3_\sigma X^\theta_h$, $\partial^3_\mu X^\theta_h$, $\partial^3_{\sigma\mu\sigma} X^\theta_h$ and $\partial^3_{\mu\mu\sigma} X^\theta_h$ the principal terms are such that their order are, respectively, $h^{\frac{1}{2}}$, h and twice $h^{\frac{3}{2}}$. Acting exactly like before, (2.92) and (2.93) follow. □

We are left to show one last proposition, before showing Propositions 16, 17 and 18:

Proposition 19. *Suppose that Assumptions 1 to 4 hold. Moreover suppose that $(Z_h)_h$ is a family of random variables such that $\mathbb{E}[|Z_h|^p | X^\theta_0 = x] \leq c(1 + |x|^c)$. Then $\forall k \geq 1$ $\forall \epsilon > 0$, we have*

$$\sup_{h \in [0, \Delta_n]} \mathbb{E}[|Z_h||\varphi^{(k)}_{h^\beta}(X^\theta_h - x)||X^\theta_0 = x] = R(\theta, h^{1-\epsilon}, x).$$

We have used $\varphi^{(k)}_{h^\beta}(y)$ in order to denote the k-th derivative $\varphi^{(k)}(\frac{y}{h^\beta})$.

Proof. Proposition 19.
Once again, $|\varphi^{(k)}_{h^\beta}(X^\theta_h - x)|$ is different from 0 only if $|X^\theta_h - x| \in [h^\beta, 2h^\beta]$. We can therefore use Holder inequality (with p big and q next to 1) and (2.65) to get, $\forall h \in [0, \Delta_n]$,

$$\mathbb{E}[|Z_h||\varphi^{(k)}_{h^\beta}(X^\theta_h - x)||X^\theta_0 = x] \leq$$

$$\leq \mathbb{E}[|Z_h|^p|X^\theta_0 = x]^{\frac{1}{p}}\mathbb{E}[1_{\left\{|X^\theta_h - x| \in [h^\beta, 2h^\beta]\right\}}|X^\theta_0 = x]^{\frac{1}{q}} \leq R(\theta, h, x)^{\frac{1}{q}} = R(\theta, h^{1-\epsilon}, x).$$

□

2.8.2.1 Proof of Proposition 16

Proof. The first point has already been showed in Proposition 8 of Chapter 1, we start proving the second. Since by the homogeneity of the equation m and m_2 depend only on the difference $t_{i+1} - t_i$ we can consider WLOG $\forall h \leq \Delta_n$

$$m(\mu, \sigma, x) := \frac{\mathbb{E}[X^\theta_h \varphi_{h^\beta}(X^\theta_h - X^\theta_0)|X^\theta_0 = x]}{\mathbb{E}[\varphi_{h^\beta}(X^\theta_h - X^\theta_0)|X^\theta_0 = x]} = \frac{\mathbb{E}[X^\theta_h \varphi_{h^\beta}(X^\theta_h - x)]}{\mathbb{E}[\varphi_{h^\beta}(X^\theta_h - x)]}. \tag{2.100}$$

Hence,

$$\partial_\sigma m(\mu, \sigma, x) = \frac{\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi_{h^\beta}(X^\theta_h - x)] + \mathbb{E}[X^\theta_h h^{-\beta}(\partial_\sigma X^\theta_h)\varphi'_{h^\beta}(X^\theta_h - x)]}{\mathbb{E}[\varphi_{h^\beta}(X^\theta_h - x)]} +$$

$$-m(\mu, \sigma, x)\frac{\mathbb{E}[h^{-\beta}(\partial_\sigma X^\theta_h)\varphi'_{h^\beta}(X^\theta_h - x)]}{\mathbb{E}[\varphi_{h^\beta}(X^\theta_h - x)]}.$$

On the numerator of the second and third term we use Proposition 19 taking as Z_h, respectively, $X^\theta_h \frac{\partial_\sigma X^\theta_h}{h^{\frac{1}{2}}}$ and $\frac{\partial_\sigma X^\theta_h}{h^{\frac{1}{2}}}$. We get, remarking moreover that from (2.31) $m(\mu, \sigma, x)$ is $R(\theta, 1, x)$ and from Theorem 1 in Chapter 1 also the denominator is lower bounded for $|x| < |h|^{-k}$,

$$|\partial_\sigma m(\mu, \sigma, x)| \leq R(\theta, 1, x)|\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi_{h^\beta}(X^\theta_h - x)]| + R(\theta, h^{\frac{3}{2} - \beta - \epsilon}, x). \tag{2.101}$$

To estimate $|\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi_{h^\beta}(X^\theta_h - x)]|$ we replace the dynamic (2.94) of $\partial_\sigma X^\theta_h$. On the first integral we use Holder inequality and (2.95) to get

$$|\mathbb{E}[\int_0^h (\partial_x b(X^{\theta,x}_s, \mu)\partial_\sigma X^{\theta,x}_s ds \varphi_{h^\beta}(X^\theta_h - x)]| \leq$$

$$\leq (ch^{p-1}\int_0^h \mathbb{E}[|\partial_\sigma X^{\theta,x}_s|^p])^{\frac{1}{p}} \leq ch^{\frac{3}{2}}(1 + |x|^c) = R(\theta, h^{\frac{3}{2}}, x),$$

where in the last inequality we have also used the first inequality of (2.90). On $\int_0^h \partial_x a(X^{\theta,x}_s, \sigma)\partial_\sigma X^{\theta,x}_s dW_s$ we use again Holder inequality, (2.96) (considering only the estimation on its first term) and the first inequality of (2.90) to obtain

$$|\mathbb{E}[\int_0^h \partial_x a(X^{\theta,x}_s, \sigma)\partial_\sigma X^{\theta,x}_s dW_s \varphi_{h^\beta}(X^\theta_h - x)]| \leq R(\theta, h, x).$$

Concerning $|\mathbb{E}[\int_0^h \partial_\sigma a(X^{\theta,x}_s, \sigma)dW_s \varphi_{h^\beta}(X^\theta_h - x)]|$, we act on it like we did in (2.83), with $\partial_\sigma a$ instead of a. We therefore get $|\mathbb{E}[\int_0^h \partial_\sigma a(X^{\theta,x}_s, \sigma)dW_s \varphi_{h^\beta}(X^\theta_h - x)]| \leq$

$R(\theta, h, x)$. To conclude the proof of this point we use on the jump part Holder inequality, (2.97) and the first inequality of (2.90). We get

$$|\mathbb{E}[\int_0^h \int_{\mathbb{R}} \partial_x \gamma(X_{s^-}^\theta) \partial_\sigma X_s^\theta z \tilde{\mu}(dz, ds) \varphi_{h^\beta}(X_h^\theta - x)]| \le$$

$$\le (c(1 + h^{\frac{p}{2}-1}) \int_0^h \mathbb{E}[|\partial_\sigma X_s^\theta|^p] ds)^{\frac{1}{p}} \le R(\theta, h^{(\frac{1}{p}+\frac{1}{2})\wedge 1}, x).$$

We can take $p = 2$, finding $|\mathbb{E}[(\partial_\sigma X_h^\theta) \varphi_{h^\beta}(X_h^\theta - x)]| = R(\theta, h, x)$. Replacing it in (2.101) and observing that $\frac{3}{2} - \beta$ is always more than 1, it follows the second point of Proposition 16.

In order to prove the third and the fourth point we first of all need to compute the derivative of m_2 with respect to both the parameters. We can just write

$$\partial_\vartheta m_2(\mu, \sigma, x) = 2\frac{\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))(\partial_\vartheta X_h^\theta - \partial_\vartheta m(\mu, \sigma, x))\varphi_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} +$$

$$+ \frac{\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))^2 h^{-\beta}(\partial_\vartheta X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} +$$

$$- m_2(\mu, \sigma, x) h^{-\beta} \frac{\mathbb{E}[(\partial_\vartheta X_h^\theta)\varphi'_{h^\beta}(X_h^\theta - x)]}{\mathbb{E}[\varphi_{h^\beta}(X_h^\theta - x)]} =: I_{1,\theta} + I_{2,\theta} + I_{3,\theta}.$$

We are going to show that, considering the derivatives with respect to both μ and σ, $I_{2,\theta}$ and $I_{3,\theta}$ are negligible compared to $I_{1,\theta}$. In order to prove it we use Theorem 1 of Chapter 1 on the denominator of $I_{2,\theta}$ and $I_{3,\theta}$, while on the numerator of $I_{2,\theta}$ we use Holder inequality, (2.89) if we consider the derivative with respect to μ or the first equation of (2.90) if we consider the derivative with respect to σ and Lemma 15. On the numerator of $I_{3,\theta}$ we use Proposition 19 and we remind that, as a consequence of Ad, m_2 is a $R(\theta, h, x)$ function. It follows

$$|I_{2,\mu} + I_{3,\mu}| \le R(\theta, h^{2+\beta-\epsilon}, x) + R(\theta, h^{3-\beta-\epsilon}, x) = R(\theta, h^{2+\beta-\epsilon}, x); \quad (2.102)$$

$$|I_{2,\sigma} + I_{3,\sigma}| \le R(\theta, h^{\frac{3}{2}+\beta-\epsilon}, x) + R(\theta, h^{\frac{5}{2}-\beta-\epsilon}, x) = R(\theta, h^{\frac{3}{2}+\beta-\epsilon}, x). \quad (2.103)$$

Concerning $I_{1,\theta}$, its numerator is

$$2\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))\partial_\theta X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] +$$

$$-2\partial_\theta m(\mu, \sigma, x))\mathbb{E}[(X_h^\theta - m(\mu, \sigma, x))\varphi_{h^\beta}(X_h^\theta - x)] =: I_{1,1}^\theta + I_{1,2}^\theta.$$

From the first two points we have already proved of Proposition 16 and the third point of Lemma 14 we get

$$|I_{1,2}^\mu| \le R(\theta, h^2, x), \qquad |I_{1,2}^\sigma| \le R(\theta, h^2, x). \quad (2.104)$$

Now we consider $I_{1,1}^\theta$ and we act differently depending on if we are dealing with the derivative with rapport to μ or those with rapport to σ. We start studying $I_{1,1}^\sigma$. Using a notation analogous to the one used in the proof of Lemma 14, we set

$$X_h^\theta - m(\mu, \sigma, x) =: \int_0^h a(X_s^\theta, \sigma) dW_s + B_h$$

and so it turns we have

$$I_{1,1}^\sigma = 2\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)\partial_\sigma X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] + 2\mathbb{E}[(B_h)\partial_\sigma X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)]. \quad (2.105)$$

On the second term here above we use Cauchy-Schwartz inequality, control analogous to (2.70) and the first estimation in (2.90), getting

$$\mathbb{E}[(B_h)\partial_\sigma X_h^\theta \varphi_{h^\beta}(X_h^\theta - x)] \le c\mathbb{E}[(B_h)^2\varphi_{h^\beta}^2(X_h^\theta - x)]^{\frac{1}{2}}\mathbb{E}[(\partial_\sigma X_h^\theta)^2]^{\frac{1}{2}} \le \tag{2.106}$$

$$\le R(\theta, h^{1+2\beta}, x)^{\frac{1}{2}} R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{1+\beta}, x).$$

To evaluate the first term on the right hand side of (2.105) we replace the dynamic (2.94) of $\partial_\sigma X_h^\theta$, isolating the principal term: $\partial_\sigma X_h^\theta := \int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s + G_\sigma$. We get

$$\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)] + \tag{2.107}$$

$$+\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)(\varphi_{h^\beta}(X_h^\theta - x) - 1)] +$$

$$+\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma)dW_s)G_\sigma \varphi_{h^\beta}(X_h^\theta - x)].$$

On the first term here above we add and substract both $a(x, \sigma)$ and $\partial_\sigma a(x, \sigma)$ getting a main term and three terms of increments. We observe it holds the following estimation, using Cauchy-Schwartz inequality, (2.50) and the first point of Lemma 25

$$\mathbb{E}[(\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]dW_s)(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)] \le \tag{2.108}$$

$$\le \mathbb{E}[(\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]dW_s)^2]^{\frac{1}{2}}\mathbb{E}[(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)^2]^{\frac{1}{2}} \le$$

$$\le c\mathbb{E}[\int_0^h (X_s - x)^2 ds]^{\frac{1}{2}} R(\theta, h^{\frac{1}{2}}, x) \le R(\theta, h, x)R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{\frac{3}{2}}, x).$$

We can act in the same way considering the increments of $\partial_\sigma a$ or the term on which we have the increments of both a and $\partial_\sigma a$. It follows that the first term of (2.107) is

$$\mathbb{E}[(\int_0^h a(x, \sigma)dW_s)(\int_0^h \partial_\sigma a(x, \sigma)dW_s)] + R(\theta, h^{\frac{3}{2}}, x) = h\, a(x, \sigma)\partial_\sigma a(x, \sigma) + R(\theta, h^{\frac{3}{2}}, x). \tag{2.109}$$

On the second term of (2.107) we use Holder inequality twice (with p big and q next to 1), (2.50) twice and (2.65). We get it is upper bounded by

$$\mathbb{E}[|(\int_0^h a(X_s^\theta, \sigma)dW_s)(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dWs)|^p]^{\frac{1}{p}}\mathbb{E}[(\varphi_{h^\beta}(X_h^\theta - x) - 1)^q]^{\frac{1}{q}} \le$$

$$\le R(\theta, h, x)R(\theta, h, x)^{\frac{1}{q}} = R(\theta, h^{2-\epsilon}, x).$$

Concerning the third term of (2.107), we first of all use Cauchy-Schwartz inequality, the fact that φ is bounded in absolute value and (2.50) in order to estimate the stochastic integral while, to estimate the 2-norm of G_σ, we use (2.95), the estimation (2.96) about the negligible part of the stochastic integral and (2.97). It follows it is upper bounded by

$$R(\theta, h^{\frac{1}{2}}, x)c(h^{\frac{3}{2}} + h) = R(\theta, h^{\frac{3}{2}}, x), \tag{2.110}$$

where we have also used the first estimation (2.90) to estimate the expected value. From (2.106), (2.109) - (2.110) it follows

$$I_{1,1}^\sigma = h\, a(x, \sigma)\partial_\sigma a(x, \sigma) + R(\theta, h^{1+\beta}, x).$$

Using also (2.103) and (2.104) we get the development of $\partial_\sigma m_2$ we wanted.
We are left to prove the third point of Proposition 16. It means, comparing it with (2.102) and (2.104), to prove that $|I^\mu_{1,1}| \le R(\theta, h^2, x)$. We observe that (2.105) still holds with the derivative with respect to μ instead of those with respect to σ. We now recall that $B_h = \int_0^h b(X_s, \mu)ds + R(\theta, h, x) + \Delta X^J_h$ and we replace it in $\mathbb{E}[B_h(\partial_\mu X^\theta_s)\varphi_{h^\beta}(X^\theta_h - x)]$, getting

$$|\mathbb{E}[(\int_0^h b(X_s,\mu)ds)(\partial_\mu X^\theta_s)\varphi_{h^\beta}(X^\theta_h - x)] + R(\theta,h,x)\mathbb{E}[(\partial_\mu X^\theta_s)\varphi_{h^\beta}(X^\theta_h - x)] + \quad (2.111)$$

$$+\mathbb{E}[(\Delta X^J_h)(\partial_\mu X^\theta_s)\varphi_{h^\beta}(X^\theta_h - x)]| \le R(\theta, h^2, x) + R(\theta, h^{1+\beta q}, x)^{\frac{1}{q}} R(\theta, h, x) =$$
$$= R(\theta, h^2, x) + R(\theta, h^{2+\beta-\epsilon}, x) = R(\theta, h^2, x),$$

where we have used Holder inequality (having taken p big and q next to 1), the fact that φ is bounded, the polynomial growth of b, the first estimation in (2.89) and Lemma 16 (in a non-conditional form). Concerning the first term of (2.105), we still replace the dynamic of $\partial_\mu X^\theta_s := \int_0^h \partial_\mu b(X^\theta_s, \mu)ds + G_\mu$, where G_μ is the negligible part in the dynamic of $\partial_\mu X^\theta_s$ and it is such that

$$\mathbb{E}[|G_\mu|^p] \le c(1 + h^{\frac{p}{2}-1} + h^{p-1}) \int_0^h \mathbb{E}[|\partial_\mu X^\theta_s|^p]ds \le c(h^{p+1} + h^{\frac{3}{2}p} + h^{2p})(1 + |x|^c)$$

(see Lemma 9 in Chapter 1). We have also used the first inequality of (2.89). It follows $\mathbb{E}[|G_\mu|^p]^{\frac{1}{p}} \le R(\theta, h^{1+\frac{1}{p}}, x)$, for $p \ge 2$. The first term of (2.105) is therefore

$$\mathbb{E}[(\int_0^h a(X^\theta_s,\sigma)dW_s)(\int_0^h \partial_\mu b(X^\theta_s,\mu)ds)] + \quad (2.112)$$

$$+\mathbb{E}[(\int_0^h a(X^\theta_s,\sigma)dW_s)(\int_0^h \partial_\mu b(X^\theta_s,\mu)ds)(1 - \varphi_{h^\beta}(X^\theta_h - x))] +$$

$$+\mathbb{E}[(\int_0^h a(X^\theta_s,\sigma)dW_s)G_\mu \varphi_{h^\beta}(X^\theta_h - x)].$$

Now on the first term here above we act like we did for the estimation of the derivative with respect to σ: we get

$$h\,\partial_\mu b(x,\mu)\mathbb{E}[\int_0^h a(X^\theta_s,\sigma)dW_s] + \mathbb{E}[(\int_0^h a(X^\theta_s,\sigma)dW_s)(\int_0^h [b(X^\theta_s,\mu) - b(x,\mu)]ds)]$$

Now we observe that the first expected value is 0, while on the others we can use Cauchy-Schwartz inequality, (2.50) and we estimate the increments like in (2.108). Hence, the first term of (2.112) is in module upper bounded by $R(\theta, h^2, x)$. On the second term of (2.112) we act like we did on the second term of (2.107), using Holder inequality twice (with p big and q next to 1), (2.50), the polynomial growth of both a and b and (2.65). We get

$$\mathbb{E}[(\int_0^h a(X^\theta_s,\sigma)dW_s)(\int_0^h \partial_\mu b(X^\theta_s,\mu)ds)(1 - \varphi_{h^\beta}(X^\theta_h - x))] \le$$
$$\le R(\theta, h^{\frac{1}{2}}, x)R(\theta, h, x)R(\theta, h^{1-\epsilon}, x) = R(\theta, h^{\frac{5}{2}-\epsilon}, x).$$

Using on the third term of (2.112) Cauchy-Schwartz inequality, (2.50), the fact that $|\varphi|$ is bounded and the estimation of the L^p norm of G_μ given above (2.112) for $p = 2$, we get it is upper bounded in absolute value by $R(\theta, h^2, x)$.
It follows $|I^\mu_{1,1}| \le R(\theta, h^2, x)$, as we wanted. □

2.8.2.2 Proof of Proposition 17

Proof. We first of all write $\partial^2_{\mu\sigma}m$. Again, since by the homogeneity of the equation m and m_2 depend only on the difference $t_{i+1}-t_i$ we can consider, for all $h\le\Delta_n$, $m(\mu,\sigma,x)$ as in (2.100). Hence, writing φ for $\varphi_{h^\beta}(X^\theta_h-x)$ (and $\varphi^{(k)}$ for $\varphi^{(k)}_{h^\beta}(X^\theta_h-x)$), we have $\partial^2_{\mu\sigma}m(\mu,\sigma,x)=$

$$\frac{\mathbb{E}[(\partial^2_{\mu\sigma}X^\theta_h)\varphi]}{\mathbb{E}[\varphi]}+\frac{2h^{-\beta}\mathbb{E}[(\partial_\mu X^\theta_h)(\partial_\sigma X^\theta_h)\varphi']}{\mathbb{E}[\varphi]}-\frac{h^{-\beta}\mathbb{E}[(\partial_\mu X^\theta_h)\varphi']\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi]}{(\mathbb{E}[\varphi])^2}+$$
$$+\frac{h^{-\beta}\mathbb{E}[X^\theta_h(\partial^2_{\mu\sigma}X^\theta_h)\varphi']}{\mathbb{E}[\varphi]}+\frac{h^{-2\beta}\mathbb{E}[X^\theta_h(\partial_\sigma X^\theta_h)(\partial_\mu X^\theta_h)\varphi'']}{\mathbb{E}[\varphi]}-\partial_\mu m\frac{h^{-\beta}\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi']}{\mathbb{E}[\varphi]}+$$
$$-\frac{h^{-2\beta}\mathbb{E}[X^\theta_h(\partial_\sigma X^\theta_h)\varphi']\mathbb{E}[(\partial_\mu X^\theta_h)\varphi']}{(\mathbb{E}[\varphi])^2}-\frac{h^{-\beta}m\mathbb{E}[(\partial^2_{\mu\sigma}X^\theta_h)\varphi']}{\mathbb{E}[\varphi]}+ \tag{2.113}$$
$$-\frac{h^{-2\beta}m\mathbb{E}[(\partial_\mu X^\theta_h)(\partial_\sigma X^\theta_h)\varphi'']}{\mathbb{E}[\varphi]}+\frac{h^{-2\beta}m\mathbb{E}[(\partial_\sigma X^\theta_h)\varphi']\mathbb{E}[(\partial_\mu X^\theta_h)\varphi']}{(\mathbb{E}[\varphi])^2}=:\sum_{j=1}^{10}I^n_j.$$

Now on I^n_2 and $\sum_{j=4}^{10}I^n_j$ we use Proposition 19, where Z_h are respectively $\frac{\partial_\mu X^\theta_h}{h}\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$, $X^\theta_h\frac{\partial^2_{\mu\sigma}X^\theta_h}{h^{\frac32}}$, $X^\theta_h\frac{\partial_\mu X^\theta_h}{h}\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$, $\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$, $X^\theta_h\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$ in the first and $\frac{\partial_\mu X^\theta_h}{h}$ in the second expected value of I^n_7, $\frac{\partial^2_{\mu\sigma}X^\theta_h}{h^{\frac32}}$, $\frac{\partial_\mu X^\theta_h}{h}\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$, $\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$ in the first expected value of I^n_{10} and $\frac{\partial_\mu X^\theta_h}{h^{\frac12}}$ in the second one. Recalling moreover that $|m(\mu,\sigma,x)|\le R(\theta,1,x)$, $|\partial_\mu m(\mu,\sigma,x)|\le R(\theta,h,x)$ and the denominator gives always $R(\theta,1,x)$ it follows

$$|I^n_2+\sum_{j=4}^{10}I^n_j|\le R(\theta,h^{\frac52-\beta-\epsilon},x)+R(\theta,h^{\frac52-2\beta-\epsilon},x)+R(\theta,h^{3-\beta-\epsilon},x)+R(\theta,h^{\frac72-2\beta-\epsilon},x)= \tag{2.114}$$
$$=R(\theta,h^{\frac52-2\beta-\epsilon},x).$$

On I^n_1 we use Holder inequality, the fact that $|\varphi|$ is bounded, second inequality in (2.90) and still the fact that the denominator is $R(\theta,1,x)$ to get

$$|I^n_1|\le R(\theta,h^{\frac32},x). \tag{2.115}$$

We now deal with the numerator of I^n_3, the denominator is still $R(\theta,1,x)$ as a consequence of Lemma 23. In the second expected value we apply Holder inequality, the boundedness of $|\varphi|$ and first inequality of (2.90) while on the first we use Proposition 19 with Z_h that this time is $\frac{\partial_\mu X^\theta_h}{h}$. We get

$$|I^n_3|\le R(\theta,h^{\frac52-\beta-\epsilon},x). \tag{2.116}$$

From (2.114), (2.115) and (2.116) it follows the first inequality in (2.27).
In order to prove the second inequality in (2.27) we compute $\partial^2_\sigma m(\mu,\sigma,x)$, getting 10 terms as in (2.113) in which the derivatives are always with respect to σ. We therefore say that $\partial^2_\sigma m(\mu,\sigma,x):=\sum_{j=1}^{10}\tilde{I}^n_j$. On $\tilde{I}^n_2$ and $\sum_{j=4}^{10}\tilde{I}^n_j$ we still use Proposition 19 taking as Z respectively $\frac{(\partial_\sigma X^\theta_h)^2}{h}$, $X^\theta_h\frac{\partial^2_\sigma X^\theta_h}{h^{\frac12}}$, $X^\theta_h\frac{(\partial_\sigma X^\theta_h)^2}{h}$, $\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$, $X^\theta_h\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$ in the first and $\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$ in the second expected value of $\tilde{I}^n_7$, $\frac{\partial^2_\sigma X^\theta_h}{h^{\frac12}}$, $\frac{(\partial_\sigma X^\theta_h)^2}{h}$, $\frac{\partial_\sigma X^\theta_h}{h^{\frac12}}$ in both the first and

the second expected values of $\tilde{I}_{10}^n$. Recalling also that $|\partial_\sigma m(\mu,\sigma,x)| \le R(\theta,\Delta_{n,i},x)$ it follows

$$|\tilde{I}_2^n + \sum_{j=4}^{10} \tilde{I}_j^n| \le R(\theta, h^{2-\beta-\epsilon}, x) + R(\theta, h^{\frac{3}{2}-\beta-\epsilon}, x) + R(\theta, h^{2-2\beta-\epsilon}, x) + R(\theta, h^{\frac{5}{2}-\beta-\epsilon}, x) +$$
$$R(\theta, h^{3-2\beta-\epsilon}, x) = R(\theta, h^{\frac{3}{2}-\beta-\epsilon}, x). \tag{2.117}$$

Concerning the numerator of $\tilde{I}_3^n$, using Holder inequality, first inequality in (2.90), the boundedness of φ and Proposition 19 for $Z = \frac{\partial_\sigma X_h^\theta}{h^{\frac{1}{2}}}$ it follows

$$|\tilde{I}_3^n| \le R(\theta, h^{2-\beta-\epsilon}, x). \tag{2.118}$$

We now have to study $\tilde{I}_1^n$. We replace the dynamic (2.98) isolating the principal term: $\partial_\sigma^2 X_h^\theta =: \int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s + G_{\sigma\sigma}$.
$G_{\sigma\sigma}$ is the negligible part and, as a consequence of (2.99), (in which we recall that $ch^{\frac{p}{2}}(1+|x|^c)$ comes from the principal term), we already know that

$$\mathbb{E}[|G_{\sigma\sigma}|^p] \le c(1+|x|^c)(h^{2p} + h^{\frac{3}{2}p} + h^p + h^{p+1}) + c(h^{p-1} + h^{\frac{p}{2}-1} + 1)\int_0^h \mathbb{E}[|\partial_\sigma^2 X_s^\theta|^p] ds \le$$
$$\le c(1+|x|^c) h^{\frac{p}{2}+1}. \tag{2.119}$$

Therefore, $\forall p \ge 2$, $\mathbb{E}[|G_{\sigma\sigma}|^p]^{\frac{1}{p}} \le R(\theta, h^{\frac{1}{2}+\frac{1}{p}}, x)$.
We can see $\tilde{I}_1^n$ in the following way:

$$\tilde{I}_1^n = \mathbb{E}[\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s] +$$
$$+\mathbb{E}[(\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s)(\varphi_{h^\beta}(X_h^\theta - x) - 1)] + \mathbb{E}[G_{\sigma\sigma}\varphi_{h^\beta}(X_h^\theta - x)].$$

The first expected value is zero, on the second we use Holder inequality, (2.50) and (2.65) to get it is upper bounded by $R(\theta, h^{\frac{3}{2}-\epsilon}, x)$. On the third term here above we use Cauchy-Schwartz inequality, the boundedness of φ and (2.119) for $p=2$ getting it is $R(\theta, h, x)$.
It follows second inequality in (2.27). Equation (2.28) has already been proved in Proposition 8 in Chapter 1.
Concerning the second derivatives of m_2, it is $\partial^2_{\sigma\mu} m_2 =$

$$= \frac{2\mathbb{E}[(\partial_\mu X_h^\theta - \partial_\mu m)(\partial_\sigma X_h^\theta - \partial_\sigma m)\varphi]}{\mathbb{E}[\varphi]} + \frac{2\mathbb{E}[(X_h^\theta - m)(\partial^2_{\mu\sigma} X_h^\theta - \partial^2_{\mu\sigma} m)\varphi]}{\mathbb{E}[\varphi]} +$$
$$+ \frac{2\mathbb{E}[(X_h^\theta - m)(\partial_\mu X_h^\theta - \partial_\mu m) h^{-\beta}\, \partial_\sigma X_h^\theta\, \varphi']}{\mathbb{E}[\varphi]} +$$
$$- \frac{2\mathbb{E}[(X_h^\theta - m)(\partial_\mu X_h^\theta - \partial_\mu m) h^{-\beta}\, \varphi]\mathbb{E}[\partial_\sigma X_h^\theta\, \varphi']}{(\mathbb{E}[\varphi])^2} +$$
$$+ \frac{2\mathbb{E}[(X_h^\theta - m)(\partial_\sigma X_h^\theta - \partial_\sigma m) h^{-\beta}\, \partial_\mu X_h^\theta\, \varphi']}{\mathbb{E}[\varphi]} + \frac{\mathbb{E}[(X_h^\theta - m)^2(\partial_\mu X_h^\theta) h^{-2\beta}\, (\partial_\sigma X_h^\theta)\, \varphi'']}{\mathbb{E}[\varphi]} +$$

$$+\frac{\mathbb{E}[(X_h^\theta - m)^2(\partial^2_{\mu\sigma}X_h^\theta)h^{-\beta}\varphi']}{\mathbb{E}[\varphi]} - \frac{\mathbb{E}[(X_h^\theta - m)^2(\partial_\mu X_h^\theta)h^{-\beta}\,\varphi']\mathbb{E}[h^{-\beta}\partial_\sigma X_h^\theta\,\varphi']}{(\mathbb{E}[\varphi])^2}+$$

$$-\partial_\sigma m_2\frac{\mathbb{E}[h^{-\beta}\partial_\mu X_h^\theta\,\varphi']}{\mathbb{E}[\varphi]} - m_2\frac{\mathbb{E}[h^{-2\beta}(\partial_\mu X_h^\theta)(\partial_\sigma X_h^\theta)\,\varphi'']}{\mathbb{E}[\varphi]}+ \tag{2.120}$$

$$-m_2\frac{\mathbb{E}[h^{-\beta}(\partial^2_{\mu\sigma}X_h^\theta)\,\varphi']}{\mathbb{E}[\varphi]} + m_2\frac{\mathbb{E}[h^{-2\beta}(\partial_\mu X_h^\theta)\,\varphi']\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi']}{(\mathbb{E}[\varphi])^2} =: \sum_{j=1}^{12} I_j^n.$$

On I_1^n we use the first and the second point of Proposition 16 to say that both the derivatives of m are $R(\theta, h, x)$. From the boundedness of φ, Lemma 23 and the estimation of the derivatives of X gathered in Lemma 24 it follows

$$|-2\mathbb{E}[(\partial_\mu X_h^\theta - \partial_\mu m)\partial_\sigma m\varphi]| \le R(\theta, h^2, x).$$

We now have to study

$$2\mathbb{E}[\partial_\mu X_h^\theta \partial_\sigma X_h^\theta \varphi] - 2\partial_\mu m\mathbb{E}[\partial_\sigma X_h^\theta \varphi] =: I_{1,1}^n + I_{1,2}^n.$$

We start considering $I_{1,2}^n$. As we have already done after (2.106), we see $\partial_\sigma X_h^\theta$ as $\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s + G_\sigma$, hence

$$|I_{1,2}^n| \le R(\theta, h, x)|\mathbb{E}[(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)(\varphi - 1)] + \mathbb{E}[G_\sigma\varphi]| \le$$

$$\le R(\theta, h, x)[R(\theta, h^{\frac{3}{2}-\epsilon}, x) + R(\theta, h, x)] = R(\theta, h^2, x),$$

where in the last inequality we have used on the first term Holder inequality, Burkholder - Davis - Gundy inequality and (2.65) while on the second we have used Cauchy - Schwartz inequality and the fact that the 2- norm of G_σ is upper bounded by a $R(\theta, h, x)$ function as a consequence of (2.95), (2.96) and (2.97).
Concerning $I_{1,1}^n$, replacing again $\partial_\sigma X_h^\theta$ and using the estimation on the 2-norm of G_σ as we have already done it follows

$$|I_{1,1}^n| \le c|\mathbb{E}[(\int_0^h \partial_\sigma a(X_s^\theta, \sigma)dW_s)\partial_\mu X_h^\theta\varphi]| + R(\theta, h^2, x).$$

Now we observe we have already proved in the conclusion of Proposition 16, starting below (2.111), that $I_{1,1}^\mu := 2\mathbb{E}[[(\int_0^h a(X_s^\theta, \sigma)dW_s)\partial_\mu X_h^\theta\varphi]$ is such that $|I_{1,1}^\mu| \le R(\theta, h^2, x)$. Acting exactly in the same way, considering now $\partial_\sigma a(X_s^\theta, \sigma)$ in the stochastic integral instead of $a(X_s^\theta, \sigma)$, we get that $|I_{1,1}^n| \le R(\theta, h^2, x)$ and so, using also Lemma 23, $|I_1^n| \le R(\theta, h^2, x)$.
Considering I_2^n, it is $I_2^n =: I_{2,1}^n + I_{2,2}^n$, where $I_{2,1}^n := \frac{2\mathbb{E}[(X_h^\theta - m)(\partial^2_{\mu\sigma}X_h^\theta)\varphi]}{\mathbb{E}[\varphi]}$ and
$I_{2,2}^n =: (-\partial^2_{\mu\sigma}m)\frac{2\mathbb{E}[(X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}$.
Now on $I_{2,1}^n$ we use Cauchy-Schwartz inequality, first point of Lemma 14 and (2.90) getting $|I_{2,1}^n| \le R(\theta, h^2, x)$, while on $I_{2,2}^n$ we use the third point of Lemma 14 and (2.113) we have just proved in order to obtain $|I_{2,2}^n| \le R(\theta, h^{\frac{5}{2}}, x)$. The application of Holder inequality, Lemma 15, Lemma 24 and the first point of Proposition 16 on the numerator of I_3^n gives us

$$|I_3^n| \le h^{-\beta}R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}}R(\theta, h^{\frac{3}{2}}, x).$$

It is enough to take p next to 1 to get it is negligible compared to $R(\theta, h^2, x)$. We act on the first expected value of I_4^n like we did on I_2^n while on the second we use Proposition 19. It yields

$$|I_4^n| \le h^{-\beta} R(\theta, h^{\frac{3}{2}-\epsilon}, x) R(\theta, h^{\frac{3}{2}}, x) = R(\theta, h^{3-\beta-\epsilon}, x).$$

On I_5^n we act like we did on I_3^n, hence

$$|I_5^n| \le h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{3}{2}-\epsilon}, x),$$

that is negligible. In the same way

$$|I_6^n| \le h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{3}{2}-\epsilon}, x),$$

$$|I_7^n| \le h^{-\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{3}{2}-\epsilon}, x) \quad \text{and}$$
$$|I_8^n| \le h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^{1-\epsilon}, x) R(\theta, h^{\frac{3}{2}-\epsilon}, x).$$

We recall that $|\partial_\sigma m_2(\mu, \sigma, x)| \le R(\theta, h, x)$ and $|m_2(\mu, \sigma, x)| \le R(\theta, h, x)$, therefore

$$|I_9^n| \le h^{-\beta} R(\theta, h, x) R(\theta, h^{2-\epsilon}, x) = R(\theta, h^{3-\beta-\epsilon}, x),$$

$$|I_{10}^n| \le h^{-2\beta} R(\theta, h, x) R(\theta, h^{\frac{5}{2}-\epsilon}, x) = R(\theta, h^{\frac{7}{2}-2\beta-\epsilon}, x),$$
$$|I_{11}^n| \le h^{-\beta} R(\theta, h, x) R(\theta, h^{\frac{5}{2}-\epsilon}, x) = R(\theta, h^{\frac{7}{2}-\beta-\epsilon}, x),$$
$$|I_{12}^n| \le h^{-2\beta} R(\theta, h, x) R(\theta, h^{2-\epsilon}, x) R(\theta, h^{\frac{3}{2}-\epsilon}, x) = R(\theta, h^{\frac{9}{2}-2\beta-\epsilon}, x).$$

First inequality in (2.29) follows. In order to show the second one we should compute $\partial^2_\mu m_2(\mu, \sigma, x)$. Since it is exactly like $\partial^2_{\mu\sigma} m_2(\mu, \sigma, x)$ but with all the derivatives with respect to μ, we still refer to (2.120) and we will write $\partial^2_\mu m_2(\mu, \sigma, x) =: \sum_{j=1}^{12} \tilde{I}_j^n$. On $\tilde{I}_1^n$, $\sum_{j=3}^{12} \tilde{I}_j^n$ we act exactly like we did here above, getting

$$|\tilde{I}_1^n + \sum_{j=3}^{12} \tilde{I}_j^n| \le R(\theta, h^2, x) + h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h^2, x) + h^{-\beta} R(\theta, h^{\frac{3}{2}}, x) R(\theta, h^{2-\epsilon}, x) +$$

$$+ h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h^2, x) + h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^2, x) +$$
$$+ h^{-\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h, x) + h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h, x) R(\theta, h^{2-\epsilon}, x) +$$
$$+ h^{-\beta} R(\theta, h^2, x) R(\theta, h^{2-\epsilon}, x) + h^{-2\beta} R(\theta, h, x) R(\theta, h^{3-\epsilon}, x) +$$
$$+ h^{-\beta} R(\theta, h, x) R(\theta, h^{2-\epsilon}, x) + h^{-2\beta} R(\theta, h, x) R(\theta, h^{4-\epsilon}, x),$$

that is a $R(\theta, h^2, x)$ function as a consequence of the fact that, choosing p next to 1, all terms are negligible compared to the first one.
We now deal with $\tilde{I}_2^n$. We still need to split it in $\tilde{I}_{2,1}^n$ and $\tilde{I}_{2,2}^n$, where

$$\tilde{I}_{2,1}^n := \frac{2\mathbb{E}[(X_h^\theta - m)(\partial^2_\mu X_h^\theta)\varphi]}{\mathbb{E}[\varphi]}$$

and

$$\tilde{I}_{2,2}^n =: (-\partial^2_\mu m) \frac{2\mathbb{E}[(X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}.$$

Using on $\tilde{I}_{2,2}^n$ Lemma 14 and (2.28) we obtain $|\tilde{I}_{2,2}^n| \le R(\theta, h^2, x)$.
In order to estimate $\tilde{I}_{2,1}^n$ we isolate the principal term in the dynamic of $\partial_\mu^2 X_h^\theta$, getting $\partial_\mu^2 X_h^\theta = \int_0^h \partial_\mu^2 b(X_s^\theta, \mu) ds + G_{\mu\mu}$, where $G_{\mu\mu}$ is the negligible part and it is such that, for $p \ge 2$,

$$\mathbb{E}[|G_{\mu\mu}|^p] \le c(1 + |x|^c)(h^{2p} + h^{p+1}) = R(\theta, h^{p+1}, x), \tag{2.121}$$

as showed in the proof of Lemma 9 in Chapter 1.
Replacing in the definition of $\tilde{I}_{2,1}^n$ we have

$$|\tilde{I}_{2,1}^n| \le |\frac{2\mathbb{E}[G_{\mu\mu}(X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}| + |\frac{2\mathbb{E}[\int_0^h \partial_\mu^2 b(X_s^\theta, \mu) ds (X_h^\theta - m)\varphi]}{\mathbb{E}[\varphi]}|.$$

Now on the first term here above we use Cauchy-Schwartz inequality, the first point of Lemma 14 and (2.121) obtaining it is upper bounded by $R(\theta, h^{\frac{3}{2}}, x) R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^2, x)$, while the second one is upper bounded by

$$ch\partial_\mu^2 b(x, \mu)|\mathbb{E}[(X_h^\theta - m)\varphi]| + c|\mathbb{E}[\int_0^h [\partial_\mu^2 b(X_s^\theta, \mu) - \partial_\mu^2 b(x, \mu)] ds (X_h^\theta - m)\varphi]| \le$$

$$\le hR(\theta, h, x) + R(\theta, h^{\frac{3}{2}}, x) R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^2, x),$$

where we have also used Lemma 23 in order to say that the denominator is lowed bounded by $1 + R(\theta, h, x)$, Cauchy-Schwartz inequality, the first point of Lemma 14 and an estimation analogous to (2.76) for the increments of the derivative of b. It follows $|\tilde{I}_{2,1}^n| \le R(\theta, h^2, x)$.
Second inequality in (2.29) follows.

We now want to show (2.30). To do it, we need to compute $\partial_\sigma^2 m_2(\mu, \sigma, x)$. Again, we still refer to (2.120) considering all the derivatives with respect to σ, writing $\partial_\sigma^2 m_2(\mu, \sigma, x) =: \sum_{j=1}^{12} \hat{I}_j^n$. From $\hat{I}_3^n$ to $\hat{I}_{12}^n$ all the terms are negligible. Indeed, acting like we did on both $\sum_{j=3}^{12} I_j^n$ and $\sum_{j=3}^{12} \tilde{I}_j^n$ we have

$$|\sum_{j=3}^{12} \hat{I}_j^n| \le h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h, x) + h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h^{2-\epsilon}, x) +$$

$$+ h^{-\beta} R(\theta, h^{1+\beta p}, x)^{\frac{1}{p}} R(\theta, h, x) + h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h, x) +$$

$$+ h^{-\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{1}{2}}, x) + h^{-2\beta} R(\theta, h^{1+2\beta p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{1}{2}}, x) R(\theta, h^{\frac{3}{2}-\epsilon}, x) +$$

$$+ h^{-\beta} R(\theta, h, x) R(\theta, h^{\frac{3}{2}-\epsilon}, x) + h^{-2\beta} R(\theta, h, x) R(\theta, h^{2-\epsilon}, x) +$$

$$+ h^{-\beta} R(\theta, h, x) R(\theta, h^{\frac{3}{2}-\epsilon}, x) + h^{-2\beta} R(\theta, h, x) R(\theta, h^{3-2\epsilon}, x),$$

that is always negligible compared to $R(\theta, h^{\frac{3}{2}, x})$ taking p next to 1. We now consider $\hat{I}_1^n$. Its numerator is

$$\mathbb{E}[(\partial_\sigma X_h^\theta)^2 \varphi] + (\partial_\sigma m)^2 \mathbb{E}[\varphi] - 2(\partial_\sigma m)\mathbb{E}[(\partial_\sigma X_h^\theta)\varphi] =: \hat{I}_{1,1}^n + \hat{I}_{1,2}^n + \hat{I}_{1,3}^n.$$

We start considering $\hat{I}_{1,1}^n$, in which we replace the dynamic of

$$\partial_\sigma X_h^\theta := \int_0^h \partial_\sigma a(\sigma, X_s^\theta) dW_s + G_\sigma,$$

where we have already seen that, for $p \geq 2$,

$$\mathbb{E}[|G_\sigma|^p] \leq c(h^{\frac{3p}{2}} + h^p + h^{\frac{p}{2}+1})(1+|x|^c) = R(\theta, h^{\frac{p}{2}+1}, x).$$

We therefore can say that, taking $p = 2$ and using (2.50), we have

$$\hat{I}_{1,1}^n = \mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta) dW_s)^2 \varphi] + R(\theta, h^{\frac{3}{2}}, x) + R(\theta, h^2, x).$$

The first term here above can be seen as

$$\mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta) dW_s)^2(\varphi - 1)] + \mathbb{E}[\int_0^h [(\partial_\sigma a)^2(\sigma, X_s^\theta) - (\partial_\sigma a)^2(\sigma, x)]ds] + h(\partial_\sigma a)^2(\sigma, x).$$

Acting like we did in (2.67) - (2.69), we have

$$|\mathbb{E}[(\int_0^h \partial_\sigma a(\sigma, X_s^\theta) dW_s)^2(\varphi - 1)] + \mathbb{E}[\int_0^h [(\partial_\sigma a(\sigma, X_s^\theta))^2 - (\partial_\sigma a(\sigma, x))^2]ds]| \leq R(\theta, h^{\frac{3}{2}}, x)$$

and so we get

$$\hat{I}_{1,1}^n = h(\partial_\sigma a(\sigma, x))^2 + R(\theta, h^{\frac{3}{2}}, x).$$

On $\hat{I}_{1,2}^n$ we use the second point of Proposition 16 and Lemma 23 to get $|\hat{I}_{1,2}^n| \leq R(\theta, h^2, x)$. In the same way, from second point of Proposition 16, the boundedness of φ and (2.90) we get $|\hat{I}_{1,3}^n| \leq R(\theta, h^{\frac{3}{2}}, x)$. It follows $\hat{I}_1^n = h(\partial_\sigma a(\sigma, x))^2 + R(\theta, h^{\frac{3}{2}}, x)$. We now study $\hat{I}_2^n$, that the denominator is still $1 + R(\theta, h, x)$ as a consequence of Lemma 23 while the numerator can be seen as

$$2\mathbb{E}[(X_h^\theta - m)\partial_\sigma^2 X_h^\theta \varphi] - 2\partial_\sigma^2 m \mathbb{E}[(X_h^\theta - m)\varphi] =: \hat{I}_{2,1}^n + \hat{I}_{2,2}^n.$$

In order to deal with $\hat{I}_{2,1}^n$ we consider the reformulation (2.66), so we have
$2\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)(\partial_\sigma^2 X_h^\theta)\varphi] + 2\mathbb{E}[(B_h)(\partial_\sigma^2 X_h^\theta)\varphi]$.
The second, as a consequence of Holder inequality, the definition of B_h, Lemma 16 used to estimate the jumps and the first inequality in (2.90) is upper bounded by $R(\theta, h^{(1+\beta p)\wedge p}, x)^{\frac{1}{p}} R(\theta, h^{\frac{1}{2}}, x)$. Taking p next to 1 it follows its order is $h^{\frac{3}{2}}$.
In order to study $\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)(\partial_\sigma^2 X_h^\theta)\varphi]$ we introduce once again the notation used above (2.119), for which $\partial_\sigma^2 X_h^\theta = \int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s + G_{\sigma\sigma}$. Hence, we have

$$\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)(\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s)\varphi] + \mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)(G_{\sigma\sigma})\varphi].$$

The second term here above is, using Cauchy-Schwartz inequality, (2.50), the boundedness of φ and (2.119), just $R(\theta, h^{\frac{3}{2}}, x)$ while the first one is

$$\mathbb{E}[(\int_0^h a(X_s^\theta, \sigma) dW_s)(\int_0^h \partial_\sigma^2 a(\sigma, X_s^\theta) dW_s)(\varphi - 1)] + \mathbb{E}[\int_0^h a(X_s^\theta, \sigma)\partial_\sigma^2 a(\sigma, X_s^\theta) ds] =$$

$$= R(\theta, h^2, x) + \mathbb{E}[\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]\partial_\sigma^2 a(\sigma, X_s^\theta) ds] +$$

$$+ a(x, \sigma)\mathbb{E}[\int_0^h [\partial_\sigma^2 a(\sigma, X_s^\theta) - \partial_\sigma^2 a(\sigma, x)]ds)] + h a(x, \sigma)\partial_\sigma^2 a(\sigma, x),$$

where we have also used Holder inequality, (2.50) and (2.65). Now we observe that

$$|\mathbb{E}[\int_0^h [a(X_s^\theta, \sigma) - a(x, \sigma)]\partial_\sigma^2 a(\sigma, X_s^\theta) ds]| \leq c \int_0^h \mathbb{E}[|X_s^\theta - x|^2]^{\frac{1}{2}} \mathbb{E}[|\partial_\sigma^2 a(\sigma, X_s^\theta)|^2]^{\frac{1}{2}} ds \leq$$

$$\leq \int_0^h R(\theta, h^{\frac{1}{2}}, x) = R(\theta, h^{\frac{3}{2}}, x).$$

Acting in the same way on $a(x,\sigma)\mathbb{E}[\int_0^h [\partial_\sigma^2 a(\sigma, X_s^\theta) - \partial_\sigma^2 a(\sigma, x)]ds)]$ it follows that

$$\hat{I}_{2,1}^n = ha(x,\sigma)\partial_\sigma^2 a(\sigma, x) + R(\theta, h^{\frac{3}{2}}, x).$$

On $\hat{I}_{2,2}^n$ we use the second inequality in (2.27) and the third point of Lemma 14 and so we obtain

$$|\hat{I}_{2,2}^n| \leq R(\theta, h^2, x),$$

the result follows. □

2.8.2.3 Proof of Proposition 18

Proof. We define $\partial_\mu^2 m_2(\mu,\sigma,x)$ as $\sum_{j=1}^{12} \tilde{I}_j^n =: \sum_{j=1}^{12} \frac{N_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}}$ and we recall that, as stated in Proposition 17, $|\tilde{I}_j^n| \leq R(\theta, \Delta_{n,i}^2, x)$. We can therefore see its derivative with respect to μ in the following way:

$$\partial_\mu^3 m_2(\mu,\sigma,x) = \sum_{j=1}^{12} (\frac{\partial_\mu N_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}} - \tilde{I}_j^n \frac{\partial_\mu D_{\tilde{I}_j^n}}{D_{\tilde{I}_j^n}}). \tag{2.122}$$

We start considering the second term here above: we remind that $D_{\tilde{I}_j^n}$ can be $\mathbb{E}[\varphi]$ or $(\mathbb{E}[\varphi])^2$. In both cases its derivative is, using Proposition 19, a $R(\theta, h^{2-\beta-\epsilon}, x)$ function; which makes the second term of (2.122) upper bounded by

$$\sum_{j=1}^{12} |\tilde{I}_j^n| R(\theta, h^{2-\beta-\epsilon}, x) \leq R(\theta, h^{4-\beta-\epsilon}, x),$$

as a consequence of the second inequality of (2.29).
Concerning the first term of (2.122), we know that $D_{\tilde{I}_j^n}$ is a $R(\theta, 1, x)$ function because of Lemma 23, while the magnitude of the derivative of the numerator does not get bigger since the order of X_h^θ and m remains the same by deriving them once more (as gathered in the second inequality of (2.92) and in Remark 10 of Chapter 1) and the fact that, by deriving φ, there comes out $h^{-\beta}\varphi' \partial_\mu X_h^\theta$. From the first inequality in (2.89) such terms are negligible compared to the ones we have already studied.
The first term of (2.122) is therefore still a $R(\theta, h^2, x)$ function.
The same reasoning applies to the study of the other seven third derivatives. □

2.8.3 Development of $m_2(\mu, \sigma, x)$

In order to prove our main results we need a development of m_2, that we stated in Ad and we find, under the choice of a particular oscillating function φ, in Propositions 10, 11 and 12.
The main tools is the iteration of the Dynkin's formula that provides us the following expansion for every function f: $\mathbb{R} \to \mathbb{R}$ such that f is in $C^{2(k+1)}$:

$$\mathbb{E}[f(X_{t_{i+1}}^\theta)|X_{t_i}^\theta = x] = \tag{2.123}$$

$$= \sum_{j=0}^{k} \frac{\Delta_{n,i}^j}{j!} A^j f(x) + \int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} ... \int_{t_i}^{u_k} \mathbb{E}[A^{k+1} f(X_{u_{k+1}}^\theta) | X_{t_i}^\theta = x] \, du_{k+1} ... du_2 \, du_1$$

where A denotes the generator of the diffusion, setting $A^0 = Id$. A is the sum of the continuous and discrete part: $A := A_c + A_d$, with

$$A_c f(x) = \frac{1}{2} a^2(x, \sigma) f''(x) + \bar{b}(x, \mu) f'(x);$$

$\bar{b}(x, \mu) = b(x, \mu) - \int_{\mathbb{R}} z \gamma(x) F(z) dz$ and

$$A_d f(x) = \int_{\mathbb{R}} (f(x + \gamma(x) z) - f(x)) \, F(z) dz.$$

2.8.3.1 Proof of Proposition 10.

Proof. We first of all observe that, by the definition of m_2 and m it is

$$m_2(\mu, \sigma, x) = \frac{\mathbb{E}[(X_{t_{i+1}}^\theta)^2 \varphi_{\Delta_{n,i}^\beta} (X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta} (X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]} - (m(\mu, \sigma, x))^2.$$

It has already been showed in Proposition 2 of Chapter 1 that

$$m(\mu, \sigma, x) = x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!} + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x),$$

where $A_{K_1}^{(k)}(x) = \bar{A}^c(h_1)(x)$, with $h_1(y) = (y - x)$.
Acting like we did in Proposition 2 of Chapter 1, we want to find a development for

$$\frac{\mathbb{E}[(X_{t_{i+1}}^\theta)^2 \varphi_{\Delta_{n,i}^\beta} (X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta} (X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]} =: \frac{n_{\Delta_{n,i}}(x)}{d_{\Delta_{n,i}}(x)}.$$

First we focus on the expression of n. We define the set of rest functions $\mathcal{R}^p$ in the following way:

$$\mathcal{R}^p := \Big\{ r(x, y, \Delta_{n,i}^p, \theta) \text{ s. t. } \forall l \geq 0, \forall l' \in \{0, 1\}$$
$$\exists c, | \frac{\partial^l \, \partial^{l'}}{\partial y^l \, \partial \vartheta^{l'}} r(x, y, \Delta_{n,i}^p, \theta) | \leq c(1 + |x|^c + |y|^c) \Delta_{n,i}^p \Big\}$$

for $\vartheta = \mu$ and $\vartheta = \sigma$. It is worth noting that, if $r \in \mathcal{R}^p$, then both $A_c r$ and $A_d r$ are in $\mathcal{R}^p$, where the integral-differential operators A_c and A_d are applied with respect to the second variable y (see details in proof of Proposition 2 in Chapter 1).
We also introduce the set $\tilde{\mathcal{F}}^p$:

$$\tilde{\mathcal{F}}^p := \left\{ \tilde{g}(y, \theta) \; s.t. \, \tilde{g}(y, \theta) = \sum_{k=0}^{p} \varphi^{(k)}((y - x) \Delta_{n,i}^{-\beta}) \Delta_{n,i}^{-k\beta} (\sum_{j=0}^{k} h_{k,j}(x, y, \theta) \Delta_{n,i}^{\beta j}) \right\}$$

where, $\forall k, j$, $\forall l \geq 0$, $\forall l' \in \{0, 1\}$, $\exists c$ such that, for $\vartheta = \mu$ and $\vartheta = \sigma$,

$$\sup_{\theta \in \Theta} | \frac{\partial^l \, \partial^{l'}}{\partial y^l \, \partial \vartheta^{l'}} h_{k,j}(x, y, \theta) | \leq c(1 + |x|^c + |y|^c),$$

with c that depends on k, j, l and l'. By the same proof as in Proposition 2 of Chapter 1 it is possible to prove that, if $\tilde{g} \in \tilde{\mathcal{F}}^p$ then $A_c\tilde{g} \in \tilde{\mathcal{F}}^{p+2}$ and, for all $\tilde{g} \in \tilde{\mathcal{F}}^p$,

$$A_d\tilde{g}(y,\theta) = -\lambda\tilde{g}(y,\theta) + r(x, y, \Delta_{n,i}^{\beta(M+2-p)}, \theta). \tag{2.124}$$

It has also been proved that, as a consequence of the equation here above, the following relation holds:

$$A_{i_N} \circ ... \circ A_{i_1}\tilde{g}(y) = A_c^{l(i_1,...,i_N)}\tilde{g}(y)(-\lambda)^{N-l(i_1,...,i_N)} + r(x, y, \Delta_{n,i}^{\beta(M+2)-2\beta l(i_1,...,i_N)}),$$

with $l(i_1, ..., i_N)$ the number of c in $\{i_1, ..., i_N\}$, the iterations considered.
We observe that $\bar{g}(y) := y^2\varphi((y-x)\Delta_{n,i}^{-\beta})$ belongs to $\mathcal{F}_0$. To find its development through Dynkin's formula we can act exactly like we did in Proposition 2 of Chapter 1.
We get that the principal term in the development of the numerator is

$$\sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + r(x, x, \Delta_{n,i}^{\beta(M+2)}, \theta), \tag{2.125}$$

with $A_{K_2}(x) = \bar{A}_c(h_2)(x)$ for $h_2(y) = y^2$ and $r(x, y, \Delta_{n,i}^{\beta(M+2)}, \theta) \in \mathcal{R}^{\beta(M+2)}$. The integral rest term in the Dynkin formula is

$$|\int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} ... \int_{t_i}^{u_N} \mathbb{E}[A^{N+1}h_2(X_{u_{N+1}})|X_{t_i} = x]du_{N+1}...du_2du_1| \le R(\theta, \Delta_{n,i}^{(1-2\beta)(N+1)}, x). \tag{2.126}$$

Using (2.125) and (2.126) we have the following development:

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{N} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{N-l} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x) + R(\theta, \Delta_{n,i}^{(1-2\beta)(N+1)}, x). \tag{2.127}$$

If $(N+1)(1-2\beta) \ge \beta(M+2)$, it entails

$$n_{\Delta_{n,i}}(x) = \sum_{l=0}^{\lfloor\beta(M+2)\rfloor} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) \sum_{k'=0}^{\lfloor\beta(M+2)\rfloor} \frac{\Delta_{n,i}^{k'}(-\lambda)^{k'}}{k'!} + R(\theta, \Delta_{n,i}^{\beta(M+2)}, x). \tag{2.128}$$

To get the control (2.12) on the derivatives of (2.128), we show that one can differentiate with respect to the parameters the remainder term.
We remark that, as $r(x, y, \Delta_{n,i}^{\beta(M+2)}, \theta)$ is in $\mathcal{R}^{\beta(M+2)}$ and because of the definition of such a set, we clearly have that for $\vartheta = \mu$ and $\vartheta = \sigma$, the function $\partial_\vartheta r(x, y, \Delta_{n,i}^{\beta(M+2)}, \theta)$ is still a $R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)$ function.
Concerning the derivatives of the integral rest, we have that, since $A^{N+1}\bar{g} \in \tilde{\mathcal{F}}^{2(N+1)}$,

$$A^{N+1}\bar{g}(y) = \sum_{k=0}^{2N} \varphi^{(k)}((y-x)\Delta_{n,i}^{-\beta})\Delta_{n,i}^{-\beta k}(\sum_{j=0}^{k} \tilde{h}_{k,j}(x, y, \theta)\Delta_{n,i}^{\beta j});$$

where $\tilde{h}_{k,j}$ are polynomial functions of a, b and their derivatives. Then, for $\vartheta = \mu$ and $\vartheta = \sigma$ it is

$$\partial_\vartheta \mathbb{E}[A^{N+1}\bar{g}(X_{u_{N+1}})|X_{t_i} = x] =$$
$$= \mathbb{E}[\partial_\vartheta A^{N+1}\bar{g}(X_{u_{N+1}}) + \partial_X A^{N+1}\bar{g}(X_{u_{N+1}})\partial_\vartheta X_{u_{N+1}}|X_{t_i} = x].$$

We use an upper bound on the conditional moments of the derivative of X with respect to both the parameters (see Lemma 24 in the Appendix) and we act as for (2.126) to get a control of the integral rest. From the computation of $\partial_X A^{N+1}\bar{g}(X_{u_{N+1}})$ it shows up an extra $\Delta_{n,i}^{-\beta}$ but we can always choose N such that the derivatives of the integral rest remain negligible compared to $R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)$.
We now consider the denominator $d_{\Delta_{n,i}}(x)$: since it is exactly like it was in the development of m, we have already proved in Proposition 2 of Chapter 1 that its expansion is

$$d_{\Delta_{n,i}}(x) = \sum_{k=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^k}{k!}(-\lambda)^k + r(x, x, \Delta_{n,i}^{\beta(M+2)}), \tag{2.129}$$

where λ is the intensity of jumps defined in the fourth point of Assumption 4 and r is a rest function which belongs to $\mathcal{R}^{\beta(M+2)}$.
Acting exactly as we have done on the numerator we get that the derivatives of $r(x, x, \Delta_{n,i}^{\beta(M+2)})$ are still $R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)$ functions and that the derivatives of the integral rest remain negligible compared to $R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)$.
From the expansion of $d_{\Delta_{n,i}}(x)$ we can say that there exists $k_0 > 0$ such that for $|x| \leq \Delta_{n,i}^{-k_0}$, $d_{\Delta_{n,i}}(x) \geq \frac{1}{2}$ $\forall n, i \leq n$: we are avoiding the possibility that the denominator is in the neighborhood of 0. Hence, for $|x| \leq \Delta_{n,i}^{k_0}$, the development of m_2 is

$$\frac{n_{\Delta_{n,i}}(x)}{d_{\Delta_{n,i}}(x)} - (m(\mu, \sigma, x))^2 = \tag{2.130}$$

$$= \sum_{l=0}^{\lfloor \beta(M+2) \rfloor} \frac{\Delta_{n,i}^l}{l!} A_{K_2}^{(l)}(x) - (x + \sum_{k=1}^{\lfloor \beta(M+2) \rfloor} A_{K_1}^{(k)}(x) \frac{\Delta_{n,i}^k}{k!})^2 + r(x, x, \Delta_{n,i}^{\beta(M+2)}, \theta).$$

The expansion (2.11) follows after remarking that $A_{K_2}^{(0)}(x) = 0$.
Moreover, since the rest term in the development of m comes from the same place as the rest in the fraction just studied, also its derivatives with respect to μ and σ remain $R(\theta, \Delta_{n,i}^{\beta(M+2)}, x)$ functions. The result follows. □

We now prove Proposition 12, Proposition 11 is a consequence of it.

2.8.3.2 Proof of Proposition 12

We recall that

$$m_2(\mu, \sigma, x) = \frac{\mathbb{E}[(X_{t_{i+1}}^\theta - m(\mu, \sigma, X_{t_i}))^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]}{\mathbb{E}[\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}^\theta - X_{t_i}^\theta) | X_{t_i}^\theta = x]} := \frac{\tilde{n}_{\Delta_{n,i}}(x)}{\tilde{d}_{\Delta_{n,i}}(x)}.$$

The expansion (2.15) in Proposition 12 is a consequence of the following two expansions for $\tilde{n}_{\Delta_{n,i}}(x)$ and $\tilde{d}_{\Delta_{n,i}}(x)$,

$$\tilde{n}_{\Delta_{n,i}}(x) =$$

$$= \Delta_{n,i} a^2(x, \sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)} \int_{\mathbb{R}} v^2 \varphi(v) F(\frac{v \Delta_{n,i}^\beta}{\gamma(x)}) dv + \Delta_{n,i}^2 [3\bar{b}^2(x, \mu) + h_2(x, \theta) - \lambda a^2(x, \sigma)] +$$

$$+ \frac{\Delta_{n,i}^{2+\beta} a^2(x, \sigma)}{2\gamma(x)} \int_{u:|u| \leq 2} [2\varphi(u) + u\varphi'(u) + u^2 \varphi''(u)] F(\frac{u \Delta_{n,i}^\beta}{\gamma(x)}) du + R(\theta, \Delta_{n,i}^{(3-2\beta) \wedge (2+\beta)}, x), \tag{2.131}$$

$$\tilde{d}_{\Delta_{n,i}}(x) = 1 - \Delta_{n,i}\lambda + \frac{\Delta_{n,i}^{1+\beta}}{\gamma(x)} \int_{u:|u|\le 2} \varphi(u) F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)}) du + R(\theta, \Delta_{n,i}^{2-2\beta}, x). \quad (2.132)$$

The fact that the order of the remainder term in (2.15) is unchanged by derivation with respect to the parameters μ and σ, will be a consequence of a similar property for the remainder terms appearing in (2.131)–(2.132).

Proof of (2.131). In order to develop the numerator $\tilde{n}_{\Delta_{n,i}}(x)$, we define the function $h_{i,n}(y) := (y - m(\mu,\sigma,x))^2 \varphi_{\Delta_{n,i}^{\beta}}(y-x)$ and we use Dynkin formula (2.123) on it, up to third order. It becomes

$$\mathbb{E}[h_{i,n}(X_{t_{i+1}}^{\theta})|X_{t_i}^{\theta} = x] = h_{i,n}(x) + \Delta_{n,i} A h_{i,n}(x) + \frac{1}{2}\Delta_{n,i}^2 A^2 h_{i,n}(x) + \quad (2.133)$$

$$\frac{1}{6}\int_{t_i}^{t_{i+1}} \int_{t_i}^{u_1} \int_{t_i}^{u_2} \mathbb{E}[A^3 h_{i,n}(X_{u_3}^{\theta})|X_{t_i}^{\theta} = x] du_3 du_2 du_1.$$

We now successively study the contribution of each term in the Dynkin's development.

By the definition of $h_{i,n}$, we have $h_{i,n}(x) = (x - m(\mu,\sigma,x))^2$.

We recall that $Ah_{i,n}(x) = A_c h_{i,n}(x) + A_d h_{i,n}(x)$, where $A_c h_{i,n}(x) = \frac{1}{2}a^2(x,\sigma) h''_{i,n}(x) + \bar{b}(x,\mu) h'_{i,n}(x)$ and

$$A_d h_{i,n}(x) = \int_{\mathbb{R}} [h_{i,n}(x + z\gamma(x)) - h_{i,n}(x)] F(z) dz =$$

$$\int_{\mathbb{R}} (x + z\gamma(x) - m(\mu,\sigma,x))^2 \varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) F(z) dz - (x - m(\mu,\sigma,x))^2 \int_{\mathbb{R}} \varphi_{\Delta_{n,i}^{\beta}}(0) F(z) dz =$$

$$= (x - m(\mu,\sigma,x))^2 \int_{\mathbb{R}} [\varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) - \varphi_{\Delta_{n,i}^{\beta}}(0)] F(z) dz + \quad (2.134)$$

$$+\gamma^2(x) \int_{\mathbb{R}} z^2 \varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) F(z) dz + 2(x - m(\mu,\sigma,x))\gamma(x) \int_{\mathbb{R}} z \varphi_{\Delta_{n,i}^{\beta}}(z\gamma(x)) F(z) dz.$$

Using now the boundedness of φ, the fact that $\int_{\mathbb{R}} F(z)dz = \lambda$ and the development of m (2.31), we have that the first term here above is a $R(\theta, \Delta_{n,i}^2, x)$ function and, from Proposition 16, also its derivatives with respect to both the parameters are $R(\theta, \Delta_{n,i}^2, x)$ functions.

On the second term of (2.134) we apply the change of variable $v := \frac{\gamma(x)z}{\Delta_{n,i}^{\beta}}$, getting $\frac{\Delta_{n,i}^{3\beta}}{\gamma(x)} \int_{\mathbb{R}} v^2 \varphi(v) F(\frac{v\Delta_{n,i}^{\beta}}{\gamma(x)}) dv$. On the third we use the definition of φ for which $\varphi(\zeta) = 0$ for ζ such that $|\zeta| \ge 2$ and again the development of m to get it is upper bounded by

$$|c(x - m(\mu,\sigma,x))\gamma(x) \int_{z:|z| \le \frac{2\Delta_{n,i}^{\beta}}{\gamma(x)}} zF(z)dz| \le R(\theta, \Delta_{n,i}^{1+\beta}, x). \quad (2.135)$$

Again, we can calculate the derivatives with respect to ϑ for both $\vartheta = \mu$ and $\vartheta = \sigma$, getting a term that is still a $R(\theta, \Delta_{n,i}^{1+\beta}, x)$ function by Proposition 16.

It follows

$$A_d h_{i,n}(x) = R(\theta, \Delta_{n,i}^2, x) + \Delta_{n,i}^{2\beta} \int_{\mathbb{R}} v^2 \varphi(v) F(\frac{v\Delta_{n,i}^{\beta}}{\gamma(x)}) dv + R(\theta, \Delta_{n,i}^{1+\beta}, x).$$

In order to compute $A_c h_{i,n}(x)$ we need the derivatives of $h_{i,n}$, they are

$$h'_{i,n}(y) = 2(y - m(\mu,\sigma,x))\varphi_{\Delta_{n,i}^{\beta}}(y-x) + (y - m(\mu,\sigma,x))^2 \Delta_{n,i}^{-\beta} \varphi'_{\Delta_{n,i}^{\beta}}(y-x),$$

$$h''_{i,n}(y) = 2\varphi_{\Delta^\beta_{n,i}}(y-x) + 4(y-m(\mu,\sigma,x))\Delta^{-\beta}_{n,i}\varphi'_{\Delta^\beta_{n,i}}(y-x)+ \tag{2.136}$$

$$+(y-m(\mu,\sigma,x))^2\Delta^{-2\beta}_{n,i}\varphi''_{\Delta^\beta_{n,i}}(y-x),$$

$$h^{(3)}_{i,n}(y) = 6\Delta^{-\beta}_{n,i}\varphi'_{\Delta^\beta_{n,i}}(y-x) + 6(y-m(\mu,\sigma,x))\Delta^{-2\beta}_{n,i}\varphi''_{\Delta^\beta_{n,i}}(y-x)+$$

$$+(y-m(\mu,\sigma,x))^2\Delta^{-3\beta}_{n,i}\varphi^{(3)}_{\Delta^\beta_{n,i}}(y-x),$$

$$h^{(4)}_{i,n}(y) = 12\Delta^{-2\beta}_{n,i}\varphi''_{\Delta^\beta_{n,i}}(y-x) + 8(y-m(\mu,\sigma,x))\Delta^{-3\beta}_{n,i}\varphi^{(3)}_{\Delta^\beta_{n,i}}(y-x)+$$

$$+(y-m(\mu,\sigma,x))^2\Delta^{-4\beta}_{n,i}\varphi^{(4)}_{\Delta^\beta_{n,i}}(y-x);$$

we have calculated the derivatives up to the fourth because they will be useful in the sequel.
Replacing the first two derivatives, calculated in x, it follows

$$A_c h_{i,n}(x) = a^2(x,\sigma) + 2(x-m(\mu,\sigma,x))\bar{b}(x,\mu). \tag{2.137}$$

Therefore we have

$$\mathbb{E}_i[h_{i,n}(X^\theta_{t_{i+1}})] = (x-m(\mu,\sigma,x))^2 + \Delta_{n,i}a^2(x,\sigma) + \frac{\Delta^{1+3\beta}_{n,i}}{\gamma(x)}\int_{\mathbb{R}} v^2\varphi(v)F(\frac{v\Delta^\beta_{n,i}}{\gamma(x)})dv+ \tag{2.138}$$

$$+2\Delta_{n,i}(x-m(\mu,\sigma,x))\bar{b}(x,\mu) + R(\theta,\Delta^{2+\beta}_{n,i},x) + \frac{1}{2}\Delta^2_{n,i}A^2h_{i,n}(x)+$$

$$+\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X^\theta_{u_3})|X^\theta_{t_i}=x]du_3du_2du_1.$$

We now study $A^2h_{i,n}(x)$. We recall it is

$$A^2h_{i,n}(x) = A_c^2h_{i,n}(x) + A_cA_dh_{i,n}(x) + A_dA_ch_{i,n}(x) + A_d^2h_{i,n}(x).$$

We observe that we can write $A_c^2h_{i,n}(x)$ as $\sum_{j=1}^4 h_j(x,\theta)h^{(j)}_{i,n}(x)$, where, for each $j\in\{1,2,3,4\}$, h_j is a function of a, $\bar{b}$ and their derivatives up to second order: $h_1 = \frac{1}{2}a^2\bar{b}'' + \bar{b}\bar{b}'$, $h_2 = \frac{1}{2}a^2(a')^2 + \frac{1}{2}a^3a'' + a^2\bar{b}' + aa'\bar{b} + \bar{b}^2$, $h_3 = a^3a' + a^2\bar{b}$ and $h_4 = \frac{1}{4}a^4$. Moreover, recalling that by the definition of φ we have $\varphi_{\Delta^\beta_{n,i}}(0) = 1$ and $\varphi^{(k)}_{\Delta^\beta_{n,i}}(0) = 0$ $\forall k\geq 1$, it follows $h'_{i,n}(x) = 2(x-m(\mu,\sigma,x))$, $h''_{i,n}(x) = 2$ and $h^{(l)}_{i,n} = 0$ $\forall l\geq 3$. We obtain

$$A_c^2h_{i,n}(x) = 2h_1(x,\theta)(x-m(\mu,\sigma,x)) + 2h_2(x,\theta) = R(\theta,\Delta_{n,i},x) + 2h_2(x,\theta), \tag{2.139}$$

where the last equality is a consequence of the development (2.31) of m.
Concerning $A_cA_dh_{i,n}(x)$, it is $\frac{1}{2}a^2(x,\sigma)(A_dh_{i,n}(x))'' + \bar{b}(x,\mu)(A_dh_{i,n}(x))'$. We start considering

$$(A_dh_{i,n}(x))' = \int_{\mathbb{R}}[h'_{i,n}(x+z\gamma(x))(1+z\gamma'(x)) - h'_{i,n}(x)]F(z)dz.$$

We now observe that, $\forall k\geq 1$ and $\forall y\in\mathbb{R}$, $|y-m(\mu,\sigma,x)|^k|\varphi_{\Delta^\beta_{n,i}}(y-x)| \leq (|y-x|^k + |x-m(\mu,\sigma,x)|^k)|\varphi_{\Delta^\beta_{n,i}}(y-x)|$. We have that $|y-x|^k|\varphi_{\Delta^\beta_{n,i}}(y-x)| \leq c\Delta^{k\beta}_{n,i}|\varphi_{\Delta^\beta_{n,i}}(y-x)|$ as a consequence of the definition of φ while, using the development (2.31) of m , it follows $|x-m(\mu,\sigma,x)|^k|\varphi_{\Delta_{n,i}^\beta}(y-x)| \le R(\theta,\Delta_{n,i}^k,x)$. Putting the pieces together it is

$$|y-m(\mu,\sigma,x)|^k|\varphi_{\Delta_{n,i}^\beta}(y-x)| \le R(\theta,\Delta_{n,i}^{\beta k},x)+R(\theta,\Delta_{n,i}^k,x) = R(\theta,\Delta_{n,i}^{\beta k},x); \quad (2.140)$$

it is easy to remark that the same reasoning holds with the derivatives of φ instead of φ. We underline that from the estimation (2.140) here above and the computation of the derivatives of $h_{i,n}$ we get that, $\forall l \ge 0$, each term of the l-derivative of $h_{i,n}$ has the same size. Indeed, each time we derive φ an extra $\Delta_{n,i}^{-\beta}$ turns out but we can recover it on the basis of (2.140). In particular, $\forall l \ge 0$ it follows

$$\left\| h_{i,n}^{(l)} \right\|_\infty \le R(\theta,\Delta_{n,i}^{\beta(2-l)},x). \quad (2.141)$$

Therefore we obtain

$$|(A_d h_{i,n}(x))'| \le c\Delta_{n,i}^\beta(2\lambda + c\left\|\gamma'\right\|_\infty). \quad (2.142)$$

Concerning the second derivative of $A_d h_{i,n}(x)$, it is

$$(A_d h_{i,n}(x))'' = \int_{\mathbb{R}} [h''_{i,n}(x+z\gamma(x))(1+z\gamma'(x))^2 + h'_{i,n}(x+z\gamma(x))z\gamma''(x) - h''_{i,n}(x)]F(z)dz =$$

$$= \int_{\mathbb{R}} [h''_{i,n}(x+z\gamma(x)) - h''_{i,n}(x) + 2z\gamma'(x)h''_{i,n}(x+z\gamma(x)) + z^2(\gamma'(x))^2 h''_{i,n}(x+z\gamma(x)) +$$

$$+ h'_{i,n}(x+z\gamma(x))z\gamma''(x)]F(z)dz = \sum_{j=1}^5 I_j.$$

On I_3 and I_4 we act like we did on the integral in (2.135), observing that in the computation of $h''_{i,n}$ we always have φ or its derivatives which make the integrals different from 0 only for $|z| \le c\Delta_{n,i}^\beta$. On I_5 we use (2.141) for $l=1$, getting $|I_5| \le R(\theta,\Delta_{n,i}^\beta,x)$. We observe that, by the computation of $h''_{i,n}(x)$ we obtain $I_2 = -2\lambda$.
To conclude the study of $A_c A_d h_{i,n}(x)$ we have to deal with I_1:

$$\int_{\mathbb{R}} h''_{i,n}(x+z\gamma(x))F(z)dz = \int_{\mathbb{R}} 2\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz +$$

$$+4\Delta_{n,i}^{-\beta}(x-m(\mu,\sigma,x))\int_{\mathbb{R}} \varphi'_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz + \Delta_{n,i}^{-\beta}\int_{\mathbb{R}} z\gamma(x)\varphi'_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz + \quad (2.143)$$

$$+\Delta_{n,i}^{-2\beta}\int_{\mathbb{R}} ((x-m(\mu,\sigma,x)) + z\gamma(x))^2\varphi''_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz.$$

Applying the change of variable $u := \frac{z\gamma(x)}{\Delta_{n,i}^\beta}$ and recalling that from the development of m it follows $|x-m(\mu,\sigma,x)|^k \le R(\theta,\Delta_{n,i}^k,x)$ for each $k \ge 1$, we obtain

$$\int_{\mathbb{R}} h''_{i,n}(x+z\gamma(x))F(z)dz =$$

$$= \frac{\Delta_{n,i}^\beta}{\gamma(x)}\int_{\mathbb{R}} (2\varphi(u) + u\varphi'(u) + u^2\varphi''(u))F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du + R(\theta,\Delta_{n,i}^{(1-\beta)\wedge(2-2\beta)\wedge(1-2\beta)},x).$$

It is worth noting that the magnitude of the first term in the left hand side of the equation here above depends on the density F.
Remarking also that $\varphi(u) = 0$ for each u such that $|u| \geq 2$, it follows

$$(A_d h_{i,n}(x))'' = \tag{2.144}$$

$$= \frac{\Delta_{n,i}^{\beta}}{\gamma(x)} \int_{u:|u|\leq 2} (2\varphi(u) + u\varphi'(u) + u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du - 2\lambda + R(\theta, \Delta_{n,i}^{(1-2\beta)\wedge\beta}, x).$$

From the definition of $A_c A_d h_{i,n}(x)$, (2.142) and (2.144) we get

$$A_c A_d h_{i,n}(x) = \tag{2.145}$$

$$= \frac{\Delta_{n,i}^{\beta} a^2(x,\sigma)}{2\gamma(x)} \int_{\mathbb{R}} (2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du - a^2(x,\sigma)\lambda + R(\theta, \Delta_{n,i}^{(1-2\beta)\wedge\beta}, x).$$

Now we deal with

$$A_d A_c h_{i,n}(x) = \int_{\mathbb{R}} [A_c h_{i,n}(x + z\gamma(x)) - A_c h_{i,n}(x)]F(z)dz. \tag{2.146}$$

From (2.137) it follows

$$\int_{\mathbb{R}} A_c h_{i,n}(x)F(z)dz = \lambda a^2(x,\sigma) + 2\lambda(x - m(\mu,\sigma,x))\bar{b}(x,\mu) = \lambda a^2(x,\sigma) + R(\theta, \Delta_{n,i}, x), \tag{2.147}$$

where we have also used the development (2.31) of m. Moreover,

$$\int_{\mathbb{R}} A_c h_{i,n}(x + z\gamma(x))F(z)dz = \tag{2.148}$$

$$= \int_{\mathbb{R}} [\frac{1}{2} a^2(x + z\gamma(x), \sigma) h_{i,n}''(x + z\gamma(x)) + \bar{b}(x + z\gamma(x), \mu) h_{i,n}'(x + z\gamma(x))]F(z)dz.$$

Acting on the first term of the right hand side of the equation here above exactly like we did in (2.143) we get it is equal to

$$\frac{\Delta_{n,i}^{\beta}}{2\gamma(x)} \int_{u:|u|\leq 2} a^2(x + u\Delta_{n,i}^{\beta}, \sigma)(2\varphi(u) + u\varphi'(u) + u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + R(\theta, \Delta_{n,i}^{1-2\beta}, x). \tag{2.149}$$

We upper bound the second term of the right hand side of (2.148), instead, using (2.141) for $l = 1$. It yields

$$|\int_{\mathbb{R}} \bar{b}(x + z\gamma(x), \mu) h_{i,n}'(x + z\gamma(x))F(z)dz| \leq R(\theta, \Delta_{n,i}^{\beta}, x), \tag{2.150}$$

where we have also used the polynomial growth of b. Replacing in (2.146) the equations (2.147) - (2.150) we obtain

$$A_d A_c h_{i,n}(x) = \frac{\Delta_{n,i}^{\beta}}{2\gamma(x)} \int_{u:|u|\leq 2} a^2(x + u\Delta_{n,i}^{\beta}, \sigma)(2\varphi(u) + u\varphi'(u) + u^2\varphi''(u))F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)})du + \tag{2.151}$$

$$-\lambda a^2(x,\sigma) + R(\theta, \Delta_{n,i}^{(1-2\beta)\wedge\beta}, x).$$

We have to study $A_d A_d h_{i,n}(x)$. It is

$$|A_d^2 h_{i,n}(x)| = |\int_{\mathbb{R}} [A_d h_{i,n}(x + z\gamma(x)) - A_d h_{i,n}(x)]F(z)dz| \leq \tag{2.152}$$

$$\leq \int_{\mathbb{R}} \|(A_d h_{i,n})'\|_\infty |z\gamma(x)|F(z)dz \leq R(\theta, \Delta_{n,i}^\beta, x),$$

where we have used the definition $A_d h'_{i,n}$ and (2.141), remarking that the estimation (2.142) holds also in no matter which $y \in \mathbb{R}$.
From (2.139), (2.145), (2.151) and (2.152) it follows

$$\frac{1}{2}\Delta_{n,i}^2 A^2 h_{i,n}(x) = \Delta_{n,i}^2 h_2(x,\theta) - \lambda \Delta_{n,i}^2 a^2(x,\sigma) + \tag{2.153}$$

$$+\frac{\Delta_{n,i}^{2+\beta}}{4\gamma(x)}\int_{u:|u|\leq 2}[a^2(x,\sigma)+a^2(x+u\Delta_{n,i}^\beta,\sigma)](2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du+$$

$$+R(\theta, \Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)}, x).$$

To complete the study of the numerator of m_2 we need to estimate $\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3 h_{i,n}(X_{u_3}^\theta)|X_{t_i}^\theta = x]du_3 du_2 du_1$. We introduce the following norm on $\mathcal{C}^p$ functions, with $p \geq 0$, $c > 0$:

$$\|f\|_{\infty,c,p} := \sum_{k=0}^p \sup_{y\in\mathbb{R}} |\frac{f^{(k)}(y)}{(1+|y|)^c}|.$$

We observe it is

$$\int_{\mathbb{R}} |f(y+z\gamma(y))|F(z)dz = \int_{\mathbb{R}} \frac{|f(y+z\gamma(y))|(1+|y+z\gamma(y)|)^{\tilde{c}}}{(1+|y+z\gamma(y)|)^{\tilde{c}}}F(z)dz \leq$$

$$\leq \|f\|_{\infty,\tilde{c},0}\int_{\mathbb{R}}(1+|y+z\gamma(y)|)^{\tilde{c}}F(z)dz.$$

We can therefore evaluate the norm of $A_d f$, getting

$$\|A_d f\|_{\infty,\tilde{c},0} = \left\|\frac{A_d f}{(1+|y|)^{\tilde{c}}}\right\|_\infty \leq$$

$$\leq c\|f\|_{\infty,\tilde{c},0}\left(\frac{\int_{\mathbb{R}}(1+|y|^{\tilde{c}}+(1+|y|)^{\tilde{c}}|z|^{\tilde{c}})F(z)dz}{(1+|y|)^{\tilde{c}}}+\lambda\right) \leq c\|f\|_{\infty,\tilde{c},0}.$$

By similar computations on the derivatives of $A_d f$ we obtain, $\forall p \geq 0$, with $p \leq 4$, $\|A_d f\|_{\infty,\tilde{c},p} \leq c\|f\|_{\infty,\tilde{c},p}$ (similar calculations are in Theorem 2.3 of [73]).
In order to find an upper bound for the norm of $A_c f$ we observe that, from the polynomial growth of both the coefficients a and b, it follows

$$|\frac{A_c f}{(1+|y|)^{\tilde{c}+2}}| = |\frac{\frac{1}{2}a^2(y,\sigma)f''(y)}{(1+|y|)^{\tilde{c}+2}} + \frac{b(y,\mu)f'(y)}{(1+|y|)^{\tilde{c}+2}}| \leq \frac{c}{(1+|y|)^{\tilde{c}}}|f''(y)| + \frac{c}{(1+|y|)^{\tilde{c}}}|f'(y)|.$$

Hence, we deduce that $\|A_c f\|_{\infty,\tilde{c}+2,0} \leq c\|f\|_{\infty,\tilde{c},2}$. Acting again with similar computation on the following derivatives, as done detailing in the proof of Theorem 2.3 in [73], we get $\|A_c f\|_{\infty,\tilde{c}+2,p} \leq c\|f\|_{\infty,\tilde{c},p+2}$.
We want to use the estimations on $\|A_d f\|_{\infty,\tilde{c},p}$ and $\|A_c f\|_{\infty,\tilde{c},p}$ and equation (2.141) to evaluate each term of
$\mathbb{E}[A^3 h_{i_n}(X_{u_3}^\theta)|X_{t_i}^\theta = x]$ but $A_c^3 h_{i,n}$. We observe it is, for $\tilde{c} \geq 4$,

$$\mathbb{E}_i[A_d^3 h_{i,n}(X_{u_3}^\theta)] \leq \left\|A_d^3 h_{i,n}\right\|_{\infty,\tilde{c},0}\mathbb{E}_i[(1+|X_{u_3}^\theta|)^{\tilde{c}}] \leq c\|h_{i,n}\|_{\infty,\tilde{c},0}R(\theta,1,x) \leq R(\theta,\Delta_{n,i}^{2\beta},x);$$

$$\mathbb{E}_i[A_dA_dA_ch_{i,n}(X^\theta_{u_3})] \leq \|A_dA_dA_ch_{i,n}\|_{\infty,\tilde{c},0}\,\mathbb{E}_i[(1+|X^\theta_{u_3}|)^{\tilde{c}}] \leq$$
$$\leq c\,\|h_{i,n}\|_{\infty,\tilde{c}-2,2}\,R(\theta,1,x) \leq R(\theta,1,x);$$

remarking that the same estimation holds for $A_dA_cA_dh_{i_n}$ and $A_cA_dA_dh_{i_n}$. Moreover we have

$$\mathbb{E}_i[A_cA_cA_dh_{i,n}(X^\theta_{u_3})] \leq \|A_cA_cA_dh_{i,n}\|_{\infty,\tilde{c},0}\,\mathbb{E}_i[(1+|X^\theta_{u_3}|)^{\tilde{c}}] \leq$$
$$\leq c\,\|h_{i,n}\|_{\infty,\tilde{c}-4,4}\,R(\theta,1,x) \leq R(\theta,\Delta_{n,i}^{-2\beta},x)$$

and we can upper bound $A_cA_dA_ch_{i,n}$ and $A_dA_cA_ch_{i,n}$ with the same quantity.
We are now left to study $\mathbb{E}_i[A_c^3h_{i,n}(X^\theta_{u_3})]$. As we have already done considering $A_c^2h_{i,n}(x)$, we see $A_c^3h_{i,n}$ as linear combination of the derivatives of $h_{i,n}$: $A_c^3h_{i,n}(y) := \sum_{j=1}^6 h_j(y,\theta)h_{i,n}^{(j)}(y)$ where for each $j\in\{1,...,6\}$ h_j is a function of a, b and their derivatives up to fourth order.
Using a conditional version of Proposition 19 we get $\forall\epsilon>0$, $\forall j\geq 3$ $\mathbb{E}_i[|h_j(X^\theta_{u_3},\theta)h_{i,n}^{(j)}(X^\theta_{u_3})|] \leq R(\theta,\Delta_{n,i}^{1+(2-j)\beta-\epsilon},x)$.
Concerning the first two terms of the sum, we have

$$|\mathbb{E}_i[h_1(X^\theta_{u_3},\theta)h'_{i,n}(X^\theta_{u_3}) + h_2(X^\theta_{u_3},\theta)h''_{i,n}(X^\theta_{u_3})]| \leq$$
$$\leq \left\|h'_{i,n}\right\|_\infty \mathbb{E}_i[|h_1(X^\theta_{u_3},\theta)|] + \left\|h''_{i,n}\right\|_\infty \mathbb{E}_i[|h_2(X^\theta_{u_3},\theta)|] \leq R(\theta,1,x),$$

which follows from (2.141) and from the polynomial growth of a, b and their derivatives, which constitute the functions h_1 and h_2. Putting all the pieces together we get

$$\frac{1}{6}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u_1}\int_{t_i}^{u_2}\mathbb{E}[A^3h_{i,n}(X^\theta_{u_3})|X^\theta_{t_i}=x]du_3du_2du_1 \leq$$
$$\leq R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(4-4\beta-\epsilon)},x) = R(\theta,\Delta_{n,i}^{3-2\beta},x).$$

From (2.138), (2.153) and the equation here above it follows $\mathbb{E}_i[h_{i,n}(X^\theta_{t_{i+1}})] =$

$$= (x-m(\mu,\sigma,x))^2 + \Delta_{n,i}a^2(x,\sigma) + \frac{\Delta_{n,i}^{1+3\beta}}{\gamma(x)}\int_{\mathbb{R}} v^2\varphi(v)F(\frac{v\Delta_{n,i}^\beta}{\gamma(x)})dv+ \tag{2.154}$$
$$+2\Delta_{n,i}(x-m(\mu,\sigma,x))\bar{b}(x,\mu) + \Delta_{n,i}^2h_2(x,\theta) - \lambda\Delta_{n,i}^2a^2(x,\sigma)+$$
$$+\frac{\Delta_{n,i}^{2+\beta}}{4\gamma(x)}\int_{u:|u|\leq 2}[a^2(x,\sigma)+a^2(x+u\Delta_{n,i}^\beta,\sigma)](2\varphi(u)+u\varphi'(u)+u^2\varphi''(u))F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})du+$$
$$+R(\theta,\Delta_{n,i}^{(3-2\beta)\wedge(2+\beta)},x).$$

The expansion (2.131) follows from (2.154) and (2.16), with the smoothness of $z\mapsto a^2(z,\sigma)$, and $\int_{\mathbb{R}}F(z)dz<\infty$.

□

Proof of (2.132). Concerning the denominator of m_2, we still use Dynkin formula up to third order, this time on $f_{i,n}(y) := \varphi_{\Delta_{n,i}^\beta}(y-x)$. We observe that, by the building, $f_{i,n}(x)=1$ and $f_{i,n}^{(k)}(x)=0$ for each $k\geq 1$. Hence, $A_cf_{i,n}(x)=0$ and

$$A_df_{i,n}(x) = \int_{\mathbb{R}}[f_{i,n}(x+\gamma(x)z)-1]F(z)dz = \int_{z:|z|\leq\frac{2\Delta_{n,i}^\beta}{|\gamma(x)|}}\varphi_{\Delta_{n,i}^\beta}(z\gamma(x))F(z)dz - \lambda.$$

As we have already done we can see the first term here above, after the change of variable $u := \frac{z\gamma(x)}{\Delta_{n,i}^{\beta}}$, as $\frac{\Delta_{n,i}^{\beta}}{\gamma(x)} \int_{u:|u|\le 2} \varphi(u) F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)}) du$, which order depends on the density F.
Concerning the study of $A^2 f_{i,n}(x)$, we first of all remark that $A_c^2 f_{i,n}(x) = 0$.
Moreover, we observe it is $f_{i,n}^{(k)}(y) = \Delta_{n,i}^{-\beta k} \varphi_{\Delta_{n,i}^{\beta}}^{(k)}(y - x)$ and so by the boundedness of φ and its derivatives we get that, for each $k \ge 1$,

$$\left\| f_{i,n}^{(k)} \right\|_{\infty} \le R(\theta, \Delta_{n,i}^{-\beta k}, x). \tag{2.155}$$

We therefore have, $\forall y \in \mathbb{R}$,

$$|(A_d f_{i,n}(y))'| = |\int_{\mathbb{R}} f'_{i,n}(y + z\gamma(y))(1 + z\gamma'(y)) F(z) dz - \lambda f'_{i,n}(y)| \le R(\theta, \Delta_{n,i}^{-\beta}, y)$$

and, in the same way, $|(A_d f_{i,n}(y))''| \le R(\theta, \Delta_{n,i}^{-2\beta}, y)$. It follows

$$|A_c A_d f_{i,n}(x)| = |\frac{1}{2} a^2(x, \sigma)(A_d f_{i,n}(x))'' + b(x, \mu)(A_d f_{i,n}(x))'| \le R(\theta, \Delta_{n,i}^{-2\beta}, x);$$

$$|A_d A_c f_{i,n}(x)| = |\int_{\mathbb{R}} A_c f_{i,n}(x + z\gamma(x)) F(z) dz| \le R(\theta, \Delta_{n,i}^{-2\beta}, x)$$

and, using also finite-increments theorem, $|A_d A_d f_{i,n}(x)| \le R(\theta, \Delta_{n,i}^{-\beta}, x)$. Putting pieces together we have $|\frac{1}{2}\Delta_{n,i}^2 A^2 f_{i,n}(x)| \le R(\theta, \Delta_{n,i}^{2-2\beta}, x)$.
Considering the integral rest of the Dynkin formula , we act like we did in the study of the numerator, passing through the use of the norm $\|.\|_{\infty,c,p}$ and the estimation of the derivatives of $f_{i,n}$ gathered in (2.155). It yields

$$\mathbb{E}_i[A_d^3 f_{i_n}(X_{u_3}^{\theta})] \le R(\theta, 1, x), \qquad \mathbb{E}_i[A_d A_d A_c f_{i_n}(X_{u_3}^{\theta})] \le R(\theta, \Delta_{n,i}^{-2\beta}, x),$$

$$\mathbb{E}_i[A_d A_c A_c f_{i_n}(X_{u_3}^{\theta})] \le R(\theta, \Delta_{n,i}^{-4\beta}, x), \qquad \mathbb{E}_i[A_c^3 f_{i_n}(X_{u_3}^{\theta})] \le R(\theta, \Delta_{n,i}^{1-6\beta-\epsilon}, x),$$

having also used Proposition 19 to get the last one estimation here above. It turns out the denominator is

$$1 - \Delta_{n,i}\lambda + \frac{\Delta_{n,i}^{1+\beta}}{\gamma(x)} \int_{u:|u|\le 2} \varphi(u) F(\frac{u\Delta_{n,i}^{\beta}}{\gamma(x)}) du + R(\theta, \Delta_{n,i}^{2-2\beta}, x); \tag{2.156}$$

since we can always find an $\epsilon > 0$ for which $4 - 6\beta - \epsilon > 2 - 2\beta$ and we have $3 - 4\beta > 2 - 2\beta$. This concludes the proof of the expansion (2.132). □

To conclude the proof of the Proposition 12 we are left to show that the derivatives with respect to both the parameters of the rest terms in the expansions (2.131)–(2.132) are still rest functions and their order remains the same.
We observe that up to the development of second order in Dynkin formula, the rest functions R are totally explicit in our computation and so it is possible to calculate its derivatives with respect to both μ and σ. As we have already seen during the proof, we can use the estimations on $\partial_{\vartheta} m(\mu, \sigma, x)$ for $\vartheta = \mu$ and $\vartheta = \sigma$ gathered in Proposition 16 and the fact that $(x - m(\mu, \sigma, x))$ is a $R(\theta, \Delta_{n,i}, x)$ function to get that size of $h_{i,n}$ and of the rest functions does not change after having derived with respect to the parameters.

Concerning the integral rest coming from the third order of the Dynkin formula, we have that

$$\partial_\vartheta \mathbb{E}_i[A^3 h_{i,n}(X^\theta_{u_3})] = \mathbb{E}_i[\partial_\vartheta A^3 h_{i,n}(X^\theta_{u_3})] + \mathbb{E}_i[\partial_X A^3 h_{i,n}(X^\theta_{u_3})\partial_\vartheta X].$$

On the first term of the right hand side here above we can act exactly like we did on $\mathbb{E}_i[A^3 h_{i,n}(X^\theta_{u_3})]$ getting a rest function whose order does not change, while from the computation of $\partial_X A^3 h_{i,n}(X^\theta_{u_3})$ an extra $\Delta_{n,i}^{-\beta}$ appears but, since from Lemma 24 the norm 1 of $\partial_\vartheta X$ is $R(\theta, \Delta_{n,i}, x)$ for $\vartheta = \mu$ and $R(\theta, \Delta_{n,i}^{\frac{1}{2}}, x)$ for $\vartheta = \sigma$, it is enough to use previously Holder inequality and observe that both $\frac{1}{2} - \beta$ and $1 - \beta$ are positive to get that the second term here above is negligible compared to the first.

2.8.3.3 Proof of Proposition 11

Proposition 11 is a particular case in which Proposition 12 holds. The proof relies on the fact that the intensity F is supposed to be $\mathcal{C}^1$ and so we can move from $F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)})$ to $F(0)$ through finite-increments theorem.

Proof. From (2.15) we get the proposition proved remarking that

$$\Delta_{n,i}^{1+2\beta} \int_{\mathbb{R}} u^2 \varphi(u) F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)}) du =$$

$$= \Delta_{n,i}^{1+2\beta} \int_{\mathbb{R}} u^2 \varphi(u) F(0) du + \Delta_{n,i}^{1+2\beta} \int_{\mathbb{R}} u^2 \varphi(u) [F(\frac{u\Delta_{n,i}^\beta}{\gamma(x)}) - F(0)] du$$

and, from finite-increments theorem, the last term is in absolute value upper bounded by
$c\Delta_{n,i}^{1+3\beta} \frac{1}{|\gamma(x)|} \int_{\mathbb{R}} |u|^3 |\varphi(u)||F'(\tilde{u})| du = R(\theta, \Delta_{n,i}^{1+3\beta}, x)$, where $\tilde{u} \in (0, \frac{u\Delta_{n,i}^\beta}{\gamma(x)})$.
Moreover, by the smoothness on F we have required, it follows that terms in (2.15) whose size depends on the density F are now upper bounded by a $R(\theta, \Delta_{n,i}^{2+\beta}, x)$ function. It yields (2.14). □

Part II

Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes

Introduction

In the second part of the thesis we study the estimation of the integrated volatility, which is an important problem in finance. In presence of jumps, a widely used method is to consider the quadratic variation where we remove the contribution of the increments bigger than a threshold, in which we judge that at least a jump occurs.
In the case where the jump part of the stochastic differential equation admits a Blumenthal Getoor index $\alpha > 1$, it has been showed in [47] that the convergence rate deteriorates and some approaches other than the quadratic variation are proposed (see [49]).

In this chapter we consider the stochastic differential equation defined in (1), for $d = 1$, and which jump part is of Stable type with index α.
Using Malliavin calculus we extend Mancini's work [65] by characterizing in detail the bias introduced by the presence of jumps in the truncated quadratic variation. We can therefore correct the original estimator in order to remove such a bias and we prove that the convergence rate not always deteriorates for $\alpha > 1$.
Furthermore, we show numerically that our corrections reduce drastically the bias and that our unbiased truncated quadratic variation performs well also when the index of jump activity α is bigger than 1.
The chapter is based on the paper "Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes" [5], to appear on Stochastic Processes and Their Applications.

Chapter 3

Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes

Abstract :

The problem of integrated volatility estimation for the solution X of a stochastic differential equation with Lévy-type jumps is considered under discrete high-frequency observations in both short and long time horizon.
We provide an asymptotic expansion for the integrated volatility that gives us, in detail, the contribution deriving from the jump part. The knowledge of such a contribution allows us to build an unbiased version of the truncated quadratic variation, in which the bias is visibly reduced.
In earlier results the condition $\beta > \frac{1}{2(2-\alpha)}$ on β (that is such that $(\frac{1}{n})^\beta$ is the threshold of the truncated quadratic variation) and on the degree of jump activity α was needed to have the original truncated realized volatility well-performed (see [65], [45]). In this chapter we theoretically relax this condition and we show that our unbiased estimator achieves excellent numerical results for any couple (α, β).

Keys words : Lévy-driven SDE, integrated variance, threshold estimator, convergence speed, high frequency data.

3.1 Introduction

The class of solutions of Lévy-driven stochastic differential equations has many applications in various area such as neuroscience, physics and finance. Indeed, it includes the stochastic Morris-Lecar neuron model [26] as well as important examples taken from finance such as the Barndorff-Nielsen-Shephard model [8], the Kou model [56] and the Merton model [70]; to name just a few.
In this chapter we aim at estimating the integrated volatility in short and long time based on discrete observations $X_{t_0}, ..., X_{t_n}$; with $t_0 = 0 \leq t_1 \leq ... \leq t_n = T_n$, of the process X given by

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz), \quad t \in \mathbb{R}_+,$$

where $W = (W_t)_{t\geq 0}$ is a one dimensional Brownian motion and $\tilde{\mu}$ is a compensated Poisson random measure, with a possible infinity jump activity.

We consider here the setting of high frequency observations , i.e.
$\Delta_n = \sup_{i=0,...,n-1} \Delta_{n,i} \to 0$ as $n \to \infty$, with $\Delta_{n,i} = (t_{i+1} - t_i)$. Both cases $T_n \in [0, \infty[$ fixed and $\lim_{n\to\infty} T_n = \infty$ are dealt and so we want to estimate, respectively, $IV_1 := \frac{1}{T}\int_0^T a^2(X_s)f(X_s)ds$ and $IV_2 := \int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$, where π is an invariant measure and f a polynomial growth function. If on one side the estimation of IV_2, to our knowledge, has never been considered before, on the other the estimation of IV_1 has been widely studied because of its great importance in finance. Indeed, taking $f \equiv 1$, IV_1 turns out to be the so called integrated volatility that has particular relevance in measuring and forecasting the asset risks; its estimation on the basis of discrete observations of X is one of the long-standing problems.
In the sequel we will present some known results denoting by IV_1 the classical integrated volatility, that is we are assuming that f equals to 1.

When X is continuous, the canonical way for estimating the integrated volatility is to use the realized volatility or approximate quadratic variation at time T:

$$[X, X]_T^n := \sum_{i=0}^{n-1} (\Delta X_i)^2, \qquad \text{where } \Delta X_i = X_{t_{i+1}} - X_{t_i}.$$

Under very weak assumptions on b and a (namely when $\int_0^T b^2(X_s)ds$ and $\int_0^T a^4(X_s)ds$ are finite for all $t \in (0, T]$), we have a central limit theorem (CLT) with rate $\sqrt{n}$: the processes $\sqrt{n}([X, X]_T^n - IV_1)$ converge in the sense of stable convergence in law for processes, to a limit Z which is defined on an extension of the space and which conditionally is a centered Gaussian variable whose conditional law is characterized by its (conditional) variance $V_T := 2\int_0^T a^4(X_s)ds$.

When X has jumps, the variable $[X, X]_T^n$ no longer converges to IV_1. However, there are other known methods to estimate the integrated volatility.
The first type of jump-robust volatility estimators are the *Multipower variations* (cf [9], [10], [46]), which we do not explicitly recall here. These estimators satisfy a CLT with rate $\sqrt{n}$ but with a conditional variance bigger than V_T (so they are rate-efficient but not variance-efficient).
The second type of volatility estimators, introduced by Jacod and Todorov in [49], is based on estimating locally the volatility from the empirical characteristic function of the increments of the process over blocks of decreasing length but containing an increasing number of observations, and then summing the local volatility estimates.

Another method to estimate the integrated volatility in jump diffusion processes, introduced by Mancini in [64], is the use of the *truncated realized volatility* or *truncated quadratic variance* (see [46], [65]):

$$\hat{IV}_T^n := \sum_{i=0}^{n-1} (\Delta X_i)^2 1_{\{|\Delta X_i| \le v_n\}},$$

where v_n is a sequence of positive truncation levels, typically of the form $(\frac{1}{n})^\beta$ for some $\beta \in (0, \frac{1}{2})$.
Below we focus on the estimation of IV_1 through the implementation of the truncated quadratic variation, that is based on the idea of summing only the squared increments of X whose absolute value is smaller than some threshold v_n.
It is shown in [45] that $\hat{IV}_T^n$ has exactly the same limiting properties as $[X,X]_T^n$ does for some $\alpha \in [0,1)$ and $\beta \in [\frac{1}{2(2-\alpha)}, \frac{1}{2})$, where α is the degree of jump activity or Blumenthal-Getoor index, that is the supremum of r for which $\int_{\mathbb{R}} (|z|^r \wedge 1) F(z) dz$ is almost surely finite; F is a Lévy measure which accounts for the jumps of the process and it is such that the compensator $\bar{\mu}$ has the form $\bar{\mu}(dt, dz) = F(z) dz dt$.
Mancini has proved in [65] that, when the jumps of X are those of a stable process with index $\alpha \ge 1$, the truncated quadratic variation is such that

$$(\hat{IV}_T^n - IV_1) \overset{\mathbb{P}}{\sim} (\frac{1}{n})^{\beta(2-\alpha)}. \tag{3.1}$$

This rate is less than $\sqrt{n}$ and no proper CLT is available in this case.

In this chapter, in order to estimate $IV_1 := \frac{1}{T}\int_0^T a^2(X_s) f(X_s) ds$ and $IV_2 := \int_{\mathbb{R}} a^2(x) f(x) \pi(dx)$, we consider in particular the truncated quadratic variation defined in the following way:

$$Q_n := \frac{1}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} (X_{t_{i+1}} - X_{t_i})^2 \varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}),$$

where φ is a C^∞ function that vanishes when the increments of the data are too large compared to the typical increments of a continuous diffusion process, and thus can be used to filter the contribution of the jumps.
We aim to extend the results proved in short time in [65] characterising precisely the noise introduced by the presence of jumps in both short and long time and finding consequently some corrections to reduce such a noise.
The main result of this chapter is the asymptotic expansion for the integrated volatility in short and long time. Compared to earlier results, which exists only in short time case, our asymptotic expansion provides us precisely the limit to which $\frac{1}{\Delta_n^{\beta(2-\alpha)}}(Q_n - IV_1)$ converges when $\Delta_n^{\beta(2-\alpha)} > \sqrt{n}$, that in case of uniform discretization steps (for which $\Delta_n = \frac{T}{n}$) matches with the condition $\beta < \frac{1}{2(2-\alpha)}$.
In the case where the discretization step is uniform our work extends [65]. Indeed, we find

$$Q_n - IV_1 = \frac{Z_n}{\sqrt{n}} + (\frac{1}{n})^{\beta(2-\alpha)} c_\alpha \int_{\mathbb{R}} \varphi(u) |u|^{1-\alpha} du \int_0^T |\gamma|^\alpha (X_s) f(X_s) ds + o_{\mathbb{P}}((\frac{1}{n})^{\beta(2-\alpha)}),$$

where $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_0^T a^4(X_s) f^2(X_s) ds)$ stably with respect to X. In Theorem 12 and 13 below the result is extended to non uniform sampling step as well. The

asymptotic expansion here above allows us to deduce the behaviour of the truncated quadratic variation for each couple (α, β), that is a plus compared to (3.1).
Furthermore, providing we know α (and if we don't it is enough to estimate it previously, see for example [91]), we can improve the performance of the truncated quadratic variation subtracting the noise due to the presence of jumps to the original estimator or taking particular functions φ that make the bias derived from the jump part equal to zero. Using the asymptotic expansion of the integrated volatility we also provide the rate of the error left after having applied the corrections.
Moreover, in the case where the volatility is constant, we show numerically that the corrections gained by the knowledge of the asymptotic expansion for the integrated volatility in short time allows us to reduce visibly the noise for any $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0, 2)$. It is a clear improvement because, if the original truncated quadratic variation was a well-performed estimator only if $\beta > \frac{1}{2(2-\alpha)}$ (condition that never holds for $\alpha \geq 1$), the unbiased truncated quadratic variation achieves excellent results for any couple (α, β).

The outline of the chapter is the following. In Section 2 we present the assumptions on the process X. In Section 3.1 we define the truncated quadratic variation, while Section 3.2 contains the main results of the chapter. In Section 4 we show the numerical performance of the unbiased estimator. The Section 5 is devoted to the statement of propositions useful for the proof of the main results, that is given in Section 6. In Section 7 we give some technical tools about Malliavin calculus, required for the proof of some propositions, while other proofs and some technical results are presented in the Appendix.

3.2 Model, assumptions

Let X be a solution to

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz), \quad t \in \mathbb{R}_+, \tag{3.2}$$

where $W = (W_t)_{t\geq 0}$ is a one dimensional Brownian motion and μ is a Poisson random measure on $[0, \infty) \times \mathbb{R}$ associated to the Lévy process $L = (L_t)_{t\geq 0}$, with $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$. The compensated measure is $\tilde{\mu} = \mu - \bar{\mu}$; we suppose that the compensator has the following form: $\bar{\mu}(dt, dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.
We denote $(\Omega, \mathcal{F}, \mathbb{P})$ the probability space on which W and μ are defined. The initial condition X_0, W and L are independent.

3.2.1 Assumptions

We suppose that the functions $b : \mathbb{R} \to \mathbb{R}$, $a : \mathbb{R} \to \mathbb{R}$ and $\gamma : \mathbb{R} \to \mathbb{R}$ satisfy the following assumptions:

ASSUMPTION 1: *The functions $b(x)$, $\gamma(x)$ and $a(x)$ are globally Lipschitz.*

Under Assumption 1 the equation (4.3) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [7] (Theorems 6.2.9. and 6.4.6.).

ASSUMPTION 2: *There exists a constant* $t > 0$ *such that* X_t *admits a density* $p_t(x, y)$ *with respect to the Lebesgue measure on* $\mathbb{R}$*; bounded in* $y \in \mathbb{R}$ *and in* $x \in K$ *for every compact* $K \subset \mathbb{R}$*. Moreover, for every* $x \in \mathbb{R}$ *and every open ball* $U \in \mathbb{R}$*, there exists a point* $z = z(x, U) \in supp(F)$ *such that* $\gamma(x)z \in U$*.*

The last assumption was used in [67] to prove the irreducibility of the process X. Other sets of conditions, sufficient for irreducibility, can be found in the same source.

ASSUMPTION 3 (Ergodicity):

1. *For all* $q > 0$*,* $\int_{|z|>1} |z|^q F(z)dz < \infty$*.*
2. *There exists* $C > 0$ *such that* $xb(x) \leq -C|x|^2$*, if* $|x| \to \infty$*.*
3. $|a(x)|/|x| \to 0$ *as* $|x| \to \infty$*.*
4. $\forall q > 0$ *we have* $\mathbb{E}|X_0|^q < \infty$*.*

The points 2 - 3 and 4 of the Assumption 3 here above are required only in the case of long time observation.
Assumption 2 ensures, together with the Assumption 3 and the fifth point of Assumption 4 below, the existence of unique invariant distribution π, as well as the ergodicity of the process X, as stated in the Lemma 26 below.

ASSUMPTION 4 (Jumps):

1. *The jump coefficient* γ *is bounded from below, that is*
$$\inf_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{min} > 0$$
2. *The Lévy measure* F *is absolutely continuous with respect to the Lebesgue measure and we denote* $F(z) = \frac{F(dz)}{dz}$*.*
3. *The Lévy measure* F *satisfies* $F(dz) = \frac{g(z)}{|z|^{1+\alpha}}dz$*, where* $\alpha \in (0, 2)$ *and* $g : \mathbb{R} \to \mathbb{R}$ *is a continuous symmetric nonnegative bounded function with* $g(0) = 1$*.*
4. *The function* g *is differentiable on* $\{0 < |z| \leq \eta\}$ *for some* $\eta > 0$ *with continuous derivative such that* $\sup_{0<|z|\leq\eta} |\frac{g'}{g}| < \infty$*.*
5. *The jump coefficient* γ *is upper bounded, i.e.* $\sup_{x\in\mathbb{R}} |\gamma(x)| := \gamma_{max} < \infty$*.*

Assumptions 4.1 and 4.5 are useful to compare size of jumps of X and L. Assumption 4.4 is satisfied by a large class of processes: α- stable process ($g = 1$), truncated α-stable processes ($g = \tau$, a truncation function), tempered stable process ($g(z) = e^{-\lambda|z|}$, $\lambda > 0$).

We will use some moment inequalities for jump diffusions, gathered in the following lemma:

Lemma 25. *Let X satisfies Assumptions 1-4. Let $L_t := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}(ds, dz)$ and let $\mathcal{F}_s := \sigma\{(W_u)_{0<u\leq s}, (L_u)_{0<u\leq s}, X_0\}$.*
Then, for all $t > s$,
1) for all $p \geq 2$, $\mathbb{E}[|X_t - X_s|^p]^{\frac{1}{p}} \leq c|t-s|^{\frac{1}{p}}$,
2) for all $p \geq 2$, $p \in \mathbb{N}$, $\mathbb{E}[|X_t - X_s|^p|\mathcal{F}_s] \leq c|t-s|(1+|X_s|^p)$.
3) for all $p \geq 2$, $p \in \mathbb{N}$, $\sup_{h\in[0,1]} \mathbb{E}[|X_{s+h}|^p|\mathcal{F}_s] \leq c(1+|X_s|^p)$.
4) for all $p > 1$, $\mathbb{E}[|X_t^c - X_s^c|^p]^{\frac{1}{p}} \leq |t-s|^{\frac{1}{2}}$ and $\mathbb{E}[|X_t^c - X_s^c|^p|\mathcal{F}_s]^{\frac{1}{p}} \leq c|t-s|^{\frac{1}{2}}(1+|X_s|^p)$, where we have denoted by X^c the continuous part of the process X.

The first two points follow from Theorem 66 of [80] and Proposition 3.1 in [88]. The third point is showed in Chapter 1, below Lemma 1, and the last one in Section 8 of [38].

The following Lemma states that Assumptions $1-4$ are sufficient for the existence of an invariant measure π such that an ergodic theorem holds and moments of all order exist.

Lemma 26. *Under assumptions 1 to 4, the process X admits a unique invariant distribution π and the ergodic theorem holds:*

1. *For every measurable function $h : \mathbb{R} \to \mathbb{R}$ satisfying $\pi(h) < \infty$, we have a.s.*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t h(X_s)ds = \pi(h).$$

2. *For all $q > 0$, $\pi(|x|^q) < \infty$.*

3. *For all $q > 0$, $\sup_{t\geq 0} \mathbb{E}[|X_t|^q] < \infty$.*

A proof is in [38] (Section 8 of Supplement) in the case $\alpha \in (0,1)$ and the proof relies on [67]. The case $\alpha \in (0,2)$ is studied in Chapter 1.

3.3 Setting and main results

Let X be the solution to (4.3). Suppose that we observe a finite sample

$$X_{t_0}, ..., X_{t_n}; \qquad 0 = t_0 \leq t_1 \leq ... \leq t_n = T.$$

Every observation time point depends also on n, but to simplify the notation we suppress this index. We will be working in a high-frequency setting, i.e.

$$\Delta_n := \sup_{i=0,\dots,n-1} \Delta_{n,i} \longrightarrow 0, \quad n \to \infty,$$

with $\Delta_{n,i} := (t_{i+1} - t_i)$.
We study both the cases $T \in \mathbb{R}$ fixed and $\lim_{n\to\infty} T = \infty$.

We denote by IV_1 the quantity $\frac{1}{T}\int_0^T a^2(X_s)f(X_s)ds$ and by IV_2 $\int_{\mathbb{R}} a^2(x)f(x)\pi(dx)$, where π is the invariant measure introduced in Lemma 26 and f a polynomial growth function.
In order to estimate IV_1 and IV_2 we introduce Q_n, based on the idea of summing

only some of the squared increments of X, those whose absolute value is smaller than $2\Delta_{n,i}^{\beta}$, with $\beta \in (0, \frac{1}{2})$. Indeed, we set

$$Q_n := \frac{1}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} (X_{t_{i+1}} - X_{t_i})^2 \varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}), \tag{3.3}$$

where

$$\varphi_{\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) = \varphi(\frac{X_{t_{i+1}} - X_{t_i}}{\Delta_{n,i}^{\beta}}),$$

with φ a smooth version of the indicator function, such that $\varphi(\zeta) = 0$ for each ζ, with $|\zeta| \geq 2$ and $\varphi(\zeta) = 1$ for each ζ, with $|\zeta| \leq 1$.
It is worth noting that, if we consider an additional constant k in φ (that becomes $\varphi_{k\Delta_{n,i}^{\beta}}(X_{t_{i+1}} - X_{t_i}) = \varphi(\frac{X_{t_{i+1}} - X_{t_i}}{k\Delta_{n,i}^{\beta}})$), the only difference is the interval on which the function is 1 or 0: it will be 1 for $|X_{t_{i+1}} - X_{t_i}| \leq k\Delta_{n,i}^{\beta}$; 0 for $|X_{t_{i+1}} - X_{t_i}| \geq 2k\Delta_{n,i}^{\beta}$. Hence, for shortness in notations, we restrict the theoretical analysis to the situation where $k = 1$ while, for applications, we may take the threshold level as $k\Delta_{n,i}^{\beta}$ with $k \neq 1$.

3.3.1 Conditions on the step discretization

In this paragraph we introduce all the assumptions on the step discretization that we will need and we will use, a little at a time, in the proofs of the main results. We consider both the cases T fixed and $\lim_{n\to\infty} T = \infty$.

ASSUMPTION S1 (Step Discretization, T fixed):

1. There exists a measurable function $s \mapsto H(s, 0)$ such that for all function h continuous and bounded,

$$\eta_{sampling}^{(n)}(h) = \frac{1}{n} \sum_{i=0}^{n-1} h(X_{t_i}) \to \eta(h) = \int_0^T h(X_s) H(s, 0) ds.$$

2. For $\delta \in [0, 1)$, there exists a measurable function $s \mapsto H(s, \delta)$ such that, for every continuous function $h : \mathbb{R} \to \mathbb{R}$,

$$\frac{1}{\Delta_n^{\delta}} \frac{1}{n} \sum_{i=0}^{n-1} h(X_{t_i}) \Delta_{n,i}^{\delta} \to \int_0^T h(X_s) H(s, \delta) ds \tag{3.4}$$

3. $\exists \delta_0 > 0 : |\frac{1}{n} - \frac{\Delta_{n,i}}{\sum_{i=0}^{n-1} \Delta_{n,i}}| \leq \frac{\Delta_n^{\beta(2-\alpha)+\delta_0}}{n}$, $\forall i \in \{0, ..., n-1\}$, for $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0, 2)$.

We observe that, for $\delta = 0$, the point 2 fall back into point 1. It is therefore a condition stronger than the first one, but it is not always required. Conditions on the sampling step analogous to those stated in first and second points are introduced in Section 2.6 of Mykland and Zhang [75], related to the existence of quadratic variation in time (see also Example 2.24 in [52]).
We remark that, considering a uniform steps discretization, the three conditions here above clearly hold.

ASSUMPTION S2 (Step Discretization, $T \to \infty$):

1. $\exists\, c_1, c_2 > 0$ such that $c_1 \leq \frac{\max_{i\in\{0,\dots,n-1\}} \Delta_{n,i}}{\min_{i\in\{0,\dots,n-1\}} \Delta_{n,i}} \leq c_2$.

2. For $\delta \in [0,1)$ there exists c_δ such that $\forall n$, $(\min_{i\in\{0,\dots,n-1\}} \Delta_{n,i})^{1-\delta} \sum_{j=1}^{n} |\frac{1}{\Delta_{n,j}^{1-\delta}} - \frac{1}{\Delta_{n,j-1}^{1-\delta}}| < c_\delta$.

3. $\exists \delta_0 > 0 : |\frac{1}{n} - \frac{\Delta_{n,i}}{\sum_{i=0}^{n-1}\Delta_{n,i}}| \leq \frac{\Delta_n^{\beta(2-\alpha)+\delta_0}}{n}$, $\forall i \in \{0,...,n-1\}$, for $\beta \in (0,\frac{1}{2})$ and $\alpha \in (0,2)$.

Again, if we consider a uniform discretization, the three conditions here above hold. The second point is an assumption of regularity on the function $j \mapsto \Delta_{n,j}^{\delta-1}$. It comes naturally from the proof of the lemma below.
We observe that, when $T \to \infty$, it doesn't make sens to add a condition as (3.4) because its left hand side always converges to the same quantity for all $\delta \in [0,1)$, as consequence of the following lemma, that we will prove in the appendix:

Lemma 27. *Suppose that Assumptions 1 to 4 and the points 1 and 2 of S2 hold. Then, for every measurable function $h : \mathbb{R} \to \mathbb{R}$ with bounded derivative such that $\pi(h) < \infty$ and for $\delta \in [0,1)$ we have the following convergence in probability:*

$$\frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}^\delta}\sum_{i=0}^{n-1}\Delta_{n,i}^\delta h(X_{t_i}) \xrightarrow{\mathbb{P}} \int_{\mathbb{R}} h(x)\pi(dx). \tag{3.5}$$

3.3.2 Main results

3.3.2.1 Decomposition of the truncated quadratic variation

In this section we enunciate theorems that explain the asymptotic behavior of Q_n. First of all we define

$$\tilde{Q}_n := \frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}}\gamma(X_{s^-})\, z\, \tilde{\mu}(ds,dz))^2\frac{f(X_{t_i})}{\Delta_{n,i}}\varphi_{\Delta_{n,i}^\beta}(X_{t_{i+1}}-X_{t_i}). \tag{3.6}$$

in order to write the decomposition of the truncated quadratic variation into two parts: the continuous quadratic variation and the quadratic variation deriving from jumps.

Theorem 10. *Suppose that Assumptions 1 to 4 hold and that $\beta \in (0,\frac{1}{2})$ and $\alpha \in (0,2)$ are given in definition* (3.3) *and in the third point of Assumption 4, respectively. Then, as $\Delta_n \to 0$,*

$$Q_n = \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(X_{t_{i+1}}^c - X_{t_i}^c)^2 + \Delta_n^{\beta(2-\alpha)}\tilde{Q}_n + \mathcal{E}_n = \tag{3.7}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}}a(X_s)dW_s)^2 + \Delta_n^{\beta(2-\alpha)}\tilde{Q}_n + \mathcal{E}_n, \tag{3.8}$$

where X_s^c is the continuous part of the process X_s, $\mathcal{E}_n$ is both $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ and, for each $\tilde{\epsilon} > 0$, $o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$; with $o_{\mathbb{P}}(\Delta_n^k)$ such that $\frac{o_{\mathbb{P}}(\Delta_n^k)}{\Delta_n^k} \xrightarrow{\mathbb{P}} 0$.

We now consider the difference between the truncated quadratic variation and the discretized volatility and we make explicit its decomposition into the statistical error and the noise term due to the jumps. To do that, we introduce

$$\hat{Q}_n := \frac{1}{n\Delta_n^{\beta(2-\alpha)}} \sum_{i=0}^{n-1} f(X_{t_i})\gamma^2(X_{t_i})\Delta_{n,i}^{\frac{2}{\alpha}-1} d(\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}-\beta}), \tag{3.9}$$

where $d(\zeta) := \mathbb{E}[(S_1^\alpha)^2\varphi(S_1^\alpha\zeta)]$; $(S_t^\alpha)_{t\geq 0}$ is an α-stable process.

Theorem 11. *Suppose that Assumptions 1 to 4 hold and that $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0,2)$ are given in Definition 3.3 and in the third point of Assumption 4, respectively. Then, as $\Delta_n \to 0$,*

$$Q_n - \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) = \frac{Z_n}{\sqrt{n}} + \Delta_n^{\beta(2-\alpha)}\hat{Q}_n + \mathcal{E}_n, \tag{3.10}$$

where $\mathcal{E}_n$ is always $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ and, adding the condition $\beta > \frac{1}{4-\alpha}$, it is also $o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$.
Moreover,

1. *If T is fixed we suppose moreover that point 1 of Assumption S1 holds, then Z_n here above is such that $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_0^T a^4(X_s)f^2(X_s)H(s,0)ds)$ stably with respect to X.*

2. *If otherwise we are in the case $\lim_{n\to\infty} T = \infty$, we suppose that points 1 and 2 of Assumption S2 hold. In this case $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_{\mathbb{R}} a^4(x)f^2(x)\pi(dx))$.*

We recognize in the expansion (3.10) the statistical error of model without jumps given by Z_n, whose variance is equal to the so called quadricity. The term $\hat{Q}_n$ is a bias term arising from the presence of jumps and given by (3.9). From this explicit expression it is possible to remove the bias term (see Section 3.4).
The term $\mathcal{E}_n$ is an additional error term that is always negligible compared to the bias deriving from the jump part $\Delta_n^{\beta(2-\alpha)}\hat{Q}_n$ (that is of order $\Delta_n^{\beta(2-\alpha)}$ by Theorems 12 and 13 below). It also gives us an upper bound to the order of the error we get after having removed the bias. In particular, if $\alpha\beta$ is small enough (that is $\alpha\beta < \frac{1}{2}$), we get that the error term $\mathcal{E}_n$ is $o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}})$ and so it is upper bounded by a term whose order is roughly the same as the statistical error's one.
The bias term admits a first order expansion that does not require the knowledge of the density of S^α.

Proposition 20. *Suppose that Assumptions 1 to 4 hold and that $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0,2)$ are given in Definition 3.3 and in the third point of Assumption 4, respectively. Then*

$$\hat{Q}_n = \frac{1}{n\Delta_n^{\beta(2-\alpha)}} c_\alpha \sum_{i=0}^{n-1} f(X_{t_i})|\gamma|^\alpha(X_{t_i})\Delta_{n,i}^{\beta(2-\alpha)}(\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du) + \tilde{\mathcal{E}}_n, \tag{3.11}$$

with

$$c_\alpha = \begin{cases} \frac{\alpha(1-\alpha)}{4\Gamma(2-\alpha)\cos(\frac{\alpha\pi}{2})} & \text{if } \alpha \neq 1,\ \alpha < 2 \\ \frac{1}{2\pi} & \text{if } \alpha = 1. \end{cases} \tag{3.12}$$

$\tilde{\mathcal{E}}_n = o_{\mathbb{P}}(1)$ *and, if* $\alpha < \frac{4}{3}$*, it is also*

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}} o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})}) = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon})\wedge(1-2\beta-\tilde{\epsilon})}).$$

We underline that we have not replaced directly the right hand side of (3.11) in (3.10), observing that $\Delta_n^{\beta(2-\alpha)}\tilde{\mathcal{E}}_n = \mathcal{E}_n$, because $\Delta_n^{\beta(2-\alpha)}\tilde{\mathcal{E}}_n$ is always $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ but to get it is also $o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$ the additional condition $\alpha < \frac{4}{3}$ is required.
In the case $\alpha < \frac{4}{3}$ we get the following corollary:

Corollary 1. *Suppose that Assumptions 1 to 4 and point 1 of Assumption S1 (or points 1 and 2 of Assumption S2, if* $\lim_{n\to\infty} T = \infty$*) hold and that* $\alpha \in (0, \frac{4}{3})$*,* $\beta \in (\frac{1}{4-\alpha}, (\frac{1}{2\alpha} \wedge \frac{1}{2}))$*. If* φ *is such that* $\int_{\mathbb{R}} |u|^{1-\alpha}\varphi(u)du = 0$ *then,* $\forall\tilde{\epsilon} > 0$*,*

$$Q_n - \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) = \frac{Z_n}{\sqrt{n}} + o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}), \tag{3.13}$$

with Z_n *defined as in Theorem 11 here above.*

It is always possible to build a function φ for which the condition here above is respected (see Section 3.4).
We observe that, if $\alpha \geq \frac{4}{3}$ but $\gamma = k \in \mathbb{R}$, the result still holds choosing φ such that $\int_{\mathbb{R}} u^2\varphi(u)\, f_\alpha(\frac{1}{k}u\Delta_{n,i}^{\beta-\frac{1}{\alpha}})du$ is equal to 0, where f_α is the density of the α-stable process. Indeed, starting from (3.10), we have that $\hat{Q}_n$ is now zero: by its definition (3.9) it is equal to $\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1} f(X_{t_i})k^\alpha\Delta_{n,i}^{\frac{2}{\alpha}-1}\int_{\mathbb{R}} z^2\varphi(zk\Delta_{n,i}^{\frac{1}{\alpha}-\beta})f_\alpha(z)dz =$ $= \frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1} f(X_{t_i})k^{\alpha-3}\Delta_{n,i}^{3\beta-\frac{1}{\alpha}-1}\int_{\mathbb{R}} u^2\varphi(u)f_\alpha(\frac{1}{k}u\Delta_{n,i}^{\beta-\frac{1}{\alpha}})du = 0$, where we have used a change of variable.
Equation (3.13) gives us the behaviour of the unbiased estimator, that is the truncated quadratic variation after having removed the noise derived from the presence of jumps. Taking α and β as in the corollary here above we also have reduced the error term $\mathcal{E}_n$ to be $o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}})$, which means that after having applied the corrections we get an error that is upper bounded by a term whose order is, in the case of finite time horizon, roughly the same as the statistical error's one.

3.3.2.2 Asymptotic expansion for the integrated volatility in short and long time

The limits of $\hat{Q}_n$ are given below in both cases T fixed and $T \to \infty$.
When T is fixed we have the following result:

Theorem 12. *Suppose that Assumptions 1 2, 4 and points 1 and 5 of Assumption 3 hold. Moreover we suppose that* T *is fixed and that points 1 and 2 of Assumption S1 hold. Then, as* $\Delta_n \to 0$*,*

$$\hat{Q}_n \xrightarrow{\mathbb{P}} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha}du \int_0^T |\gamma(X_s)|^\alpha f(X_s)H(s,\beta(2-\alpha))ds. \tag{3.14}$$

Moreover, if we add the third point of Assumption S1, we have

$$Q_n - IV_1 = \tag{3.15}$$

$$= \frac{Z_n}{\sqrt{n}} + \Delta_n^{\beta(2-\alpha)} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du \int_0^T |\gamma(X_s)|^\alpha f(X_s) H(s, \beta(2-\alpha)) ds + o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)}),$$

where $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_0^T a^4(X_s) f^2(X_s) H(s,0) ds)$ *stably with respect to* X.

It is worth noting that, in both [47] and [65], the integrated volatility estimation in short time is dealt and they show that the truncated quadratic variation has rate $\sqrt{n}$ if $\beta > \frac{1}{2(2-\alpha)}$.
We remark that the jump part is negligible compared to the statistic error if $\Delta_n^{\beta(2-\alpha)} < n^{-\frac{1}{2}}$, it follows the condition $\Delta_n < n^{-\frac{1}{2\beta(2-\alpha)}}$ on the discretization step. If we use, in particular, an uniform step discretization such that $\forall i \in \{0, ..., n-1\}$ $\Delta_{n,i} = \Delta_n = \frac{1}{n}$, then the condition becomes $n^{-1} < n^{-\frac{1}{2\beta(2-\alpha)}}$ and so $\beta > \frac{1}{2(2-\alpha)}$, that is the same condition given in [47].
However, if we take (α, β) for which such a condition doesn't hold, we can still use that we know in detail the noise deriving from jumps to implement corrections that still make the unbiased estimator well-performed (see Section 3.4).

We also study the asymptotic expansion for the integrated volatility in long time that, to our knowledge, hasn't never been dealt before. We have the following result:

Theorem 13. *Suppose that Assumptions 1 to 4 and points 1 and 2 of Assumption S2 hold. We assume moreover that* $\lim_{n\to\infty} T = \infty$ *and* $n\Delta_n = O(T)$. *Then, as* $\Delta_n \to 0$,

$$\hat{Q}_n \xrightarrow{\mathbb{P}} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du \int_{\mathbb{R}} |\gamma(x)|^\alpha f(x) \pi(dx). \tag{3.16}$$

Moreover, if we add the third condition of Assumption S2 we have

$$Q_n - \frac{1}{T}\int_0^T f(X_s) a^2(X_s) ds = \tag{3.17}$$

$$= \frac{Z_n}{\sqrt{n}} + \Delta_n^{\beta(2-\alpha)} c_\alpha \int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du \int_{\mathbb{R}} |\gamma(x)|^\alpha f(x) \pi(dx) + o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)}),$$

where $Z_n \xrightarrow{\mathcal{L}} N(0, 2\int_{\mathbb{R}} a^4(x) f^2(x) \pi(dx))$.

Because of the ergodic theorem, $\frac{1}{T}\int_0^T f(X_s) a^2(X_s) ds$ converges to IV_2, but slowly (with rate $\sqrt{T}$). Anyway for the applications the convergence to IV_2 is not required.

We observe that, if we take a discretization step that is $\Delta_n = n^{-\rho}$, with $\rho \in (0,1)$, the jump part is negligible compared to the statistical error if $n^{-\rho\beta(2-\alpha)} < n^{-\frac{1}{2}}$ and so if $\rho > \frac{1}{2\beta(2-\alpha)}$. Since β is always less than $\frac{1}{2}$ it means that ρ must be more than $\frac{1}{2-\alpha}$ or, equivalently, $\alpha < 2 - \frac{1}{\rho}$.
It is worth noting that smaller is ρ and less choice we have on α. In particular for $\rho < \frac{1}{2}$ there is no α for which the condition here above holds. On the other side, for ρ close to 1, we fall back on the condition $\alpha < 1$.

3.4 Unbiased estimation in the case of constant volatility

In this section we consider a concrete application of the unbiased volatility estimator in a jump diffusion model and we investigate its numerical performance.

We consider our model (4.3) in which we assume, in addition, that the functions a and γ are both constants.
Suppose that we are given a discrete sample $X_{t_0}, ..., X_{t_n}$ with $t_i = i\Delta_n = \frac{i}{n}$ for $i = 0, ..., n$. We remark that, with such a discretization step, all the points of Assumption S1 and S2 clearly hold.
We now want to analyze the estimation improvement; to do it we compare the classical error committed using the truncated quadratic variation with the unbiased estimation derived by our main results.
We define the estimator we are going to use, in which we have clearly taken $f \equiv 1$ and we have introduced a threshold k in the function φ, so it is

$$Q_n = \frac{1}{n}\sum_{i=0}^{n-1} \frac{(X_{t_{i+1}} - X_{t_i})^2}{\Delta_{n,i}} \varphi_{k\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}) = \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 \varphi_{k\Delta_{n,i}^\beta}(X_{t_{i+1}} - X_{t_i}). \tag{3.18}$$

If normalized, the error committed estimating the volatility is $E_1 := (Q_n - \sigma^2)\sqrt{n}$.
We start from (3.11) that in our case, taking into account the presence of k, is

$$\hat{Q}_n = c_\alpha \gamma^\alpha k^{2-\alpha} (\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du) + \tilde{\mathcal{E}}_n. \tag{3.19}$$

We now get different methods to make the error smaller.
First of all we can replace (3.19) in (3.10) and so we can reduce the error by subtracting a correction term, building the new estimator

$$Q_n^c := Q_n - \Delta_n^{\beta(2-\alpha)} c_\alpha \gamma^\alpha k^{2-\alpha} (\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du).$$

The error committed estimating the volatility with such a corrected estimator is $E_2 := (Q_n^c - \sigma^2)\sqrt{n}$.
Another approach consists of taking a particular function $\tilde{\varphi}$ that makes the main contribution of $\hat{Q}_n$ equal to 0. We define $\tilde{\varphi}(\zeta) = \varphi(\zeta) + c\psi(\zeta)$, with ψ a $\mathcal{C}^\infty$ function such that $\psi(\zeta) = 0$ for each ζ, $|\zeta| \geq 2$ or $|\zeta| \leq 1$. In this way, for any $c \in \mathbb{R}\backslash\{0\}$, $\tilde{\varphi}$ is still a smooth version of the indicator function such that $\tilde{\varphi}(\zeta) = 0$ for each ζ, $|\zeta| \geq 2$ and $\tilde{\varphi}(\zeta) = 1$ for each ζ, $|\zeta| \leq 1$. We can therefore leverage the arbitrariness in c to make the main contribution of $\hat{Q}_n$ equal to zero, choosing $\tilde{c} := -\frac{\int_{\mathbb{R}} \varphi(u)|u|^{1-\alpha} du}{\int_{\mathbb{R}} \psi(u)|u|^{1-\alpha} du}$, which is such that $\int_{\mathbb{R}} (\varphi + \tilde{c}\psi(u))|u|^{1-\alpha} du = 0$.
Hence, it is possible to achieve an improved estimation of the volatility by used the truncated quadratic variation $Q_{n,c} := \sum_{i=0}^{n-1} (X_{t_{i+1}} - X_{t_i})^2 (\varphi + \tilde{c}\psi)(\frac{X_{t_{i+1}} - X_{t_i}}{k\Delta_{n,i}^\beta})$. To make it clear we will analyze the quantity $E_3 := (Q_{n,c} - \sigma^2)\sqrt{n}$.
Another method widely used in numerical analysis to improve the rate of convergence of a sequence is the so-called Richardson extrapolation. We observe that the first term on the right hand side of (3.19) does not depend on n and so we can just write $\hat{Q}_n = \hat{Q} + \tilde{\mathcal{E}}_n$. Replacing it in (3.10) we get

$$Q_n = \sigma^2 + \frac{Z_n}{\sqrt{n}} + \frac{1}{n^{\beta(2-\alpha)}}\hat{Q} + \mathcal{E}_n \qquad \text{and}$$

$$Q_{2n} = \sigma^2 + \frac{Z_{2n}}{\sqrt{2n}} + \frac{1}{2^{\beta(2-\alpha)}} \frac{1}{n^{\beta(2-\alpha)}}\hat{Q} + \mathcal{E}_{2n},$$

where we have also used that $\Delta_n^{\beta(2-\alpha)} \tilde{\mathcal{E}}_n = \mathcal{E}_n$. We can therefore use $\frac{Q_n - 2^{\beta(2-\alpha)} Q_{2n}}{1 - 2^{\beta(2-\alpha)}}$ as improved estimator of σ^2.

We give simulation results for E_1, E_2 and E_3 in the situation where $\sigma = 1$. The given mean and the deviation standard are each based on 500 Monte Carlo samples. We choose to simulate a tempered stable process (that is F satisfies $F(dz) = \frac{e^{-|z|}}{|z|^{1+\alpha}}$) in the case $\alpha < 1$ while, in the interest of computational efficiency, we will exhibit results gained from the simulation of a stable Lévy process in the case $\alpha \geq 1$ ($F(dz) = \frac{1}{|z|^{1+\alpha}}$).
We have taken the smooth functions φ and ψ as below:

$$\varphi(x) = \begin{cases} 1 & \text{if } |x| < 1 \\ e^{\frac{1}{3} + \frac{1}{|x|^2 - 4}} & \text{if } 1 \leq |x| < 2 \\ 0 & \text{if } |x| \geq 2 \end{cases} \tag{3.20}$$

$$\psi_M(x) = \begin{cases} 0 & \text{if } |x| \leq 1 \text{ or } |x| \geq M \\ e^{\frac{1}{3} + \frac{1}{|3-x|^2 - 4}} & \text{if } 1 < |x| \leq \frac{3}{2} \\ e^{\frac{1}{|x|^2 - M} - \frac{5}{21} + \frac{4}{4M^2 - 9}} & \text{if } \frac{3}{2} < |x| < M; \end{cases} \tag{3.21}$$

choosing opportunely the constant M in the definition of ψ_M we can make its decay slower or faster. We observe that the theoretical results still hold even if the support of $\tilde{\varphi}$ changes as M changes and so it is $[-M, M]$ instead of [-2, 2].
Concerning the constant k in the definition of φ, we fix it equal to 3 in the simulation of the tempered stable process, while its value is 2 in the case $\alpha > 1$, $\beta = 0.2$ and, in the case $\alpha > 1$ and $\beta = 0.49$, it increases as α and γ increase.
The results of the simulations are given in columns 3-6 of Table 3.1 for $\beta = 0.2$ and in columns 3-6 of Table 3.2 for $\beta = 0.49$.

α	γ	**Mean** E_1	**Rms** E_1	**Mean** E_2	**Mean** E_3
0.1	1	3.820	3.177	0.831	0.189
	3	5.289	3.388	1.953	-0.013
0.5	1	15.168	9.411	0.955	1.706
	3	14.445	5.726	2.971	0.080
0.9	1	13.717	4.573	4.597	0.311
	3	42.419	6.980	13.664	-0.711
1.2	1	32.507	11.573	0.069	2.137
	3	112.648	21.279	-0.915	0.800
1.5	1	50.305	12.680	0.195	0.923
	3	250.832	27.170	-5.749	3.557
1.9	1	261.066	20.729	-0.530	9.139
	3	2311.521	155.950	-0.304	-35.177

Table 3.1 – Monte Carlo estimates of E_1, E_2 and E_3 from 500 samples. We have here fixed $n = 700$ and $\beta = 0.2$.

It appears that the estimation we get using the truncated quadratic variation performs worse as soon as α and γ become bigger (see column 3 in both Tables 3.1 and 3.2). However, after having applied the corrections, the error seems visibly reduced. A proof of which lies, for example, in the comparison between the error and the root mean square: before the adjustment in both Tables 3.1 and 3.2 the third column

α	γ	**Mean** E_1	**Rms** E_1	**Mean** E_2	**Mean** E_3
0.1	1	1.092	1.535	0.307	-0.402
	3	1.254	1.627	0.378	-0.372
0.5	1	2.503	1.690	0.754	-0.753
	3	4.680	2.146	1.651	-0.824
0.9	1	2.909	1.548	0.217	0.416
	3	8.042	1.767	0.620	-0.404
1.2	1	7.649	1.992	-0.944	-0.185
	3	64.937	9.918	-1.692	-2.275
1.5	1	25.713	3.653	-1.697	3.653
	3	218.591	21.871	-4.566	-13.027
1.9	1	238.379	14.860	-6.826	16.330
	3	2357.553	189.231	3.827	-87.353

Table 3.2 – Monte Carlo estimates of E_1, E_2 and E_3 from 500 samples. We have here fixed $n = 700$ and $\beta = 0.49$.

dominates the fourth one, showing that the bias of the original estimator dominates the standard deviation while, after the implementation of our main results, we get E_2 and E_3 for which the bias is much smaller.
We observe that for $\alpha < 1$, in both cases $\beta = 0.2$ and $\beta = 0.49$, it is possible to choose opportunely M (on which ψ's decay depends) to make the error E_3 smaller than E_2. On the other hand, for $\alpha > 1$, the approach who consists of subtracting the jump part to the error results better than the other, since E_3 is in this case generally bigger than E_2, but to use this method the knowledge of γ is required. It is worth noting that both the approaches used, that lead us respectively to E_2 and E_3, work well for any $\beta \in (0, \frac{1}{2})$.
We recall that, in [47], the condition found on β to get a well-performed estimator was

$$\beta > \frac{1}{2(2-\alpha)}, \tag{3.22}$$

that is not respected in the case $\beta = 0.2$. Our results match the ones in [47], since the third column in Table 3.2 (where $\beta = 0.49$) is generally smaller than the third one in Table 3.1 (where $\beta = 0.2$). We emphasise nevertheless that, comparing columns 5 and 6 in the two tables, there is no evidence of a dependence on β of E_2 and E_3. The price you pay is that, to implement our corrections, the knowledge of α is request. Such corrections turn out to be a clear improvement also because for α that is less than 1 the original estimator (3.18) is well-performed only for those values of the couple (α, β) which respect the condition (3.22) while, for $\alpha \geq 1$, there is no $\beta \in (0, \frac{1}{2})$ for which such a condition can hold. That's the reason why, in the lower part of both Tables 3.1 and 3.2, E_1 is so big.
Using our main results, instead, we get E_2 and E_3 that are always small and so we obtain two corrections which make the unbiased estimator always well-performed without adding any requirement on α or β.

3.5 Developments in small time

In order to prove our main results we need some developments in small time.

In the sequel, for $\delta \geq 0$, we will denote $R(\Delta_{n,i}^{\delta}, x)$ for any function $R(\Delta_{n,i}^{\delta}, x) = R_{i,n}(x)$, where $R_{i,n} : \mathbb{R} \longrightarrow \mathbb{R}$, $x \mapsto R_{i,n}(x)$ is such that

$$\exists c > 0 \qquad |R_{i,n}(x)| \leq c(1 + |x|^c)\Delta_{n,i}^{\delta} \tag{3.23}$$

with c independent of i, n.
The functions R represent the term of rest and have the following useful property, consequence of the just given definition:

$$R(\Delta_{n,i}^{\delta}, x) = \Delta_{n,i}^{\delta} R(\Delta_{n,i}^{0}, x). \tag{3.24}$$

We point out that it does not involve the linearity of R, since the functions R on the left and on the right side are not necessarily the same but only two functions on which the control (3.23) holds with $\Delta_{n,i}^{\delta}$ and $\Delta_{n,i}^{0}$, respectively.

We now state a proposition in which we prove a bound for the total variation distance between the conditional law of the rescaled Levy process and the α-stable distribution. It will be shown in Section 3.7.

Proposition 21. *Suppose that Assumptions 1 to 4 hold. Let $(S_t^{\alpha})_{t\geq 0}$ be an α-stable process. Let h be a measurable bounded function such that $\|h\|_{pol} := \sup_{x\in\mathbb{R}}(\frac{|h(x)|}{1+|x|^p}) < \infty$, for some $p \geq 1$, $p \geq \alpha$ hence*

$$|h(x)| \leq \|h\|_{pol} (|x|^p + 1). \tag{3.25}$$

Moreover we denote $\|h\|_{\infty} := \sup_{x\in\mathbb{R}} |h(x)|$. Then, for any $\epsilon > 0$,

$$|\mathbb{E}[h(\Delta_n^{-\frac{1}{\alpha}} L_{\Delta_n})] - \mathbb{E}[h(S_1^{\alpha})]| \leq C_{\epsilon}\Delta_n log(\Delta_n^{-\frac{1}{\alpha}}) \|h\|_{\infty} + \tag{3.26}$$

$$+C_{\epsilon}\Delta_n^{\frac{1}{\alpha}} \|h\|_{\infty}^{1-\frac{\alpha}{p}-\epsilon} \|h\|_{pol}^{\frac{\alpha}{p}+\epsilon} log(\Delta_n^{-\frac{1}{\alpha}}) + C_{\epsilon}\Delta_n^{\frac{1}{\alpha}} \|h\|_{\infty}^{1+\frac{1}{p}-\frac{\alpha}{p}+\epsilon} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p}-\epsilon} log(\Delta_n^{-\frac{1}{\alpha}})1_{\{\alpha>1\}},$$

where C_{ϵ} is a constant independent of n.

Remark 12. *The previous theorem is an extension of Theorem 4.2 in [19], it will be useful when $\|h\|_{\infty}$ is large, compared to $\|h\|_{pol}$. For instance, it is the case if consider a function $h(x) := |x|^2 1_{|x|\leq M}$ for M large.*

The next proposition will be useful for the proof of main results. It will be shown in the appendix.

Proposition 22. *Suppose that Assumptions 1 to 4 hold. We define, for $i \in \{0, ..., n-1\}$,*

$$\Delta X_i^J := \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz) \quad \text{and} \quad \Delta \tilde{X}_i^J := \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{t_i})\, z\, \tilde{\mu}(ds, dz).$$

1. *Then we have*

$$(\Delta X_i^J)^2 \varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i) = (\Delta \tilde{X}_i^J)^2 \varphi_{\Delta_{n,i}^{\beta}}(\Delta \tilde{X}_i^J) + o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1}), \tag{3.27}$$

where $o_{L^1}(\Delta_{n,i}^k)$ is such that $\mathbb{E}_i[|o_{L^1}(\Delta_{n,i}^k)|] = R(\Delta_{n,i}^k, x)$, with the notation $\mathbb{E}_i[.] = \mathbb{E}[.|\mathcal{F}_{t_i}]$, $(\mathcal{F}_{t_i})_{t\in[0,T]}$ has been defined in Lemma 25.
Moreover, for each $\tilde{\epsilon} > 0$ and f the function introduced in the definition of Q_n,

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\Delta X_i^J)^2\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) = \tag{3.28}$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\Delta \tilde{X}_i^J)^2\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) + o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})}).$$

2. *We also have*

$$(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)\Delta X_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) = \tag{3.29}$$

$$= (\int_{t_i}^{t_{i+1}} a(X_s)dW_s)\Delta \tilde{X}_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) + o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1})$$

and

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)\Delta X_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) =$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)\Delta \tilde{X}_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) + o_{L^1}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})}). \tag{3.30}$$

3.6 Proof of main results

In our proofs, the following lemma will be useful:

Lemma 28. *Let us denote by $\Delta X_i^J := \int_{t_i}^{t_{i+1}}\int_{\mathbb{R}\setminus\{0\}}\gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz)$ and let $\mathcal{F}_s$ be the filtration defined in Lemma 25. Then*

1. *For each $q \geq 2$ $\exists \epsilon > 0$ such that*

$$\mathbb{E}[|\Delta X_i^J 1_{\{|\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\}}|^q|\mathcal{F}_{t_i}] = R(\Delta_{n,i}^{1+\beta(q-\alpha)}, X_{t_i}) = R(\Delta_{n,i}^{1+\epsilon}, X_{t_i}). \tag{3.31}$$

$$\mathbb{E}[|\Delta \tilde{X}_i^J 1_{\{|\Delta \tilde{X}_i^J| \leq 4\Delta_{n,i}^\beta\}}|^q|\mathcal{F}_{t_i}] = R(\Delta_{n,i}^{1+\beta(q-\alpha)}, X_{t_i}) = R(\Delta_{n,i}^{1+\epsilon}, X_{t_i}). \tag{3.32}$$

2. *For each $q \geq 1$ we have*

$$\mathbb{E}[|\Delta X_i^J 1_{\left\{\frac{\Delta_{n,i}^\beta}{4} \leq |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}|^q|\mathcal{F}_{t_i}] = R(\Delta_{n,i}^{1+\beta(q-\alpha)}, X_{t_i}). \tag{3.33}$$

Proof. Reasoning as in Lemma 10 in Chapter 1 we easily get (3.31). Observing that $\Delta \tilde{X}_i^J$ is a particular case of ΔX_i^J where γ is fixed, evaluated in X_{t_i}, it follows that (3.32) can be obtained in the same way of (3.31). Using the bound on ΔX_i^J obtained from the indicator function we get that the left hand side of (3.33) is upper bounded by

$$c\Delta_{n,i}^{\beta q}\mathbb{E}[1_{\left\{\frac{\Delta_{n,i}^\beta}{4} \leq |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}|\mathcal{F}_{t_i}] \leq \Delta_{n,i}^{\beta q}R(\Delta_{n,i}^{1-\alpha\beta}, X_{t_i}),$$

where in the last inequality we have used Lemma 11 in Chapter 1 on the interval $[t_i, t_{i+1}]$ instead of on $[0, h]$. From property (3.24) of R we get (3.33). □

3.6.1 Proof of Theorem 10.

We observe that, using the dynamic (4.3) of X and the definition of the continuous part X^c, we have that

$$X_{t_{i+1}} - X_{t_i} = (X^c_{t_{i+1}} - X^c_{t_i}) + \int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz). \tag{3.34}$$

Replacing (3.34) in definition (3.3) of Q_n we have

$$Q_n = \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(X^c_{t_{i+1}} - X^c_{t_i})^2 + \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(X^c_{t_{i+1}} - X^c_{t_i})^2(\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) - 1) +$$
$$+ \frac{2}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(X^c_{t_{i+1}} - X^c_{t_i})(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz))\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) + \tag{3.35}$$
$$+ \frac{1}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}\setminus\{0\}} \gamma(X_{s^-})\, z\, \tilde{\mu}(ds, dz))^2\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) =: \sum_{j=1}^{4} I_j^n.$$

Comparing (3.35) with (3.7), using also definition (3.6) of $\tilde{Q}_n$, it follows that our goal is to show that $I_2^n + I_3^n = \mathcal{E}_n$, that is both $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ and $o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})})$. In the sequel the constant c may change value from line to line. By the definition of X^c we have

$$|I_2^n| \le \frac{c}{n}\sum_{i=0}^{n-1} |\frac{f(X_{t_i})}{\Delta_{n,i}}|[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^2 + |\int_{t_i}^{t_{i+1}} b(X_s)ds|^2]|\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) - 1| =$$
$$=: |I_{2,1}^n| + |I_{2,2}^n|.$$

Concerning $I_{2,1}^n$, using Holder inequality we have

$$\mathbb{E}[|I_{2,1}^n|] \le \frac{c}{n}\sum_{i=0}^{n-1} \mathbb{E}[|\frac{f(X_{t_i})}{\Delta_{n,i}}|\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^{2p}]^{\frac{1}{p}}\mathbb{E}_i[|\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) - 1|^q]^{\frac{1}{q}}], \tag{3.36}$$

where $\mathbb{E}_i$ is the conditional expectation wit respect to $\mathcal{F}_{t_i}$.
We now use Burkholder-Davis-Gundy inequality to get, for $p_1 \ge 2$,

$$\mathbb{E}_i[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^{p_1}]^{\frac{1}{p_1}} \le \mathbb{E}_i[|\int_{t_i}^{t_{i+1}} a^2(X_s)ds|^{\frac{p_1}{2}}]^{\frac{1}{p_1}} \le R(\Delta_{n,i}^{\frac{p_1}{2}}, X_{t_i})^{\frac{1}{p_1}} = R(\Delta_{n,i}^{\frac{1}{2}}, X_{t_i}), \tag{3.37}$$

where in the last inequality we have used the polynomial growth of a and third point of Lemma 25. We now observe that, from the definition of φ we know that $\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) - 1$ is different from 0 only if $|\Delta X_i| > \Delta^\beta_{n,i}$. We consider two different sets: $|\Delta X_i^J| < \frac{1}{2}\Delta^\beta_{n,i}$ and $|\Delta X_i^J| \ge \frac{1}{2}\Delta^\beta_{n,i}$. We recall that $\Delta X_i = \Delta X_i^c + \Delta X_i^J$ and so, if $|\Delta X_i| > \Delta^\beta_{n,i}$ and $|\Delta X_i^J| < \frac{1}{2}\Delta^\beta_{n,i}$, then it means that $|\Delta X_i^c|$ must be more than $\frac{1}{2}\Delta^\beta_{n,i}$. Using a conditional version of Tchebychev inequality we have that, $\forall r > 1$,

$$\mathbb{P}_i(|\Delta X_i^c| \ge \frac{1}{2}\Delta^\beta_{n,i}) \le c\frac{\mathbb{E}_i[|\Delta X_i^c|^r]}{\Delta^{\beta r}_{n,i}} \le R(\Delta_{n,i}^{(\frac{1}{2}-\beta)r}, X_{t_i}), \tag{3.38}$$

where $\mathbb{P}_i$ is the conditional probability with respect to $\mathcal{F}_{t_i}$; the last inequality follows from the fourth point of Lemma 25. If otherwise $|\Delta X_i^J| \ge \frac{1}{2}\Delta_{n,i}^\beta$, then we introduce the set $N_{i,n} := \left\{ |\Delta L_s| \le \frac{2\Delta_{n,i}^\beta}{\gamma_{min}}; \forall s \in (t_i, t_{i+1}] \right\}$. We have

$$\mathbb{P}_i(\left\{ |\Delta X_i^J| \ge \frac{1}{2}\Delta_{n,i}^\beta \right\} \cap (N_{i,n})^c) \le \mathbb{P}_i((N_{i,n})^c),$$

with

$$\mathbb{P}_i((N_{i,n})^c) = \mathbb{P}_i(\exists s \in (t_i, t_{i+1}] : |\Delta L_s| > \frac{\Delta_{n,i}^\beta}{2\gamma_{min}}) \le c \int_{t_i}^{t_{i+1}} \int_{\frac{\Delta_{n,i}^\beta}{2\gamma_{min}}}^{\infty} F(z)dzds \le c\Delta_{n,i}^{1-\alpha\beta}, \tag{3.39}$$

where we have used the third point of Assumption 4. Furthermore, using Markov inequality,

$$\mathbb{P}_i(\left\{ |\Delta X_i^J| \ge \frac{1}{2}\Delta_{n,i}^\beta \right\} \cap N_{i,n}) \le c\mathbb{E}_i[|\Delta X_i^J|^r 1_{N_{i,n}}]R(\Delta_{n,i}^{-\beta r}, X_{t_i}) \le \tag{3.40}$$

$$\le R(\Delta_{n,i}^{-\beta r+1+\beta(r-\alpha)}, X_{t_i}) = R(\Delta_{n,i}^{1-\beta\alpha}, X_{t_i}),$$

where in the last equality we have used the first point of Lemma 28, observing that $1_{N_{i,n}}$ acts like the indicator function in (3.31) (see also Lemma 10 Chapter 1). Now using (3.38), (3.39), (3.40) and the arbitrariness of r we have

$$\mathbb{P}_i(|\Delta X_i| > \Delta_{n,i}^\beta) =$$

$$= \mathbb{P}_i(|\Delta X_i| > \Delta_{n,i}^\beta, |\Delta X_i^J| < \frac{1}{2}\Delta_{n,i}^\beta) + \mathbb{P}_i(|\Delta X_i| > \Delta_{n,i}^\beta, |\Delta X_i^J| \ge \frac{1}{2}\Delta_{n,i}^\beta) \le$$

$$\le R(\Delta_{n,i}^{[(\frac{1}{2}-\beta)r]\wedge[1-\alpha\beta]}, X_{t_i}) = R(\Delta_{n,i}^{1-\alpha\beta}, X_{t_i}). \tag{3.41}$$

Taking p big and q next to 1 in (3.36) and replacing there (3.37) with $p_1 = 2p$ and (3.41) we get, $\forall \epsilon > 0$,

$$\mathbb{E}[|I_{2,1}^n|] \le \frac{1}{n}\sum_{i=1}^{n-1} \mathbb{E}[|\frac{f(X_{t_i})}{\Delta_{n,i}}|R(\Delta_{n,i}, X_{t_i})R(\Delta_{n,i}^{1-\alpha\beta-\epsilon}, X_{t_i})] \le$$

$$\le \Delta_n^{1-\alpha\beta-\epsilon}\frac{c}{n}\sum_{i=1}^{n-1} \mathbb{E}[|f(X_{t_i})|R(1, X_{t_i})].$$

Now, for each $\tilde{\epsilon} > 0$, we can always find an ϵ smaller than it, that is enough to get $I_{2,1}^n = o_{\mathbb{P}}(\Delta_n^{1-\alpha\beta-\tilde{\epsilon}})$ and so

$$I_{2,1}^n = o_{\mathbb{P}}(\Delta_n^{(1-\alpha\beta-\tilde{\epsilon})\wedge(\frac{1}{2}-\tilde{\epsilon})}). \tag{3.42}$$

Moreover $I_{2,1}^n = o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$, indeed

$$\frac{\mathbb{E}[|I_{2,1}^n|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{1-\alpha\beta-\beta(2-\alpha)-\epsilon}\frac{1}{n}\sum_{i=1}^{n-1} \mathbb{E}[|f(X_{t_i})|R(1, X_{t_i})]. \tag{3.43}$$

Since we can always find an $\epsilon > 0$ such that $\beta < \frac{1}{2} - \epsilon$, we observe that the exponent on Δ_n is positive. Using the polynomial growth of both f and R and the third point of Lemma 26 we get that (3.43) goes to zero in norm 1 and so in probability.

Let us now consider $I^n_{2,2}$. We observe that $|\varphi_{\Delta^\beta_{n,i}}(\Delta X_i) - 1| \le c$. Moreover, by adding and subtracting $b(X_{t_i})$ in the integral we get

$$(\int_{t_i}^{t_{i+1}} b(X_s)ds)^2 \le c\Delta^2_{n,i} b^2(X_{t_i}) + c(\int_{t_i}^{t_{i+1}} [b(X_s) - b(X_{t_i})]ds)^2.$$

Using Jensen inequality and the regularity of b we get

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} b(X_s)ds)^2] \le R(\Delta^2_{n,i}, X_{t_i}) + \int_{t_i}^{t_{i+1}} \|b'\|^2_\infty \mathbb{E}_i[|X_s - X_{t_i}|^2]ds \le$$

$$\le R(\Delta^2_{n,i}, X_{t_i}) + c\int_{t_i}^{t_{i+1}} \Delta_{n,i}(1 + |X_{t_i}|^2)ds = R(\Delta^2_{n,i}, X_{t_i}), \tag{3.44}$$

where in the last inequality we have used the second point of Lemma 25. Using (3.44) we get

$$\mathbb{E}[|I^n_{2,2}|] \le \Delta_n \frac{1}{n}\sum_{i=1}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})] \tag{3.45}$$

and so $I^n_{2,2} = o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ since

$$\frac{\mathbb{E}[|I^n_{2,2}|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{1-\beta(2-\alpha)} \frac{1}{n}\sum_{i=1}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})], \tag{3.46}$$

that goes to 0 because the exponent on Δ_n is always more than zero, f and R have polynomial growth and we can use the third point of Lemma 26. Moreover, using (3.45), we have that $I^n_{2,2} = o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}})$ and so it is $o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$. From (3.42), (3.43), (3.45) and (3.46) we get $I^n_2 = \mathcal{E}_n$.
Let us now consider I^n_3. From the definition of the process (X^c_t) it is

$$\frac{2}{n}\sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}}[\int_{t_i}^{t_{i+1}} b(X_s)ds + \int_{t_i}^{t_{i+1}} a(X_s)dW_s]\Delta X^J_i \varphi_{\Delta^\beta_{n,i}}(\Delta X_i) =: I^n_{3,1} + I^n_{3,2}.$$

We use on $I^n_{3,1}$ Cauchy-Schwartz inequality, (3.44) and Lemma 10 in [3], getting

$$\mathbb{E}[|I^n_{3,1}|] \le \frac{2}{n}\sum_{i=0}^{n-1} \mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}} R(\Delta_{n,i}^{1+\beta(2-\alpha)}, X_{t_i})^{\frac{1}{2}} R(\Delta^2_{n,i}, X_{t_i})^{\frac{1}{2}}] \le$$

$$\le \Delta_n^{\frac{1}{2}+\frac{\beta}{2}(2-\alpha)} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})],$$

where we have also used property (3.24) on R. We observe it is $\frac{1}{2} + \beta - \frac{\alpha\beta}{2} > \frac{1}{2}$ if and only if $\beta(1 - \frac{\alpha}{2}) > 0$, that is always true. We can therefore say that $I^n_{3,1} = o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}})$ and so

$$I^n_{3,1} = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.47}$$

Moreover,

$$\frac{\mathbb{E}[|I^n_{3,1}|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{\frac{1}{2}-\beta+\frac{\alpha\beta}{2}} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})], \tag{3.48}$$

that goes to zero using the polynomial growth of both f and R and the third point of Lemma 26 and observing that the exponent on Δ_n is positive for $\beta < \frac{1}{2}\frac{1}{(1-\frac{\alpha}{2})}$, that is always true.

Concerning $I_{3,2}^n$, we start proving that $I_{3,2}^n = o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$. From (3.29) in Proposition 22 we have

$$\frac{I_{3,2}^n}{\Delta_n^{\beta(2-\alpha)}} = \frac{1}{\Delta_n^{\beta(2-\alpha)}} \frac{2}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} [\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) \int_{t_i}^{t_{i+1}} a(X_s) dW_s + o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1})]. \tag{3.49}$$

By the definition of o_{L^1} the last term here above goes to zero in norm 1 and so in probability. The first term of (3.49) can be seen as

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}} \frac{2}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} \Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) [\int_{t_i}^{t_{i+1}} a(X_{t_i}) dW_s + \int_{t_i}^{t_{i+1}} (a(X_s) - a(X_{t_i})) dW_s]. \tag{3.50}$$

On the first term of (3.50) here above we want to use Lemma 9 of [36] in order to get that it converges to zero in probability, so we have to show the following:

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}} \frac{2}{n} \sum_{i=0}^{n-1} \mathbb{E}_i [\frac{f(X_{t_i})}{\Delta_{n,i}} \Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) \int_{t_i}^{t_{i+1}} a(X_{t_i}) dW_s] \xrightarrow{\mathbb{P}} 0, \tag{3.51}$$

$$\frac{1}{\Delta_n^{2\beta(2-\alpha)}} \frac{4}{n^2} \sum_{i=0}^{n-1} \mathbb{E}_i [\frac{f^2(X_{t_i})}{\Delta_{n,i}^2} (\Delta \tilde{X}_i^J)^2 \varphi_{\Delta_{n,i}^\beta}^2 (\Delta \tilde{X}_i^J) (\int_{t_i}^{t_{i+1}} a(X_{t_i}) dW_s)^2] \xrightarrow{\mathbb{P}} 0, \tag{3.52}$$

where $\mathbb{E}_i[.] = \mathbb{E}[.|\mathcal{F}_{t_i}]$.
Using the independence between W and L we have that the left hand side of (3.51) is

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}} \frac{2}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} \mathbb{E}_i [\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)] \mathbb{E}_i [\int_{t_i}^{t_{i+1}} a(X_{t_i}) dW_s] = 0. \tag{3.53}$$

Now, in order to prove (3.52), we use Holder inequality with p big and q next to 1 on its left hand side, getting it is upper bounded by

$$\Delta_n^{1-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} \frac{f^2(X_{t_i})}{\Delta_{n,i}^2} \mathbb{E}_i [(\int_{t_i}^{t_{i+1}} a(X_{t_i}) dW_s)^{2p}]^{\frac{1}{p}} \mathbb{E}_i [|\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)|^{2q}]^{\frac{1}{q}} \le$$

$$\le \Delta_n^{1-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} \frac{f^2(X_{t_i})}{\Delta_{n,i}^2} R(\Delta_{n,i}, X_{t_i}) R(\Delta_{n,i}^{\frac{1}{q}+\frac{\beta}{q}(2q-\alpha)}, X_{t_i}) \le$$

$$\le \Delta_n^{1-2\beta(2-\alpha)+2\beta-\alpha\beta-\epsilon} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} f^2(X_{t_i}) R(1, X_{t_i}), \tag{3.54}$$

where we have used (3.37), (3.32) and property (3.24) of R. We observe that the exponent on Δ_n is positive if $\beta < \frac{1}{2-\alpha} - \epsilon$ and we can always find an $\epsilon > 0$ such that it is true. Hence, using also that $\frac{1}{n\Delta_n}$ is bounded, the polynomial growth of both f and R and the third point of Lemma 26, we get that (3.54) goes to zero in norm 1 and so in probability.
Concerning the second term of (3.50), using Cauchy-Schwartz inequality, (3.37) and (3.32) we have

$$\mathbb{E}_i [|\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)| | \int_{t_i}^{t_{i+1}} [a(X_s) - a(X_{t_i})] dW_s|] \le$$

$$\le \mathbb{E}_i [|\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)|^2]^{\frac{1}{2}} \mathbb{E}_i [| \int_{t_i}^{t_{i+1}} [a(X_s) - a(X_{t_i})] dW_s|^2]^{\frac{1}{2}} \le$$

$$\leq R(\Delta_{n,i}^{\frac{1}{2}+\frac{\beta}{2}(2-\alpha)}, X_{t_i})\mathbb{E}_i[\int_{t_i}^{t_{i+1}} \|a'\|_\infty |X_s - X_{t_i}|^2 ds]^{\frac{1}{2}} \leq$$

$$\leq \Delta_{n,i}^{\frac{1}{2}+\frac{\beta}{2}(2-\alpha)} R(1, X_{t_i})(\int_{t_i}^{t_{i+1}} \Delta_{n,i}(1+|X_{t_i}|^2)ds)^{\frac{1}{2}} \leq \Delta_{n,i}^{\frac{3}{2}+\frac{\beta}{2}(2-\alpha)} R(1, X_{t_i}), \quad (3.55)$$

where we have also used the third point of Lemma 25 and the property (3.24) of R. Replacing (3.55) in the second term of (3.50) we get it is upper bounded in norm 1 by

$$\Delta_n^{\frac{1}{2}-\beta+\frac{\alpha\beta}{2}} \frac{1}{n}\sum_{i=0}^{n-1} \mathbb{E}[|f(X_{t_i})|R(1, X_{t_i})], \quad (3.56)$$

that goes to zero since the exponent on Δ_n is more than 0 for $\beta < \frac{1}{2}\frac{1}{(1-\frac{\alpha}{2})}$, that is always true. Using (3.49) - (3.52) and (3.56) we get

$$\frac{I_{3,2}^n}{\Delta_n^{\beta(2-\alpha)}} \xrightarrow{\mathbb{P}} 0. \quad (3.57)$$

We now want to show that $I_{3,2}^n$ is also $o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$.
Using (3.30) in Proposition 22 we get it is enough to prove that

$$\frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}[\Delta\tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)\int_{t_i}^{t_{i+1}} a(X_s)dW_s] \xrightarrow{\mathbb{P}} 0, \quad (3.58)$$

where the left hand side here above can be seen as (3.50), with the only difference that now we have $\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}$ instead of $\Delta_n^{\beta(2-\alpha)}$. We have again, acting like we did in (3.53) and (3.54),

$$\frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\frac{2}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\mathbb{E}_i[\Delta\tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)\int_{t_i}^{t_{i+1}} a(X_{t_i})dW_s] \xrightarrow{\mathbb{P}} 0 \quad (3.59)$$

and

$$\frac{1}{\Delta_n^{2(\frac{1}{2}-\tilde{\epsilon})}}\frac{4}{n^2}\sum_{i=0}^{n-1}\mathbb{E}_i[\frac{f^2(X_{t_i})}{\Delta_{n,i}^2}(\Delta\tilde{X}_i^J)^2\varphi^2_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)(\int_{t_i}^{t_{i+1}} a(X_{t_i})dW_s)^2] \leq$$

$$\leq \Delta_n^{2\tilde{\epsilon}+2\beta-\alpha\beta-\epsilon}\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} f^2(X_{t_i})R(1, X_{t_i}), \quad (3.60)$$

that goes to zero in norm 1 and so in probability. Using also (3.55) we have that $\frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\frac{2}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\Delta\tilde{X}_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)\int_{t_i}^{t_{i+1}}[a(X_s)-a(X_{t_i})]dW_s$ is upper bounded in norm 1 by

$$\Delta_n^{\frac{\beta}{2}(2-\alpha)+\tilde{\epsilon}}\frac{1}{n}\mathbb{E}[|f(X_{t_i})||R(1, X_{t_i})|], \quad (3.61)$$

that goes to zero since the exponent on Δ_n is always positive. Using (3.58) - (3.61) we get $I_{3,2}^n = o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}})$ and so

$$I_{3,2}^n = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \quad (3.62)$$

From (3.42), (3.43), (3.45), (3.46), (3.47), (3.48), (3.57) and (3.62) it follows (3.7). Now, in order to prove (3.8), we recall the definition of X_t^c:

$$X_{t_{i+1}}^c - X_{t_i}^c = \int_{t_i}^{t_{i+1}} b(X_s)ds + \int_{t_i}^{t_{i+1}} a(X_s)dW_s. \quad (3.63)$$

Replacing (3.63) in (3.7) and comparing it with (3.8) it follows that our goal is to show that

$$A_1^n + A_2^n :=$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} b(X_s)ds)^2 + \frac{2}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} b(X_s)ds)(\int_{t_i}^{t_{i+1}} a(X_s)dW_s) = \mathcal{E}_n.$$

Using (3.44) and property (3.24) of R we know that

$$\frac{\mathbb{E}[|A_1^n|]}{\Delta_n^{\beta(2-\alpha)}} \le \frac{1}{\Delta_n^{\beta(2-\alpha)}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}}R(\Delta_{n,i}^2, X_{t_i})] \le \tag{3.64}$$

$$\le \Delta_n^{1-\beta(2-\alpha)}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f(X_{t_i})|R(1, X_{t_i})]$$

and

$$\frac{\mathbb{E}[|A_1^n|]}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}} \le \Delta_n^{\frac{1}{2}+\tilde{\epsilon}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f(X_{t_i})|R(1, X_{t_i})], \tag{3.65}$$

that go to zero since the exponent on Δ_n is always more than 0, f and R have both polynomial growth and we can use the third point of Lemma 26.
Let us now consider A_2^n. By adding and subtracting $b(X_{t_i})$ in the first integral, as we have already done, we get that

$$A_2^n = \sum_{i=0}^{n-1}\zeta_{n,i} + A_{2,2}^n := \frac{2}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} b(X_{t_i})ds)(\int_{t_i}^{t_{i+1}} a(X_s)dW_s) +$$

$$+ \frac{2}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\int_{t_i}^{t_{i+1}} [b(X_s) - b(X_{t_i})]ds)(\int_{t_i}^{t_{i+1}} a(X_s)dW_s).$$

Using Lemma 9 in [36], we want to show that

$$\sum_{i=0}^{n-1}\zeta_{n,i} = \mathcal{E}_n \tag{3.66}$$

and so that the following convergences hold:

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}\zeta_{n,i} \xrightarrow{\mathbb{P}} 0 \qquad \frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\sum_{i=0}^{n-1}\zeta_{n,i} \xrightarrow{\mathbb{P}} 0; \tag{3.67}$$

$$\frac{1}{\Delta_n^{2\beta(2-\alpha)}}\sum_{i=0}^{n-1}\zeta_{n,i}^2 \xrightarrow{\mathbb{P}} 0 \qquad \frac{1}{\Delta_n^{2(\frac{1}{2}-\tilde{\epsilon})}}\sum_{i=0}^{n-1}\zeta_{n,i}^2 \xrightarrow{\mathbb{P}} 0. \tag{3.68}$$

We have

$$\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}] = \frac{2}{\Delta_n^{\beta(2-\alpha)}}\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\Delta_{n,i}b(X_{t_i})\mathbb{E}_i[\int_{t_i}^{t_{i+1}} a(X_s)dW_s] = 0$$

and so the two convergences in (3.67) both hold. Concerning (3.68), using (3.37) we have

$$\Delta_n^{1-2\beta(2-\alpha)}\frac{1}{n\Delta_n}\frac{c}{n}\sum_{i=0}^{n-1}f^2(X_{t_i})b^2(X_{t_i})\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2] \le$$

$$\leq \Delta_n^{2-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} f^2(X_{t_i}) b^2(X_{t_i}) R(1, X_{t_i})$$

and

$$\Delta_n^{1-2(\frac{1}{2}-\tilde{\epsilon})} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} f^2(X_{t_i}) b^2(X_{t_i}) \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} a(X_s) dW_s)^2] \leq$$

$$\leq \Delta_n^{1+2\tilde{\epsilon}} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} f^2(X_{t_i}) b^2(X_{t_i}) R(1, X_{t_i}),$$

that go to zero in norm 1 and so in probability since $\frac{1}{n\Delta_n}$ is bounded and the fact that the exponent on Δ_n is always positive. It follows (3.68) and so (3.66). Concerning $A_{2,2}^n$, using Holder inequality, (3.37), the regularity of b and Jensen inequality it is

$$\mathbb{E}[|A_{2,2}^n|] \leq c\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}} \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} \|b'\|_\infty |X_s - X_{t_i}| ds)^q]^{\frac{1}{q}} R(\Delta_{n,i}^{\frac{1}{2}}, X_{t_i})] \leq$$

$$\leq c\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}} (\Delta_{n,i}^{q-1} \int_{t_i}^{t_{i+1}} \mathbb{E}_i[|X_s - X_{t_i}|^q] ds)^{\frac{1}{q}} R(\Delta_{n,i}^{\frac{1}{2}}, X_{t_i})] \leq$$

$$\leq c\frac{1}{n} \sum_{i=0}^{n-1} \mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}} (\Delta_{n,i}^{q-1} \int_{t_i}^{t_{i+1}} \Delta_{n,i}(1 + |X_{t_i}|^q) ds)^{\frac{1}{q}} R(\Delta_{n,i}^{\frac{1}{2}}, X_{t_i})].$$

So we get

$$\frac{\mathbb{E}[|A_{2,2}^n|]}{\Delta_n^{\beta(2-\alpha)}} \leq \Delta_n^{\frac{1}{q}+\frac{1}{2}-\beta(2-\alpha)} \frac{c}{n} \sum_{i=0}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})] \qquad \text{and} \tag{3.69}$$

$$\frac{\mathbb{E}[|A_{2,2}^n|]}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}} \leq \Delta_n^{\frac{1}{q}+\tilde{\epsilon}} \frac{c}{n} \sum_{i=0}^{n-1} \mathbb{E}[|f(X_{t_i})| R(1, X_{t_i})]. \tag{3.70}$$

Since it holds for $q \geq 2$, the best choice is to take $q = 2$, in this way we get that (3.69) and (3.70) go to 0 in norm 1, using the polynomial growth of both f and R, the third point of Lemma 26 and the fact that the exponent on Δ_n is in both cases more than zero, because of $\beta < \frac{1}{2-\alpha}$.
From (3.64), (3.65), (3.67), (3.69) and (3.70) it follows (3.8).

3.6.2 Proof of Theorem 11

Proof. We want to prove

$$\frac{1}{n} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} (\int_{t_i}^{t_{i+1}} a(X_s) dW_s)^2 - \frac{1}{n} \sum_{i=0}^{n-1} f(X_{t_i}) a^2(X_{t_i}) = \frac{Z_n}{\sqrt{n}} + \mathcal{E}_n, \tag{3.71}$$

and

$$\tilde{Q}_n = \hat{Q}_n + \frac{1}{\Delta_n^{\beta(2-\alpha)}} \mathcal{E}_n,$$

where $\mathcal{E}_n$ is always $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ and, if $\beta > \frac{1}{4-\alpha}$, then it is also $o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon}) \wedge (1-\alpha\beta-\tilde{\epsilon})})$. We can rewrite the last equation here above as

$$\tilde{Q}_n = \hat{Q}_n + o_{\mathbb{P}}(1) \tag{3.72}$$

and, for $\beta > \frac{1}{4-\alpha}$,

$$\tilde{Q}_n = \hat{Q}_n + \frac{1}{\Delta_n^{\beta(2-\alpha)}} o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.73}$$

Using them and (3.8) it follows (3.10). Hence we are now left to prove (3.71) - (3.73).

Proof of (3.71).
We can see the left hand side of (3.71) as

$$\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}[(\int_{t_i}^{t_{i+1}} a(X_s)dW_s)^2 - \int_{t_i}^{t_{i+1}} a^2(X_s)ds]+$$

$$+\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}[a^2(X_s) - a^2(X_{t_i})]ds =: M_n^Q + B_n.$$

We want to show that $B_n = \mathcal{E}_n$, it means that it is both $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$ and $o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$. Considering the development up to second order of the function a^2 we get

$$a^2(X_s) - a^2(X_{t_i}) = 2aa'(X_{t_i})(X_s - X_{t_i}) + ((a')^2 + aa'')(\tilde{X}_{t_i})(X_s - X_{t_i})^2, \tag{3.74}$$

where $\tilde{X}_{t_i} \in [X_{t_i}, X_s]$. Replacing (3.74) in the definition of B_n it is

$$\frac{2}{n}\sum_{i=0}^{n-1}\frac{(faa')(X_{t_i})}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})ds+$$

$$+\frac{1}{n}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}((a')^2 + aa'')(\tilde{X}_{t_i})(X_s - X_{t_i})^2 ds =: B_1^n + B_2^n.$$

We start by proving that $B_2^n = o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$. Indeed, using Holder inequality taking p big and q next to 1, it is

$$\mathbb{E}[|B_2^n|] \le \frac{c}{n}\sum_{i=0}^{n-1}\mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}\mathbb{E}_i[|(a')^2 + aa''|^p(\tilde{X}_{t_i})]^{\frac{1}{p}}\mathbb{E}_i[|X_s - X_{t_i}|^{2q}]^{\frac{1}{q}}ds] \le$$

$$\le \frac{c}{n}\sum_{i=0}^{n-1}\mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}(1 + |X_{t_i}|^p)^{\frac{1}{p}}(1 + |X_{t_i}|^{2q})^{\frac{1}{q}}|s - t_i|^{\frac{1}{q}}ds] \le$$

$$\le \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}}R(\Delta_{n,i}^{1+\frac{1}{q}}, X_{t_i})] \le \frac{\Delta_n^{\frac{1}{q}}}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f((a')^2 + aa'')|(X_{t_i})R(1, X_{t_i})],$$

where we have used the third point of Lemma 1 for the first expected value and the second point on the second one. It follows

$$\frac{\mathbb{E}[|B_2^n|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{1-\epsilon-\beta(2-\alpha)}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f|(X_{t_i})R(1, X_{t_i})] \qquad \text{and} \tag{3.75}$$

$$\frac{\mathbb{E}[|B_2^n|]}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}} \le \Delta_n^{\frac{1}{2}-\epsilon+\tilde{\epsilon}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|f|(X_{t_i})R(1, X_{t_i})], \tag{3.76}$$

that go to zero using the polynomial growth of f and R . We have also used the third point of Lemma 2 and observed that the exponent on Δ_n is always more than

0.
Concerning B_1^n, we recall that from (4.3) it follows

$$X_s - X_{t_i} = \int_{t_i}^{s} b(X_u)du + \int_{t_i}^{s} a(X_u)dW_u + \int_{t_i}^{s}\int_{\mathbb{R}} \gamma(X_{u^-})z\tilde{\mu}(du,dz)$$

and so, replacing it in the definition of B_1^n, we get three terms: $B_1^n := I_1^n + I_2^n + I_3^n$. We start considering I_1^n on which we use polynomial growth of b and the third point of Lemma 25 to get

$$\mathbb{E}[|I_1^n|] \le \frac{2}{n}\sum_{i=0}^{n-1} \mathbb{E}[\frac{|f(X_{t_i})|}{\Delta_{n,i}}|aa'|(X_{t_i})\int_{t_i}^{t_{i+1}}(\int_{t_i}^{s} R(1,X_{t_i})du)ds] \le$$

$$\le \Delta_n \frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|faa'|(X_{t_i})R(1,X_{t_i})].$$

It follows

$$\frac{\mathbb{E}[|I_1^n|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{1-\beta(2-\alpha)}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|faa'|(X_{t_i})R(1,X_{t_i})] \qquad \text{and} \tag{3.77}$$

$$\frac{\mathbb{E}[|I_1^n|]}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}} \le \Delta_n^{\frac{1}{2}+\tilde{\epsilon}}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[|faa'|(X_{t_i})R(1,X_{t_i})], \tag{3.78}$$

that go to zero because of the polynomial growth of f, a, a', a'' and R and the fact that $1-\beta(2-\alpha) > 0$.
Considering I_2^n, we define $\zeta_{n,i} := \frac{2}{n}\frac{(faa')(X_{t_i})}{\Delta_{n,i}}\int_{t_i}^{t_{i+1}}(\int_{t_i}^{s} a(X_u)dW_u)ds$. We want to use Lemma 9 in [36] to get that

$$\frac{I_2^n}{\Delta_n^{\beta(2-\alpha)}} \xrightarrow{\mathbb{P}} 0 \qquad \text{and} \qquad \frac{I_2^n}{\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}} \xrightarrow{\mathbb{P}} 0 \tag{3.79}$$

and so we have to show the following :

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}] \xrightarrow{\mathbb{P}} 0, \qquad \frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}] \xrightarrow{\mathbb{P}} 0; \tag{3.80}$$

$$\frac{1}{\Delta_n^{2\beta(2-\alpha)}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}^2] \xrightarrow{\mathbb{P}} 0, \tag{3.81}$$

$$\frac{1}{\Delta_n^{2(\frac{1}{2}-\tilde{\epsilon})}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}^2] \xrightarrow{\mathbb{P}} 0. \tag{3.82}$$

By the definition of $\zeta_{n,i}$ it is $\mathbb{E}_i[\zeta_{n,i}] = 0$ and so (3.80) is clearly true. The left hand side of (3.81) is

$$\Delta_n^{1-2\beta(2-\alpha)}\frac{1}{n\Delta_n}\frac{4}{n}\sum_{i=0}^{n-1}\frac{(faa')^2(X_{t_i})}{\Delta_{n,i}^2}\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}(\int_{t_i}^{s} a(X_u)dW_u)ds)^2]. \tag{3.83}$$

Using Fubini theorem and Ito isometry we have

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}}(\int_{t_i}^{s} a(X_u)dW_u)ds)^2] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}}(t_{i+1}-s)a(X_s)dW_s)^2] = \tag{3.84}$$

$$= \mathbb{E}_i[\int_{t_i}^{t_{i+1}} (t_{i+1} - s^2) a^2(X_s) ds] \le R(\Delta_{n,i}^3, X_{t_i}),$$

where in the last inequality we have used polynomial growth of a and the third point of Lemma 25. Because of (3.84), we get that (3.83) is upper bounded by

$$\Delta_n^{1-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(\Delta_{n,i}, X_{t_i}) \le$$

$$\le \Delta_n^{2-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(1, X_{t_i}),$$

that converges to zero in norm 1 and so (3.81) follows, since $2 - 2\beta(2-\alpha) > 0$ for $\beta < \frac{1}{2-\alpha}$, that is always true. Moreover we have used that $n\Delta_n$ is bounded, the polynomial growth of f, a, a' and R and the third point of Lemma 25.
Acting in the same way we get that the left hand side of (3.82) is upper bounded by

$$\Delta_n^{1-(1-2\tilde{\epsilon})} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(\Delta_{n,i}, X_{t_i}) \le \Delta_n^{1+2\tilde{\epsilon}} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(1, X_{t_i}),$$

that goes to zero in norm 1. In order to show also

$$\frac{I_3^n}{\Delta_n^{\beta(2-\alpha)}} \xrightarrow{\mathbb{P}} 0 \qquad \text{and} \qquad \frac{I_3^n}{\Delta_n^{(\frac{1}{2}-\tilde{\epsilon}) \wedge (1-\alpha\beta-\tilde{\epsilon})}} \xrightarrow{\mathbb{P}} 0, \tag{3.85}$$

we define $\tilde{\zeta}_{n,i} := \frac{2}{n} \frac{(faa')(X_{t_i})}{\Delta_{n,i}} \int_{t_i}^{t_{i+1}} (\int_{t_i}^{s} \int_{\mathbb{R}} \gamma(X_{u^-}) z \tilde{\mu}(du, dz)) ds$. We have again $\mathbb{E}_i[\tilde{\zeta}_{n,i}] = 0$ and so (3.80) holds with $\tilde{\zeta}_{n,i}$ in place of $\zeta_{n,i}$. We now act like we did in (3.84), using Fubini theorem and Ito isometry. It follows

$$\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} (\int_{t_i}^{s} \int_{\mathbb{R}} \gamma(X_{u^-}) z \tilde{\mu}(du, dz) ds)^2] = \mathbb{E}_i[(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} (t_{i+1} - s) \gamma(X_{s^-}) z \tilde{\mu}(ds, dz))^2] =$$

$$= \mathbb{E}_i[\int_{t_i}^{t_{i+1}} (t_{i+1} - s)^2 \gamma^2(X_{s^-}) ds (\int_{\mathbb{R}} z^2 F(z) dz)] \le R(\Delta_{n,i}^3, X_{t_i}), \tag{3.86}$$

having used in the last inequality the definition of $\bar{\mu}(ds, dz)$, the fact that $\int_{\mathbb{R}} z^2 F(z) dz < \infty$, the polynomial growth of γ and the third point of Lemma 25. Replacing (3.86) in the left hand side of (3.81) and (3.82), with $\tilde{\zeta}_{n,i}$ in place of $\zeta_{n,i}$, we have

$$\frac{1}{\Delta_n^{2\beta(2-\alpha)}} \sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_{n,i}^2] \le \Delta_n^{1-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{c}{n} \sum_{i=0}^{n-1} \frac{(faa')^2(X_{t_i})}{\Delta_{n,i}^2} R(\Delta_{n,i}^3, X_{t_i}) \le$$

$$\le \Delta_n^{2-2\beta(2-\alpha)} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(1, X_{t_i})$$

$$\text{and } \frac{1}{\Delta_n^{1-2\tilde{\epsilon}}} \sum_{i=0}^{n-1} \mathbb{E}_i[\tilde{\zeta}_{n,i}^2] \le \Delta_n^{1+2\tilde{\epsilon}} \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=0}^{n-1} (faa')^2(X_{t_i}) R(1, X_{t_i}).$$

Again, they converge to zero in norm 1 and thus in probability since $2-2\beta(2-\alpha) > 0$ always holds, using also the polynomial growth of a, a', f and R and the third point of Lemma 26. Therefore, we get (3.85).
From (3.75), (3.76), (3.77), (3.78), (3.79) and (3.85) it follows that

$$B_n = \mathcal{E}_n. \tag{3.87}$$

Concerning $M_n^Q := \sum_{i=0}^{n-1} \hat{\zeta}_{n,i}$, we have to act differently for T fixed and $\lim_{n\to\infty} T = \infty$.
Case 1: T fixed
Genon - Catalot and Jacod have proved in [36] that, in the continuous framework, the following conditions are enough to get $\sqrt{n} M_n^Q \to N(0, 2\int_0^T (a^4 f^2)(X_s) H(s,0) ds)$ stably with respect to X:

- $\mathbb{E}_i[\hat{\zeta}_{n,i}] = 0$;
- $\sum_{i=0}^{n-1} \mathbb{E}_i[\hat{\zeta}_{n,i}^2] \xrightarrow{\mathbb{P}} 2\int_0^T (a^4 f^2)(X_s) H(s,0) ds$;
- $\sum_{i=0}^{n-1} \mathbb{E}_i[\hat{\zeta}_{n,i}^4] \xrightarrow{\mathbb{P}} 0$;
- $\sum_{i=0}^{n-1} \mathbb{E}_i[\hat{\zeta}_{n,i}(W_{t_{i+1}} - W_{t_i})] \xrightarrow{\mathbb{P}} 0$.

Theorem 2.2.15 in [46] adapts the previous theorem to our framework, in which there is the presence of jumps.
We observe that the conditions here above are respected, hence

$$M_n^Q = \frac{Z_n}{\sqrt{n}}, \text{ where } Z_n \xrightarrow{n} N(0, 2\int_0^T (a^4 f^2)(X_s) H(s,0) ds), \tag{3.88}$$

stably with respect to X.
Case 2: $\lim_{n\to\infty} T = \infty$.
In order to show the asymptotic normality we have to prove that $\hat{\zeta}_{n,i}$ is a martingal difference array such that $\sum_{i=0}^{n-1} \mathbb{E}_i[\hat{\zeta}_{n,i}^2] \xrightarrow{\mathbb{P}} 2\int_{\mathbb{R}} a^4(x) f^2(x) \pi(dx)$ and that $\sum_{i=0}^{n-1} \mathbb{E}_i[|\hat{\zeta}|_{n,i}^{2+\delta}] \xrightarrow{\mathbb{P}} 0$, for a constant $\delta > 0$. The previous conditions are true as a consequence of the the building of our sequence $\hat{\zeta}_{n,i}$ and using Lemma 27 with $\delta = 0$. So we get

$$M_n^Q = \frac{Z_n}{\sqrt{n}}, \text{ where } Z_n \xrightarrow{n} N(0, 2\int_{\mathbb{R}} a^4(x) f^2(x) \pi(dx)). \tag{3.89}$$

From (3.87), (3.88) and (3.89), it follows (3.71).

Proof of (3.72).
We use Proposition 22 replacing (3.27) in the definition (3.6) of $\tilde{Q}_n$. Recalling that the convergence in norm 1 implies the convergence in probability it is clear that we have to prove the result on

$$\frac{1}{n\Delta_n^{\beta(2-\alpha)}} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} (\Delta \tilde{X}_i^J)^2 \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) =$$

$$= \frac{1}{n\Delta_n^{\beta(2-\alpha)}} \sum_{i=0}^{n-1} \frac{f(X_{t_i})}{\Delta_{n,i}} \gamma^2(X_{t_i}) \Delta_{n,i}^{\frac{2}{\alpha}} (\frac{\Delta \tilde{X}_i^J}{\gamma(X_{t_i}) \Delta_{n,i}^{\frac{1}{\alpha}}})^2 \varphi_{\Delta_{n,i}^\beta}(\frac{\Delta \tilde{X}_i^J}{\gamma(X_{t_i}) \Delta_{n,i}^{\frac{1}{\alpha}}} \gamma(X_{t_i}) \Delta_{n,i}^{\frac{1}{\alpha}}), \tag{3.90}$$

where we have also rescaled the process in order to apply Proposition 21. We now define

$$g_{i,n}(y) := y^2 \varphi_{\Delta_{n,i}^\beta}(y \gamma(X_{t_i}) \Delta_{n,i}^{\frac{1}{\alpha}}), \tag{3.91}$$

hence we can rewrite (3.90) as

$$\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\gamma^2(X_{t_i})\Delta_{n,i}^{\frac{2}{\alpha}}[g_{i,n}(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})-\mathbb{E}[g_{i,n}(S_1^\alpha)]]+$$

$$+\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}\gamma^2(X_{t_i})\Delta_{n,i}^{\frac{2}{\alpha}}\mathbb{E}[g_{i,n}(S_1^\alpha)]=:\sum_{i=0}^{n-1}A_{1,i}^n+\hat{Q}_n, \tag{3.92}$$

where S_1^α is the α-stable process at time $t=1$. We want to show that $\sum_{i=0}^{n-1}A_{1,i}^n$ converges to zero in probability. With this purpose in mind, we take the conditional expectation of $A_{1,i}^n$ and we apply Proposition 21 on the interval $[t_i,t_{i+1}]$ instead of on $[0,\Delta_n]$, observing that property (3.25) holds on $g_{i,n}$ for $p=2$. By the definition (3.91) of $g_{i,n}$, we have $\|g_{i,n}\|_\infty=R(\Delta_{n,i}^{2(\beta-\frac{1}{\alpha})},X_{t_i})$ and $\|g_{i,n}\|_{pol}=R(1,X_{t_i})$. Replacing them in (3.26) we have that

$$|\mathbb{E}_i[g_{i,n}(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})]-\mathbb{E}[g_{i,n}(S_1^\alpha)]|\le c_{\epsilon,\alpha}\Delta_{n,i}|\log(\Delta_{n,i})|R(\Delta_{n,i}^{2(\beta-\frac{1}{\alpha})},X_{t_i})+$$

$$+c_{\epsilon,\alpha}\Delta_{n,i}^{\frac{1}{\alpha}}|\log(\Delta_{n,i})|R(\Delta_{n,i}^{2(\beta-\frac{1}{\alpha})(1-\frac{\alpha}{2}-\epsilon)},X_{t_i})+$$

$$+c_{\epsilon,\alpha}\Delta_{n,i}^{\frac{1}{\alpha}}|\log(\Delta_{n,i})|R(\Delta_{n,i}^{2(\beta-\frac{1}{\alpha})(\frac{3}{2}-\frac{\alpha}{2}-\epsilon)},X_{t_i})1_{\alpha>1}.$$

To get $\sum_{i=0}^{n-1}A_{1,i}^n:=o_{\mathbb{P}}(1)$, we want to use Lemma 9 of [36]. We have

$$\sum_{i=0}^{n-1}|\mathbb{E}_i[A_{1,i}^n]|\le\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})||\log(\Delta_{n,i})|[\Delta_{n,i}^{\frac{2}{\alpha}+2(\beta-\frac{1}{\alpha})}R(1,X_{t_i})+$$

$$+\Delta_{n,i}^{\frac{3}{\alpha}-1+(2-\alpha-\epsilon)(\beta-\frac{1}{\alpha})}R(1,X_{t_i})+\Delta_{n,i}^{\frac{3}{\alpha}-1+(3-\alpha-\epsilon)(\beta-\frac{1}{\alpha})}R(1,X_{t_i})1_{\alpha>1}]\le$$

$$\le(\Delta_n^{\alpha\beta}+\Delta_n^{\frac{1}{\alpha}-\epsilon})|\log(\Delta_n)|\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i})+$$

$$+\Delta_n^{\beta-\epsilon}|\log(\Delta_n)|\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i})1_{\alpha>1}, \tag{3.93}$$

where we have used property (3.24) and the monotony of the logarithmic function in order to say that $\log(\Delta_{n,i})\le\log(\Delta_n)$. Using the polynomial growth of f and R, the fifth point of Assumption 4 in order to bound γ and the third point of Lemma 26, (3.93) converges to 0 in norm 1 and so in probability since $\Delta_n^{\alpha\beta}\log(\Delta_n)\to 0$ for $n\to\infty$ and we can always find an $\epsilon>0$ such that $\Delta_n^{\frac{1}{\alpha}-\epsilon}$ does the same.
To use Lemma 9 of [36] we have also to show that

$$\Delta_n^{1-2\beta(2-\alpha)}\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}\frac{f^2(X_{t_i})}{\Delta_{n,i}^2}\gamma^4(X_{t_i})\Delta_{n,i}^{\frac{4}{\alpha}}\mathbb{E}_i[(g_{i,n}(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})-\mathbb{E}[g_{i,n}(S_1^\alpha)])^2]\xrightarrow{P}0. \tag{3.94}$$

We observe that

$$\mathbb{E}_i[(g_{i,n}(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})-\mathbb{E}[g_{i,n}(S_1^\alpha)])^2]\le c\mathbb{E}_i[g_{i,n}^2(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})]+c\mathbb{E}_i[\mathbb{E}[g_{i,n}(S_1^\alpha)]^2].$$

Now, using equation (3.32) of Lemma 28, we observe it is

$$\mathbb{E}_i[g_{i,n}^2(\frac{\Delta\tilde{X}_i^J}{\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}}})] = \frac{\Delta_{n,i}^{-\frac{4}{\alpha}}}{\gamma^4(X_{t_i})}\mathbb{E}_i[(\Delta\tilde{X}_i^J)^4\varphi^2_{\Delta_n^\beta}(\Delta\tilde{X}_i^J)] = \frac{\Delta_{n,i}^{-\frac{4}{\alpha}}}{\gamma^4(X_{t_i})}R(\Delta_{n,i}^{1+\beta(4-\alpha)}, X_{t_i}), \tag{3.95}$$

where φ acts as the indicator function. Moreover we observe that

$$\mathbb{E}[g_{i,n}(S_1^\alpha)] = \int_{\mathbb{R}} z^2\varphi(\Delta_{n,i}^{\frac{1}{\alpha}-\beta}\gamma(X_{t_i})z)f_\alpha(z)dz = d(\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}-\beta}), \tag{3.96}$$

with $f_\alpha(z)$ the density of the stable process. We now introduce the following lemma, that will be shown in the Appendix:

Lemma 29. *Suppose that Assumptions 1-4 hold. Then, for each ζ_n such that $\zeta_n \to 0$ and for each $\hat{\epsilon} > 0$,*

$$d(\zeta_n) = |\zeta_n|^{\alpha-2}c_\alpha\int_{\mathbb{R}}|u|^{1-\alpha}\varphi(u)du + o_{\mathbb{P}}(|\zeta_n|^{-\hat{\epsilon}} + |\zeta_n|^{2\alpha-2-\hat{\epsilon}}), \tag{3.97}$$

where c_α has been defined in (3.12).

Since $\frac{1}{\alpha} - \beta > 0$, $\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}-\beta}$ goes to zero for $n \to \infty$ and so we can take ζ_n as $\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}-\beta}$, getting that

$$\mathbb{E}[g_{i,n}(S_1^\alpha)] = d(\gamma(X_{t_i})\Delta_{n,i}^{\frac{1}{\alpha}-\beta}) = R(\Delta_{n,i}^{(\frac{1}{\alpha}-\beta)(\alpha-2)}, X_{t_i}). \tag{3.98}$$

Replacing (3.95) and (3.98) in the left hand side of (3.94) we get it is upper bounded by

$$\sum_{i=0}^{n-1}\mathbb{E}_i[(A_{1,i}^n)^2] = \Delta_n^{1-2\beta(2-\alpha)}\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}f^2(X_{t_i})\gamma^4(X_{t_i})\Delta_{n,i}^{\frac{4}{\alpha}-2}(R(\Delta_{n,i}^{1+\beta(4-\alpha)-\frac{4}{\alpha}}, X_{t_i})+$$

$$+R(\Delta_{n,i}^{4\beta-\frac{4}{\alpha}+2-2\alpha\beta}, X_{t_i})) \le \Delta_n^{\alpha\beta\wedge 1}\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}f^2(X_{t_i})\gamma^4(X_{t_i})R(1, X_{t_i}), \tag{3.99}$$

that converges to zero in norm 1 and so in probability in both cases T fixed and $T \to \infty$, using the polynomial growth of f and R and the fact that the exponent on Δ_n is always positive. From (3.93) and (3.99) it follows

$$\sum_{i=0}^{n-1}A_{1,i}^n = o_{\mathbb{P}}(1). \tag{3.100}$$

and so (3.72).

Proof of (3.73).
We use Proposition 22 replacing (3.28) in definition (3.6) of $\tilde{Q}_n$. Our goal is to prove that

$$\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}\frac{f(X_{t_i})}{\Delta_{n,i}}(\Delta\tilde{X}_i^J)^2\varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J) = \hat{Q}_n + o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon})\wedge(1-2\beta-\tilde{\epsilon})}).$$

On the left hand side of the equation here above we can act like we did in (3.90) - (3.92). To get (3.73) we are therefore left to show that , if $\beta > \frac{1}{4-\alpha}$, then $\sum_{i=0}^{n-1}A_{1,i}^n$ is also $o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon})\wedge(1-2\beta-\tilde{\epsilon})})$. To prove it, we want to use Lemma 9 of [36], hence we want to prove the following:

$$\frac{1}{\Delta_n^{\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon}}}\sum_{i=0}^{n-1}\mathbb{E}_i[A_{1,i}^n]\xrightarrow{\mathbb{P}}0 \quad \text{and} \tag{3.101}$$

$$\frac{1}{\Delta_n^{2(\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon})}}\sum_{i=0}^{n-1}\mathbb{E}_i[(A_{1,i}^n)^2]\xrightarrow{\mathbb{P}}0. \tag{3.102}$$

Using (3.93) we have that, if $\alpha > 1$, then the left hand side of (3.101) is in module upper bounded by

$$\frac{\Delta_n^{\beta-\epsilon}|log(\Delta_n)|}{\Delta_n^{\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon}}}\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i})=$$

$$=\Delta_n^{3\beta-\alpha\beta-\frac{1}{2}+\tilde{\epsilon}-\epsilon}|log(\Delta_n)|\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i}),$$

that goes to zero since we have chosen $\beta > \frac{1}{4-\alpha} > \frac{1}{2(3-\alpha)}$. Otherwise, if $\alpha \le 1$, then (3.93) gives us that the left hand side of (3.101) is in module upper bounded by

$$\frac{\Delta_n^{\alpha\beta}|log(\Delta_n)|}{\Delta_n^{\frac{1}{2}-2\beta+\alpha\beta-\tilde{\epsilon}}}\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i})=$$

$$=\Delta_n^{2\beta-\frac{1}{2}+\tilde{\epsilon}}|log(\Delta_n)|\frac{1}{n}\sum_{i=0}^{n-1}|f(X_{t_i})||\gamma^2(X_{t_i})|R(1,X_{t_i}),$$

that goes to zero because $\beta > \frac{1}{4-\alpha} > \frac{1}{4}$.
Using also (3.99), the left hand side of (3.102) turns out to be upper bounded by

$$\Delta_n^{-1+4\beta-2\alpha\beta+2\tilde{\epsilon}}\Delta_n^{\alpha\beta\wedge 1}\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1}f^2(X_{t_i})\gamma^4(X_{t_i})R(1,X_{t_i}),$$

that goes to zero in norm 1 and so in probability since we have chosen $\beta > \frac{1}{4-\alpha}$. It follows (3.102) and so (3.10).

□

3.6.3 Proof of Proposition 20

Proof. To prove the proposition we replace (3.97) in the definition of $\hat{Q}_n$. It follows that our goal is to show that

$$I_1^n + I_2^n :=$$

$$=\frac{1}{n\Delta_n^{\beta(2-\alpha)}}\sum_{i=0}^{n-1}f(X_{t_i})\gamma^2(X_{t_i})\Delta_{n,i}^{\frac{2}{\alpha}-1}(o_{\mathbb{P}}(|\Delta_{n,i}^{\frac{1}{\alpha}-\beta}\gamma(X_{t_i})|^{-\hat{\epsilon}}+|\Delta_{n,i}^{\frac{1}{\alpha}-\beta}\gamma(X_{t_i})|^{2\alpha-2-\hat{\epsilon}}))=\tilde{\mathcal{E}}_n,$$

where $\tilde{\mathcal{E}}_n$ is always $o_{\mathbb{P}}(1)$ and, if $\alpha < \frac{4}{3}$, it is also $\frac{1}{\Delta_n^{\beta(2-\alpha)}}o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$.
We have that $I_1^n = o_{\mathbb{P}}(1)$ since it is upper bounded by

$$\Delta_n^{\frac{2}{\alpha}-1-2\beta+\alpha\beta-\hat{\epsilon}(\frac{1}{\alpha}-\beta)}\frac{1}{n}\sum_{i=0}^{n-1}R(1,X_{t_i})o_{\mathbb{P}}(1),$$

that goes to zero in norm 1 and so in probability since we can always find an $\hat{\epsilon} > 0$ such that the exponent on Δ_n is positive.
Also I_2^n is $o_{\mathbb{P}}(1)$. Indeed it is upper bounded by

$$\Delta_n^{\frac{2}{\alpha}-1-2\beta+\alpha\beta-2(\frac{1}{\alpha}-\beta)+2(1-\alpha\beta)-\hat{\epsilon}(\frac{1}{\alpha}-\beta)} \frac{1}{n}\sum_{i=0}^{n-1} R(1, X_{t_i}) o_{\mathbb{P}}(1). \tag{3.103}$$

We observe that the exponent on Δ_n is $1-\alpha\beta-\hat{\epsilon}(\frac{1}{\alpha}-\beta)$ and we can always find $\hat{\epsilon}$ such that it is more than zero, hence (3.103) converges in norm 1 and so in probability.
In order to show that $I_1^n = \frac{1}{\Delta_n^{\beta(2-\alpha)}} o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}) = o_{\mathbb{P}}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}-\beta(2-\alpha)})$ we observe that

$$\frac{I_1^n}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}-\beta(2-\alpha)}} \le \Delta_n^{\frac{2}{\alpha}-1-\frac{1}{2}+\tilde{\epsilon}-\hat{\epsilon}(\frac{1}{\alpha}-\beta)} \frac{1}{n}\sum_{i=0}^{n-1} R(1, X_{t_i}) o_{\mathbb{P}}(1).$$

If $\alpha < \frac{4}{3}$ we can always find $\tilde{\epsilon}$ and $\hat{\epsilon}$ such that the exponent on Δ_n is more than zero, getting the convergence wanted. It follows $I_1^n = \frac{1}{\Delta_n^{\beta(2-\alpha)}} o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$.
To conclude, $I_2^n = \frac{1}{\Delta_n^{\beta(2-\alpha)}} o_{\mathbb{P}}(\Delta_n^{1-\alpha\beta-\tilde{\epsilon}}) = o_{\mathbb{P}}(\Delta_n^{1-2\beta-\tilde{\epsilon}})$. Indeed,

$$\frac{I_2^n}{\Delta_n^{1-2\beta-\tilde{\epsilon}}} \le \Delta_n^{\frac{2}{\alpha}-1-1+\alpha\beta+\tilde{\epsilon}-2(\frac{1}{\alpha}-\beta)+2(1-\alpha\beta)-\hat{\epsilon}(\frac{1}{\alpha}-\beta)} \frac{1}{n}\sum_{i=0}^{n-1} R(1, X_{t_i}) o_{\mathbb{P}}(1). \tag{3.104}$$

The exponent on Δ_n is $2\beta - \alpha\beta + \tilde{\epsilon} - \hat{\epsilon}(\frac{1}{\alpha} - \beta)$ and so we can always find $\tilde{\epsilon}$ and $\hat{\epsilon}$ such that it is positive. It follows the convergence in norm 1 and so in probability of (3.104). The proposition is therefore proved. □

3.6.4 Proof of Corollary 1

Proof. We observe that (3.13) is a consequence of (3.11) in the case where $\hat{Q}_n = 0$. Moreover, $\beta < \frac{1}{2\alpha}$ implies that $\Delta_n^{1-\alpha\beta-\tilde{\epsilon}}$ is negligible compared to $\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}$. It follows (3.13). □

3.6.5 Proof of Theorem 12.

Proof. The convergence (3.14) clearly follows from (3.11) and the second point of Assumption S1 with $\delta = \beta(2-\alpha)$.
Concerning the proof of (3.15), we can see its left hand side as

$$Q_n - \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) + \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) - IV_1$$

and so, using (3.10), it turns out that our goal is to show that

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) - IV_1 = o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)}). \tag{3.105}$$

It is

$$\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) - \frac{1}{T}\int_0^T (fa^2)(X_s)ds =$$

$$= \frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i}) - \frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}}\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}(fa^2)(X_s)ds.$$

We now act as we did in (3.74), considering this time the development up to second order of the function fa^2 instead of a^2. Replacing it in the equation here above we get

$$\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i})(\frac{1}{n} - \frac{\Delta_{n,i}}{\sum_{i=0}^{n-1}\Delta_{n,i}}) - \frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}}\sum_{i=0}^{n-1}(fa^2)'(X_{t_i})\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})ds +$$

$$-\frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}}\sum_{i=0}^{n-1}\frac{1}{2}\int_{t_i}^{t_{i+1}}(fa^2)''(\tilde{X}_{t_i})(X_s - X_{t_i})^2 ds =: I_1^n + I_2^n + I_3^n,$$

where $\tilde{X}_{t_i} \in [X_{t_i}, X_s]$. Now, using the third point of the Assumption S1, we have

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}}\mathbb{E}[|I_1^n|] \le \Delta_n^{\delta_0}\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}[(fa^2)(X_{t_i})], \tag{3.106}$$

that goes to zero because of the polynomial growth of both f and a and the third point of Lemma 26.

Concerning I_2^n, we act like we did in the proof of Theorem 11 to get that B_1^n defined below equation (3.74) was $o_{\mathbb{P}}(\Delta_n^{\beta(2-\alpha)})$. We have used the dynamic of the process X to get $\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})ds|] \le R(\Delta_{n,i}^2, X_{t_i})$. We observe moreover that, as a consequence of the third point of Assumption S1, we have

$$\frac{\Delta_{n,i}}{\sum_{i=0}^{n-1}\Delta_{n,i}} \le \frac{1}{n} + \frac{\Delta_n^{\beta(2-\alpha)+\delta_0}}{n} \le \frac{c}{n}. \tag{3.107}$$

It follows

$$\frac{\mathbb{E}[|I_2^n|]}{\Delta_n^{\beta(2-\alpha)}} \le \Delta_n^{1-\beta(2-\alpha)}\sum_{i=0}^{n-1}\mathbb{E}[(fa^2)'(X_{t_i})\frac{\Delta_{n,i}}{\sum_{i=0}^{n-1}\Delta_{n,i}}R(1, X_{t_i})] \le \tag{3.108}$$

$$\le \Delta_n^{1-\beta(2-\alpha)}\frac{c}{n}\sum_{i=0}^{n-1}\mathbb{E}[(fa^2)'(X_{t_i})R(1, X_{t_i})],$$

that goes to zero since $1-\beta(2-\alpha)$ is always more than 0.

Also on I_3^n we act like we did on B_2^n in the proof of theorem 11 to get

$$\mathbb{E}_i[|\int_{t_i}^{t_{i+1}}(fa^2)''(\tilde{X}_{t_i})(X_s - X_{t_i})^2 ds|] \le R(\Delta_{n,i}^{2-\epsilon}, X_{t_i}),$$

(see above equation (3.75)). Using also (3.107) it follows

$$\frac{1}{\Delta_n^{\beta(2-\alpha)}}\mathbb{E}[|I_3^n|] \le \Delta_n^{1-\beta(2-\alpha)-\epsilon}\frac{c}{n}\sum_{i=0}^{n-1}\mathbb{E}[(fa^2)''(X_{t_i})R(1, X_{t_i})]. \tag{3.109}$$

Again, it goes to zero in norm 1 and so in probability. From (3.106), (3.108) and (3.109) it follows (3.105) and so the theorem is proved. □

3.6.6 Proof of Theorem 13.

Proof. The convergence (3.16) is a consequence of Lemma 27, that we can apply since we have assumed that points 1 and 2 of Assumption S2 hold.
Concerning the proof of (3.17), we can again add and subtract $\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})a^2(X_{t_i})$ and so our goal is to show (3.105), with T that now goes to ∞ for $n \to \infty$. We observe that we can act like we did in the previous theorem because, having assumed the third point of the Assumption S2, the proof here above still hold. □

3.7 Proof of developments in small time: Proposition 21.

This section is dedicate to the proof of Proposition 21. Proposition 22 will be proved in the appendix.
To prove Proposition 21, it is convenient to introduce an adequate truncation function and to consider a rescaled process, as explained in the next subsections. Moreover, the proof of Proposition 21 requires some Malliavin calculus; we recall in what follows all the technical tools to make easier the understanding of the chapter.

3.7.1 Localization and rescaling

We introduce a truncation function in order to suppress the big jumps of (L_t). Let $\tau : \mathbb{R} \to [0,1]$ be a symmetric function, continuous with continuous derivative, such that $\tau = 1$ on $\left\{|z| \le \frac{1}{4}\eta\right\}$ and $\tau = 0$ on $\left\{|z| \ge \frac{1}{2}\eta\right\}$, with η defined in the fourth point of Assumption 4.
On the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ we consider the Lévy process (L_t) defined below (4.3) which measure is $F(dz) = \frac{g(z)}{|z|^{1+\alpha}}1_{\mathbb{R}\setminus\{0\}}(z)dz$, according with the third point of Assumption 4, and the truncated Lévy process (L_t^τ) with measure $F^\tau(dz)$ given by $F^\tau(dz) = \frac{g(z)\tau(z)}{|z|^{1+\alpha}}1_{\mathbb{R}\setminus\{0\}}(z)dz$. This can be done by setting $L_t := \int_0^t\int_{\mathbb{R}} z\tilde{\mu}(ds,dz)$, as we have already done, and $L_t^\tau := \int_0^t\int_{\mathbb{R}} z\tilde{\mu}^\tau(ds,dz)$, where $\tilde{\mu}$ and $\tilde{\mu}^\tau$ are the compensated Poisson random measures associated respectively to

$$\mu(A) := \int_{[0,1]}\int_{\mathbb{R}}\int_{[0,T]} 1_A(t,z)\mu^g(dt,dz,du), \qquad A \subset [0,T]\times\mathbb{R},$$

$$\mu^\tau(A) := \int_{[0,1]}\int_{\mathbb{R}}\int_{[0,T]} 1_A(t,z)1_{u\le\tau(z)}\mu^g(dt,dz,du), \qquad A \subset [0,T]\times\mathbb{R},$$

for μ^g a Poisson random measure on $[0,T]\times\mathbb{R}\times[0,1]$ with compensator $\bar{\mu}^g(dt,dz,du) = dt\frac{g(z)}{|z|^{1+\alpha}}1_{\mathbb{R}\setminus\{0\}}(z)dzdu$.
By construction, the restrictions of the measures μ and μ^τ to $[0,\Delta_n]\times\mathbb{R}$ coincide on the set
$\{(u,z)$ such that $u \le \tau(z)\}$, and thus coincide on the event

$$\Omega_n := \left\{\omega \in \Omega; \mu^g([0,\Delta_n]\times\left\{z\in\mathbb{R} : |z| \ge \frac{\eta}{4}\right\}\times[0,1]) = 0\right\}.$$

Since $\mu^g([0,\Delta_n]\times\left\{z\in\mathbb{R} : |z| \ge \frac{\eta}{4}\right\}\times[0,1])$ has a Poisson distribution with parameter

$$\lambda_n := \int_0^{\Delta_n}\int_{|z|\ge\frac{\eta}{4}}\int_0^1 \frac{g(z)}{|z|^{1+\alpha}}du\,dz\,dt \le c\Delta_n;$$

we deduce that

$$\mathbb{P}(\Omega_n^c) \leq c\Delta_n. \tag{3.110}$$

Then we have

$$\mathbb{P}((L_t)_{t\leq\Delta_n} \neq (L_t^\tau)_{t\leq\Delta_n}) \leq \mathbb{P}(\Omega_n^c) \leq c\Delta_n. \tag{3.111}$$

To prove Proposition 21 we have to rescale the process $(L_t)_{t\in[0,1]}$, we therefore introduce an auxiliary Lévy process $(L_t^n)_{t\in[0,1]}$ defined possibly on another filtered space $(\tilde{\Omega}, \tilde{\mathcal{F}}, (\tilde{\mathcal{F}}_t), \tilde{\mathbb{P}})$ and admitting the decomposition $L_t^n := \int_0^t \int_{\mathbb{R}} z\tilde{\mu}^n(dt, dz)$, with $t \in [0,1]$; where $\tilde{\mu}^n$ is a compensated Poisson random measure $\tilde{\mu}^n = \mu^n - \bar{\mu}^n$, with compensator

$$\bar{\mu}^n(dt, dz) = dt\frac{g(z\Delta_n^{\frac{1}{\alpha}})}{|z|^{1+\alpha}}\tau(z\Delta_n^{\frac{1}{\alpha}})1_{\mathbb{R}\setminus\{0\}}(z)dz. \tag{3.112}$$

By construction, the process $(L_t^n)_{t\in[0,1]}$ is equal in law to the rescaled truncated process $(\Delta_n^{-\frac{1}{\alpha}}L_{\Delta_n t}^\tau)_{t\in[0,1]}$ that coincides with $(\Delta_n^{-\frac{1}{\alpha}}L_{\Delta_n t})_{t\in[0,1]}$ on Ω_n.

3.7.2 Malliavin calculus

In this section, we recall some results on Malliavin calculus for jump processes. We refer to [14] for a complete presentation and to [19] for the adaptation to our framework. We will work on the Poisson space associated to the measure μ^n defining the process $(L_t^n)_{t\in[0,1]}$ of the previous section, assuming that n is fixed. By construction, the support of μ^n is contained in $[0,1] \times E_n$, where $E_n := \left\{ z \in \mathbb{R} : |z| < \frac{\eta}{2}\frac{1}{\Delta_n^{\frac{1}{\alpha}}} \right\}$, with η defined in the fourth point of Assumption 4. We recall that the measure μ^n has compensator

$$\bar{\mu}^n(dt, dz) = dt\frac{g(z\Delta_n^{\frac{1}{\alpha}})}{|z|^{1+\alpha}}\tau(z\Delta_n^{\frac{1}{\alpha}})1_{\mathbb{R}\setminus\{0\}}(z)dz := dtF_n(z)dz. \tag{3.113}$$

In this section we assume that the truncation function τ satisfies the additional assumption

$$\int_{\mathbb{R}} |\frac{\tau'(z)}{\tau(z)}|^p \tau(z)dz < \infty, \qquad \forall p \geq 1.$$

We now define the Malliavin operators L and Γ (omitting their dependence in n) and their basic properties (see [14] Chapter IV, sections 8-9-10). For a test function $f : [0,1]\times\mathbb{R} \to \mathbb{R}$ measurable, $\mathcal{C}^2$ with respect the second variable, with bounded derivative and such that $f \in \cap_{p\geq1} L^p(\bar{\mu}^n(dt, dz))$, we set $\mu^n(f) = \int_0^1 \int_{\mathbb{R}} f(t,z)\mu^n(dt, dz)$. As auxiliary function, we consider $\rho : \mathbb{R} \to [0, \infty)$ such that ρ is symmetric, two times differentiable and such that $\rho(z) = z^4$ if $z \in [0, \frac{1}{2}]$ and $\rho(z) = z^2$ if $z \geq 1$. Thanks to the truncation τ, we do not need that ρ vanishes at infinity. Assuming the fourth point of Assumption 4, we check that ρ, ρ' and $\rho\frac{F_n'}{F_n}$ belong to $\cap_{p\geq1} L^p(F_n(z)dz)$. With these notations, we define the Malliavin operator L on the functional $\mu^n(f)$ as follows:

$$L(\mu^n(f)) := \frac{1}{2}\mu^n(\rho' f' + \rho\frac{F_n'}{F_n}f' + \rho f''),$$

where f' and f'' are derivative with respect to the second variable. This definition permits to construct a linear operator on the space $D \subset \cap_{p\geq1} L^p(F_n(z)dz)$ which is self-adjoint: $\forall \Phi, \Psi \in D$, $\mathbb{E}\Phi L\Psi = \mathbb{E}L\Phi\Psi$ (see Section 8 in [14] for the details on the

construction of D).
We associate to L the symmetric bilinear operator Γ:

$$\Gamma(\Phi,\Psi)=L(\Phi,\Psi)-\Phi L(\Psi)-\Psi L(\Phi).$$

If f and h are two test functions, we have

$$\Gamma(\mu^n(f),\mu^n(h))=\mu^n(\rho f'h'). \tag{3.114}$$

The operators L and Γ satisfy the chain rule property:

$$LF(\Phi)=F'(\Phi)L\Phi+\frac{1}{2}F''(\Phi)\Gamma(\Phi,\Phi), \qquad \Gamma(F(\Phi),\Psi)=F'(\Phi)\Gamma(\Phi,\Psi).$$

These operators permit to establish the following integration by parts formula (see [14] Theorem 8-10 p.103).

Theorem 14. *Let Φ and Ψ be random variable in D and f be a bounded function with bounded derivatives up to order two. If $\Gamma(\Phi,\Phi)$ is invertible and $\Gamma^{-1}(\Phi,\Phi)\in\cap_{p\geq 1}L^p$, then we have*

$$\mathbb{E}f'(\Phi)\Psi=\mathbb{E}f(\Phi)\mathcal{H}_\Phi(\Psi), \tag{3.115}$$

with

$$\mathcal{H}_\Phi(\Psi)=-2\Psi\Gamma^{-1}(\Phi,\Phi)L\Phi-\Gamma(\Phi,\Psi\Gamma^{-1}(\Phi,\Phi)). \tag{3.116}$$

The random variable L_1^n belongs to the domain of the operators L and Γ. Computing $L(L_1^n)$, $\Gamma(L_1^n,L_1^n)$ and $\mathcal{H}_{L_1^n}(1)$ it is possible to deduce the following useful inequalities, proved in Lemma 4.3 in [19].

Lemma 30. *We have*

$$\sup_n \mathbb{E}|\mathcal{H}_{L_1^n}(1)|^p\leq C_p \qquad \forall p\geq 1,$$

$$\sup_n \mathbb{E}|\int_0^1\int_{|z|>1}|z|\mu^n(ds,dz)\mathcal{H}_{L_1^n}(1)|^p\leq C_p \qquad \forall p\geq 1.$$

With this background we can proceed to the proof of Proposition 21.

3.7.3 Proof of Proposition 21

Proof. The first step is to construct on the same probability space two random variables whose laws are close to the laws of $\Delta_n^{-\frac{1}{\alpha}}L_{\Delta_n}$ and S_1^α. We recall briefly the notation of Section 3.7.1: μ^n is a Poisson random measure with compensator $\bar{\mu}^n(dt,dz)$ defined in (3.112) and the process L_t^n is defined by

$$L_t^n=\int_0^t\int_{\mathbb{R}}z\tilde{\mu}^n(ds,dz)=\int_0^t\int_{|z|\leq\Delta_n^{-\frac{1}{\alpha}}\frac{\eta}{2}}z\tilde{\mu}^n(ds,dz) \tag{3.117}$$

with $\tilde{\mu}^n=\mu^n-\bar{\mu}^n$. Using triangle inequality we have

$$|\mathbb{E}[h(\Delta_n^{-\frac{1}{\alpha}}L_{\Delta_n})]-\mathbb{E}[h(S_1^\alpha)]|\leq|\mathbb{E}[h(\Delta_n^{-\frac{1}{\alpha}}L_{\Delta_n})]-\mathbb{E}[h(L_1^n)]|+|\mathbb{E}[h(L_1^n)-h(S_1^\alpha)]|. \tag{3.118}$$

By the definition of L_1^n it is

$$|\mathbb{E}[h(\Delta_n^{-\frac{1}{\alpha}} L_{\Delta_n})] - \mathbb{E}[h(L_1^n)]| = |\mathbb{E}[h(\Delta_n^{-\frac{1}{\alpha}} L_{\Delta_n}) - h(\Delta_n^{-\frac{1}{\alpha}} L_{\Delta_n}^\tau)]| \leq \tag{3.119}$$

$$\leq 2 \left\| h \right\|_\infty \mathbb{P}(\Omega_n^c) \leq c \left\| h \right\|_\infty \Delta_n,$$

where in the last inequality we have used (3.111). In order to get an estimation to the second term of (3.118) we now construct a variable approximating the law of S_1^α and based on the Poisson measure μ^n :

$$L_t^{\alpha,n} := \int_0^t \int_{|z| \leq \Delta_n^{-\frac{1}{\alpha}} \frac{\eta}{2}} h_n(z) \tilde{\mu}^n(ds, dz), \tag{3.120}$$

where h_n is an odd function built in the proof of Theorem 4.1 in [19] for which the following lemma holds:

Lemma 31. *1. For each test function f, defined as in Section 3.7.2, we have*

$$\int_0^1 \int_{|z| \leq \frac{\eta}{2} \Delta_n^{-\frac{1}{\alpha}}} f(t, h_n(z)) \bar{\mu}^n(dt, dz) = \int_0^1 \int_{|\omega| \leq \frac{\eta}{2} \Delta_n^{-\frac{1}{\alpha}}} f(t, \omega) \bar{\mu}^{\alpha,n}(dt, d\omega), \tag{3.121}$$

where $\bar{\mu}^n(dt, dz)$ is the compensator defined in (3.112) and

$$\bar{\mu}^{\alpha,n}(dt, d\omega) = dt \frac{\tau(\omega \Delta_n^{\frac{1}{\alpha}})}{|\omega|^{1+\alpha}} d\omega$$

is the compensator of a measure associated to an α- stable process whose jumps are truncated with the function τ.

2. There exists $\epsilon_0 > 0$ such that, for $|z| \leq \epsilon_0 \Delta_n^{-\frac{1}{\alpha}}$,

$$|h_n(z) - z| \leq c z^2 \Delta_n^{\frac{1}{\alpha}} + c|z|^{1+\alpha} \Delta_n \qquad \text{if } \alpha \neq 1,$$

$$|h_n(z) - z| \leq c z^2 \Delta_n |log(|z| \Delta_n)| \qquad \text{if } \alpha = 1.$$

3. The function h_n is $\mathcal{C}^1$ on $(-\epsilon_0 \Delta_n^{-\frac{1}{\alpha}}, \epsilon_0 \Delta_n^{-\frac{1}{\alpha}})$ and for $|z| < \epsilon_0 \Delta_n^{-\frac{1}{\alpha}}$,

$$|h_n'(z) - 1| \leq c|z| \Delta_n^{\frac{1}{\alpha}} + c|z|^\alpha \Delta_n \qquad \text{if } \alpha \neq 1,$$

$$|h_n'(z) - 1| \leq c|z| \Delta_n |log(|z| \Delta_n)| \qquad \text{if } \alpha = 1.$$

The second and the third point of the lemma here above are proved in Lemma 4.5 of [19], while the first point is proved in Theorem 4.1 [19] and it shows us, using the exponential formula for Poisson measure, that h_n is the function that turns our measure μ^n into the measure associated to an α-stable process truncated with the function τ. Thus $(L_t^{\alpha,n})_{t \in [0,1]}$ is a Lévy process with jump intensity $\omega \mapsto \frac{\tau(\omega \Delta_n^{\frac{1}{\alpha}})}{|\omega|^{1+\alpha}}$ and we recognize the law of an α-stable truncated process. We deduce, similarly to (3.119),

$$|\mathbb{E}[h(L_1^{\alpha,n})] - \mathbb{E}[h(S_1^\alpha)]| \leq c \left\| h \right\|_\infty \Delta_n. \tag{3.122}$$

Proposition 21 is a consequence of (3.118), (3.119), (3.122) and the following lemma:

Lemma 32. *Suppose that Assumptions 1 to 4 hold. Let h be as in Proposition 21. Then, for any $\epsilon > 0$ and for $p \geq \alpha$,*

$$|\mathbb{E}[h(L_1^n) - h(L_1^{\alpha,n})]| \leq C_\epsilon \Delta_n |log(\Delta_n)| \, \|h\|_\infty + C_\epsilon \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1-\frac{\alpha}{p}+\epsilon} \|h\|_{pol}^{\frac{\alpha}{p}-\epsilon} |log(\Delta_n)| +$$

$$+ C_\epsilon \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p}+\epsilon} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p}-\epsilon} |log(\Delta_n)| 1_{\alpha>1}.$$

Proof. The proof is based of the comparison of the representation of (3.117) and (3.120). Since in Lemma 31 the difference $h_n(z) - z$ is controlled for $|z| \leq \epsilon_0 \Delta_n^{-\frac{1}{\alpha}}$, we need to introduce a localization procedure consisting in regularizing $1_{\left\{\mu^n([0,1]\times\left\{z\in\mathbb{R}:|z|>\epsilon_0\Delta_n^{-\frac{1}{\alpha}}\right\})=0\right\}}$. Let $\mathcal{I}$ be a smooth function defined on $\mathbb{R}$ and with values in $[0,1]$, such that $\mathcal{I}(x) = 1$ for $x \leq \frac{1}{2}$ and $\mathcal{I}(x) = 0$ for $x \geq 1$. Moreover, we denote by ζ a smooth function on $\mathbb{R}$, with values in $[0,1]$ such that $\zeta(z) = 0$ for $|z| \leq \frac{1}{2}$ and $\zeta(z) = 1$ for $|z| \geq 1$ and we set

$$V^n := \int_0^1 \int_{\mathbb{R}} \zeta(\frac{z\Delta_n^{\frac{1}{\alpha}}}{\epsilon_0}) \mu^n(ds,dz) =$$

$$= \int_0^1 \int_{\left\{\frac{1}{2}\epsilon_0\Delta_n^{-\frac{1}{\alpha}} \leq |z| \leq \epsilon_0\Delta_n^{-\frac{1}{\alpha}}\right\}} \zeta(\frac{z\Delta_n^{\frac{1}{\alpha}}}{\epsilon_0}) \mu^n(ds,dz) + \int_0^1 \int_{\left\{|z|\geq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}\right\}} \mu^n(ds,dz),$$

$$W^n := \mathcal{I}(V^n).$$

From the construction, W^n is a Malliavin differentiable random variable such that $W^n \neq 0$ implies $\mu^n([0,1]\times\left\{z \in \mathbb{R} : |z| > \epsilon_0\Delta_n^{-\frac{1}{\alpha}}\right\}) = 0$. It is possible to show, acting as we did in (3.110), that $\mathbb{P}(W^n \neq 1) \leq \mathbb{P}(\mu^n \text{ has a jump of size } > \frac{1}{2}\epsilon_0\Delta_n^{-\frac{1}{\alpha}}) = O(\Delta_n)$. From the latter, it is clear that the proof of the lemma reduces in proving the result on $|\mathbb{E}[h(L_1^n)W^n] - \mathbb{E}[h(L_1^{\alpha,n})W^n]|$. Considering a regularizing sequence (h_p) converging to h in L^1 norm, such that $\forall p$ h_p is $\mathcal{C}^1$ with bounded derivative and $\|h_p\|_\infty \leq \|h\|_\infty$, we may assume that h is $\mathcal{C}^1$ with bounded derivative too. Using the integration by part formula (3.115) and denoting by H any primitive function of h we can write $\mathbb{E}[h(L_1^n)W^n] = \mathbb{E}[H(L_1^n)\mathcal{H}_{L_1^n}(W^n)]$ where the Malliavin weight can be written, using (3.116) and the chain rule property of the operator Γ, as

$$\mathcal{H}_{L_1^n}(W^n) = W^n \mathcal{H}_{L_1^n}(1) - \frac{\Gamma(W^n, L_1^n)}{\Gamma(L_1^n, L_1^n)}. \tag{3.123}$$

Using the triangle inequality, we are now left to find upper bounds for the following two terms:

$$\tilde{T}_1 := |\mathbb{E}[h(L_1^{\alpha,n})W^n] - \mathbb{E}[H(L_1^{\alpha,n})\mathcal{H}_{L_1^n}(W^n)]|,$$

$$\tilde{T}_2 := |\mathbb{E}[H(L_1^{\alpha,n})\mathcal{H}_{L_1^n}(W^n)] - \mathbb{E}[H(L_1^n)\mathcal{H}_{L_1^n}(W^n)]|.$$

Let us start considering $\tilde{T}_2$. Using the Lipschitz property of the function H and (3.123) we have it is upper bounded by

$$\mathbb{E}[|h(\hat{L}_1)||L_1^{\alpha,n} - L_1^n||\mathcal{H}_{L_1^n}(W^n)|] \leq$$

$$\leq \mathbb{E}[|h(\hat{L}_1)||L_1^{\alpha,n} - L_1^n||W^n \mathcal{H}_{L_1^n}(1)|] + \mathbb{E}[|h(\hat{L}_1)||L_1^{\alpha,n} - L_1^n||\frac{\Gamma(W^n, L_1^n)}{\Gamma(L_1^n, L_1^n)}|] =$$

$$=: \tilde{T}_{2,1} + \tilde{T}_{2,2},$$

where $\hat{L}_1$ is between $L_1^{\alpha,n}$ and L_1^n. We focus on $\tilde{T}_{2,1}$. Using the definitions (3.117) and (3.120) of L_1^n and $L_1^{\alpha,n}$ it is

$$\tilde{T}_{2,1} \leq \mathbb{E}[|h(\hat{L}_1)|| \int_0^1 \int_{\mathbb{R}} (h_n(z) - z)\tilde{\mu}^n(ds, dz)||\mathcal{H}_{L_1^n}(1) W^n|] \leq$$

$$\leq \mathbb{E}[|h(\hat{L}_1)|| \int_0^1 \int_{|z|\leq 1} (h_n(z) - z)\tilde{\mu}^n(ds, dz)||\mathcal{H}_{L_1^n}(1) W^n|] +$$

$$+ \mathbb{E}[|h(\hat{L}_1)|| \int_0^1 \int_{1\leq |z| \leq \epsilon_0 \Delta_n^{-\frac{1}{\alpha}}} (h_n(z) - z)\mu^n(ds, dz)||\mathcal{H}_{L_1^n}(1) W^n|], \tag{3.124}$$

where we have used that h_n is an odd function with the symmetry of the compensator $\bar{\mu}^n$ and the fact that on $W_n \neq 0$ we have $\mu^n([0,1] \times \left\{ z \in \mathbb{R} : |z| > \epsilon_0 \Delta_n^{-\frac{1}{\alpha}} \right\}) = 0$. For the sake of shortness, we only give the details of the proof in the case $\alpha \neq 1$. In the case $\alpha = 1$, one needs to modify this control with an additional logarithmic term. For the small jumps term, from inequality 2.1.37 in [46] and the second point of Lemma 31 we deduce $\mathbb{E}[| \int_0^1 \int_{|z|\leq 1} (h_n(z) - z)\tilde{\mu}^n(ds, dz)|^{q_1}] \leq C_{q_1}(\Delta_n + \Delta_n^{\frac{1}{\alpha}})^{q_1}$, $\forall q_1 \geq 2$. Using it and Holder inequality with q_1 big and q_2 close to 1 we have

$$\mathbb{E}[|h(\hat{L}_1)|| \int_0^1 \int_{|z|\leq 1} (h_n(z) - z)\tilde{\mu}^n(ds, dz)||\mathcal{H}_{L_1^n}(1) W^n|] \leq$$

$$\leq C_{q_1}(\Delta_n + \Delta_n^{\frac{1}{\alpha}}) \mathbb{E}[|h(\hat{L}_1)|^{q_2} |\mathcal{H}_{L_1^n}(1)|^{q_2} W^n]^{\frac{1}{q_2}} \leq$$

$$\leq C_{q_1}(\Delta_n + \Delta_n^{\frac{1}{\alpha}}) \mathbb{E}[|h(\hat{L}_1)|^{p_1 q_2} W^n]^{\frac{1}{q_2 p_1}} \mathbb{E}[|\mathcal{H}_{L_1^n}(1)|^{q_2 p_2}]^{\frac{1}{q_2 p_2}}, \tag{3.125}$$

where in the last inequality we have used again Holder inequality, with p_2 big and p_1 close to 1. Using the first point of Lemma 30, we know that $\mathbb{E}[|\mathcal{H}_{L_1^n}(1)|^{q_2 p_2}]^{\frac{1}{q_2 p_2}}$ is bounded, hence (3.125) is upper bounded by

$$C_{q_1 q_2 p_2} \Delta_n \|h\|_\infty + C_{q_1 q_2 p_2} \Delta_n^{\frac{1}{\alpha}} \mathbb{E}[|h(\hat{L}_1) W^n|^{p_1 q_2}]^{\frac{1}{q_2 p_1}}, \tag{3.126}$$

where we have bounded $|h(\hat{L}_1)|$ with its infinity norm and used that $0 \leq W^n \leq 1$. We remind that we are considering q_2 and p_1 next to 1, hence we can write $q_2 p_1$ as $1 + \epsilon$. We now introduce r in the following way:

$$\mathbb{E}[|h(\hat{L}_1)|^{1+\epsilon} W^n]^{\frac{1}{1+\epsilon}} = \mathbb{E}[|h(\hat{L}_1)|^{(1+\epsilon)r} |h(\hat{L}_1)|^{(1+\epsilon)(1-r)} W^n]^{\frac{1}{1+\epsilon}} \leq$$

$$\leq \|h\|_\infty^r \mathbb{E}[|h(\hat{L}_1)|^{(1+\epsilon)(1-r)} W^n]^{\frac{1}{1+\epsilon}} \leq \|h\|_\infty^r \|h\|_{pol}^{1-r} \mathbb{E}[(1 + |\hat{L}_1|^p)^{(1+\epsilon)(1-r)} W^n]^{\frac{1}{1+\epsilon}} \leq$$

$$\leq c \|h\|_\infty^r \|h\|_{pol}^{1-r} + c \|h\|_\infty^r \|h\|_{pol}^{1-r} \mathbb{E}[|\hat{L}_1|^{p(1+\epsilon)(1-r)} W^n]^{\frac{1}{1+\epsilon}}; \tag{3.127}$$

where we have estimated h with its norm ∞ and we have used the property (3.25) of h and that $0 \leq W^n \leq 1$. We observe that $\hat{L}_1$ is between L_1^n and $L_1^{\alpha,n}$ hence $|\hat{L}_1| \leq |L_1^n| + |L_1^{\alpha,n}|$. Moreover we choose r such that $p(1+\epsilon)(1-r) = \alpha$; therefore $r = 1 - \frac{\alpha}{p(1+\epsilon)}$. In this way we have that (3.127) is upper bounded by

$$c \|h\|_\infty^{1 - \frac{\alpha}{p(1+\epsilon)}} \|h\|_{pol}^{\frac{\alpha}{p(1+\epsilon)}} \frac{1}{1+\epsilon} log(\Delta_n^{-\frac{1}{\alpha}}). \tag{3.128}$$

Indeed, remarking that as a consequence of the second point of Lemma 31 there exists $c > 0$ such that $|h_n(z)| \le c|z|$, we can act on both L_1^n and $L_1^{\alpha,n}$ in the same way. Using also Lemma 2.1.5 in the appendix of [46] if $\alpha \in [1,2]$ and Jensen inequality if $\alpha \in [0,1)$, we have

$$\mathbb{E}[|\hat{L}_1|^\alpha W^n] \le c\mathbb{E}[(|L_1^n|^\alpha + |L_1^{\alpha,n}|^\alpha)W^n] \le$$

$$\le c\mathbb{E}[|\int_0^1\int_{|z|\le 1} z\tilde{\mu}^n(ds,dz)|] + c\mathbb{E}[|\int_0^1\int_{|z|\le 1} h_n(z)\tilde{\mu}^n(ds,dz)|] +$$

$$+ c\mathbb{E}[\int_0^1\int_{1\le|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|^\alpha\bar{\mu}^n(ds,dz)] + c\mathbb{E}[\int_0^1\int_{1\le|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |h_n(z)|^\alpha\bar{\mu}^n(ds,dz)].$$

We observe that, using Kunita inequality, the first term here above is bounded in L^p and, as a consequence of the second point of Lemma 31, the second term here above so does. Concerning the third term here above (and so, again, we act on the fourth in the same way), we have

$$c\mathbb{E}[\int_0^1\int_{1\le|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|^\alpha\bar{\mu}^n(ds,dz)] \le \tag{3.129}$$

$$\le c\int_{1\le|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|^{\alpha-1-\alpha}dz \le c\,log(\Delta_n^{-\frac{1}{\alpha}}) \le c|log(\Delta_n)|,$$

where we have also used definition (3.112) of $\bar{\mu}^n$. Replacing (3.128) in (3.126) we get

$$\mathbb{E}[|h(\hat{L}_1)||\int_0^1\int_{|z|\le 1}(h_n(z)-z)\tilde{\mu}^n(ds,dz)||\mathcal{H}_{L_1^n}(1)W^n|] \le \tag{3.130}$$

$$\le C_{q_1q_2p_2}\Delta_n\|h\|_\infty + C_{q_1q_2p_2}\Delta_n^{\frac{1}{\alpha}}\|h\|_\infty^{1-\frac{\alpha}{p}+\epsilon}\|h\|_{pol}^{\frac{\alpha}{p}-\epsilon}log(\Delta_n^{-\frac{1}{\alpha}}),$$

where we have taken another ϵ, using its arbitrariness. The constants depend also on it.

Let us now consider the large jumps term in (3.124). Using the second point of Lemma 31 and the following basic inequality

$$\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|^\delta\mu^n(ds,dz) \le$$

$$\le \int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|^{\delta-1}\mu^n(ds,dz)\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|\mu^n(ds,dz)$$

for $\delta \ge 1$, we get it is upper bounded by

$$\mathbb{E}[|h(\hat{L}_1)|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z| + \tag{3.131}$$

$$+\Delta_n|z|^\alpha)\mu^n(ds,dz)\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|\mu^n(ds,dz)|\mathcal{H}_{L_1^n}(1)|W^n].$$

We now use Holder inequality with p_2 big and p_1 next to 1 and we observe that, from the second point of Lemma 30, it follows

$$\mathbb{E}[|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|\mu^n(ds,dz)\mathcal{H}_{L_1^n}(1)|^{p_2}]^{\frac{1}{p_2}} \le C_{p_2}.$$

Hence (3.131) is upper bounded by

$$C_{p_2}\mathbb{E}[|h(\hat{L}_1)|^{p_1}|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^{\alpha})\mu^n(ds,dz)|^{p_1}W^n]^{\frac{1}{p_1}}\le \tag{3.132}$$

$$\le C_{p_2}\left\|h\right\|_\infty\Delta_n\mathbb{E}[|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|^{\alpha}\mu^n(ds,dz)|^{p_1}]^{\frac{1}{p_1}}+ \tag{3.133}$$

$$+C_{p_2}\Delta_n^{\frac{1}{\alpha}}\mathbb{E}[|h(\hat{L}_1)|^{p_1}|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|\mu^n(ds,dz)|^{p_1}W^n]^{\frac{1}{p_1}}.$$

Concerning the first term of (3.133), we use Lemma 2.1.5 in the appendix of [46] with $p_1=(1+\epsilon)\in[1,2]$ and the definition of F_n given in (3.113), getting

$$\mathbb{E}[|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|^{\alpha}\mu^n(ds,dz)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}}\le\mathbb{E}[\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|^{\alpha(1+\epsilon)}\bar{\mu}^n(ds,dz)]^{\frac{1}{1+\epsilon}}\le$$

$$\le c(\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|^{\alpha(1+\epsilon)-1-\alpha}dz)^{\frac{1}{1+\epsilon}}\le c\Delta_n^{-\frac{\epsilon}{1+\epsilon}}=c\Delta_n^{-\epsilon}, \tag{3.134}$$

where we have used the arbitrariness of ϵ in the last equality.
On the second term of (3.133) we act differently depending on whether or not α is more than 1. If it does, we act as we did in (3.127), considering $p_1=1+\epsilon<\alpha$ and introducing r, this time we set it such that the following equality holds:

$$p(1+\epsilon)(1-r)+(1+\epsilon)=\alpha. \tag{3.135}$$

We also use the property (3.25) on h, hence it is upper bounded by

$$C_{p_2}\Delta_n^{\frac{1}{\alpha}}\left\|h\right\|_\infty^r\left\|h\right\|_{pol}^{1-r}\mathbb{E}[(1+|\hat{L}_1|^{p(1+\epsilon)(1-r)})|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|\mu^n(ds,dz)|^{1+\epsilon}W^n]^{\frac{1}{1+\epsilon}}. \tag{3.136}$$

Now on the first term here above we use that $0\le W^n\le 1$ and Lemma 2.1.5 in the appendix of [46] as we did in (3.134) in order to get

$$\mathbb{E}[|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|\mu^n(ds,dz)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}}\le c. \tag{3.137}$$

Moreover we observe, as we have already done, that $|\hat{L}_1|\le|L_1^n|+|L_1^{\alpha,n}|$ and that, from the second point of Lemma 31, there exists $c>0$ such that $|h_n(z)|\le c|z|$; so we get

$$\mathbb{E}[|\hat{L}_1|^{p(1+\epsilon)(1-r)}|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|\mu^n(ds,dz)|^{1+\epsilon}W^n]^{\frac{1}{1+\epsilon}}\le$$

$$\le c+\mathbb{E}[|\int_0^1\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|\mu^n(ds,dz)|^{p(1+\epsilon)(1-r)+(1+\epsilon)}]^{\frac{1}{1+\epsilon}}\le$$

$$\le c(\int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}|z|^{\alpha}|z|^{-1-\alpha}dz)^{\frac{1}{1+\epsilon}}\le c\frac{1}{1+\epsilon}log(\Delta_n^{-\frac{1}{\alpha}})\le c|log(\Delta_n)|, \tag{3.138}$$

having chosen a particular r just in order to have the exponent here above equal to α and so having found out the same computation of (3.7.3). We haven't considered the integral on $|z|\le 1$ only because, as we have already seen above (3.7.3), the integral is bounded in L^p and so we simply get (3.137) again. From (3.135) we obtain $r = 1 + \frac{1}{p} - \frac{\alpha}{p(1+\epsilon)}$. Replacing it and using (3.137) and (3.138) we get (3.136) is upper bounded by

$$C_{p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p(1+\epsilon)}} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p(1+\epsilon)}} (c + |log(\Delta_n)|) = \tag{3.139}$$

$$= C_{p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p(1+\epsilon)}} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p(1+\epsilon)}} |log(\Delta_n)|.$$

If otherwise α is less than 1, then the second term of (3.133) is upper bounded by

$$C_{p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty \mathbb{E}[|\int_0^1 \int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|\mu^n(ds,dz)|^{p_1} W^n]^{\frac{1}{p_1}} \le \tag{3.140}$$

$$\le C_{p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty \Delta_n^{\frac{1}{1+\epsilon}-\frac{1}{\alpha}} = C_{p_2} \Delta_n^{\frac{1}{1+\epsilon}} \|h\|_\infty ,$$

where we have taken $p_1 = 1 + \epsilon$ and we have used the fact that $0 \le W^n \le 1$ and that, for $\alpha < 1$,

$$\mathbb{E}[|\int_0^1 \int_{1<|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} |z|\mu^n(ds,dz)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le c\Delta_{n,i}^{\frac{1}{1+\epsilon}-\frac{1}{\alpha}}.$$

Using (3.133), (3.134), (3.139) and (3.140) it follows

$$\mathbb{E}[|h(\hat{L}_1)|| \int_0^1 \int_{1\le|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} (h_n(z) - z)\mu^n(ds,dz)||\mathcal{H}_{L_1^n}(1)W^n|] \le$$

$$\le C_{p_2} \Delta_n^{1-\epsilon} \|h\|_\infty + C_{p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p(1+\epsilon)}} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p(1+\epsilon)}} |log(\Delta_n)| 1_{\alpha>1}. \tag{3.141}$$

Now from (3.124), (3.130), and (3.141) it follows

$$\tilde{T}_{2,1} \le C_{q_1q_2p_2} \Delta_n^{1-\epsilon} \|h\|_\infty + C_{q_1q_2p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1-\frac{\alpha}{p}+\epsilon} \|h\|_{pol}^{\frac{\alpha}{p}-\epsilon} |log(\Delta_n)| + \tag{3.142}$$

$$+ C_{q_1q_2p_2} \Delta_n^{\frac{1}{\alpha}} \|h\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p}+\epsilon} \|h\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p}-\epsilon} |log(\Delta_n)| 1_{\alpha>1}.$$

Concerning $\tilde{T}_{2,2}$, it is already proved in Theorem 4.2 in [19] that

$$\tilde{T}_{2,2} \le c\Delta_n \|h\|_\infty . \tag{3.143}$$

Let us now consider $\tilde{T}_1$. Using (3.114) and (3.116) we can write

$$\mathcal{H}_{L_1^n}(W^n) = \frac{-W^n L(L_1^n)}{\Gamma(L_1^n, L_1^n)} + L(\frac{W^n}{\Gamma(L_1^n, L_1^n)}) L_1^n - L(\frac{L_1^n W^n}{\Gamma(L_1^n, L_1^n)}).$$

With computations using that L is a self-adjoint operator we get

$$\tilde{T}_1 = |\mathbb{E}[h(L_1^{\alpha,n})W^n] - \mathbb{E}[h(L_1^{\alpha,n}) \frac{\Gamma(L_1^{\alpha,n}, L_1^n)}{\Gamma(L_1^n, L_1^n)} W^n]| \le \mathbb{E}[|h(\hat{L}_1)|| \frac{\Gamma(L_1^n - L_1^{\alpha,n}, L_1^n)}{\Gamma(L_1^n, L_1^n)} |W^n]. \tag{3.144}$$

Using equation (3.114), we have

$$\Gamma(L_1^n - L_1^{\alpha,n}, L_1^n) = \int_0^1 \int_{|z|<\frac{\eta}{2}\Delta_n^{-\frac{1}{\alpha}}} \rho(z)(1 - h_n'(z))\mu^n(ds,dz).$$

Using the third point of Lemma 31 we deduce the following on the event $W^n \neq 0$:

$$|\Gamma(L_1^n - L_1^{\alpha,n}, L_1^n)| \le c \int_0^1 \int_{|z|\le\epsilon_0\Delta_n^{-\frac{1}{\alpha}}} \rho(z)(\Delta_n^{\frac{1}{\alpha}}|z| + \Delta_n|z|^\alpha)\mu^n(ds,dz) \le$$

$$\leq c\int_0^1\int_{|z|\leq 1}\rho(z)(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^\alpha)\mu^n(ds,dz)+$$
$$+c\int_0^1\int_{1<|z|\leq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}\rho(z)\mu^n(ds,dz)\int_0^1\int_{1<|z|\leq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^\alpha)\mu^n(ds,dz)\leq$$
$$\leq c\int_0^1\int_{\mathbb{R}}\rho(z)\mu^n(ds,dz)(\Delta_n^{\frac{1}{\alpha}}+\Delta_n)+$$
$$+c\int_0^1\int_{\mathbb{R}}\rho(z)\mu^n(ds,dz)\int_0^1\int_{1<|z|\leq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^\alpha)\mu^n(ds,dz)=$$
$$=c(\Delta_n^{\frac{1}{\alpha}}+\Delta_n)\Gamma(L_1^n,L_1^n)+c\Gamma(L_1^n,L_1^n)(\int_0^1\int_{1<|z|\leq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^\alpha)\mu^n(ds,dz)),\tag{3.145}$$

where we have used that z is always less than 1 in the first integral and that, since ρ is a positive function, we can upper bound the integrals considering whole set $\mathbb{R}$. Moreover, we have used the definition of $\Gamma(L_1^n,L_1^n)$. Replacing (3.145) in (3.144) we get

$$\tilde{T}_1\leq c(\Delta_n^{\frac{1}{\alpha}}+\Delta_n)\mathbb{E}[|h(\hat{L}_1)|]+c\mathbb{E}[|h(\hat{L}_1)|\int_0^1\int_{1<|z|\leq\epsilon_0\Delta_n^{-\frac{1}{\alpha}}}(\Delta_n^{\frac{1}{\alpha}}|z|+\Delta_n|z|^\alpha)\mu^n(ds,dz))]=\tag{3.146}$$
$$=:\tilde{T}_{1,1}+\tilde{T}_{1,2}.$$

Concerning $\tilde{T}_{1,1}$, we have

$$\tilde{T}_{1,1}\leq c\Delta_n\left\|h\right\|_\infty+c\Delta_n^{\frac{1}{\alpha}}\mathbb{E}[|h(\hat{L}_1)|]\leq c\Delta_n\left\|h\right\|_\infty+c\Delta_n^{\frac{1}{\alpha}}\left\|h\right\|_\infty^{1-\frac{\alpha}{p}}\left\|h\right\|_{pol}^{\frac{\alpha}{p}}|log(\Delta_n)|),\tag{3.147}$$

where in the last inequality we have acted exactly like we did in (3.127) and (3.128) with the exponent on h that is exactly equal to 1 instead of $1+\epsilon$ and so we have choosen r such that $p(1-r)=\alpha$. Let us now consider $\tilde{T}_{1,2}$. We observe that it is exactly like (3.132) but with $p_1=1$ instead of $p_1=1+\epsilon$, with the only difference that computing (3.134) now we get $c\,log(\Delta_n^{-\frac{1}{\alpha}})$ instead of $c\Delta_n^{-\epsilon}$ and in the definition (3.135) we choose r such that $p(1-r)+1=\alpha$. Acting exactly like we did above it follows

$$\tilde{T}_{1,2}\leq C_{p_2}\Delta_n|log(\Delta_n)|\left\|h\right\|_\infty+C_{p_2}\Delta_n^{\frac{1}{\alpha}}\left\|h\right\|_\infty^{1+\frac{1}{p}-\frac{\alpha}{p}}\left\|h\right\|_{pol}^{-\frac{1}{p}+\frac{\alpha}{p}}|log(\Delta_n)|1_{\alpha>1}.\tag{3.148}$$

Using (3.142), (3.143), (3.147) and (3.148), the lemma is proved. □

It follows Proposition 21, using also (3.118), (3.119) and (3.122). □

3.8 Appendix

In this section we will prove the technical proposition and lemmas we have used.

3.8.1 Proof of Proposition 3

Proof. Proposition 3. In order to show (3.27), we reformulate $(\Delta X_i^J)^2\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)$ as

$$(\Delta X_i^J)^2[\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)]+(\Delta X_i^J)^2[\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)-\varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)]+\tag{3.149}$$

$$+ + (\Delta X_i^J - \Delta \tilde{X}_i^J)^2 \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) +$$

$$+2\Delta \tilde{X}_i^J (\Delta X_i^J - \Delta \tilde{X}_i^J) \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) + (\Delta \tilde{X}_i^J)^2 \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) =: \sum_{k=1}^5 I_k^n(i).$$

Comparing (3.27) with (3.149) it turns out that our goal is to show that $\sum_{k=1}^4 I_k^n(i) = o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1})$. In the sequel will prove that $\sum_{k=1}^4 \mathbb{E}[|I_k^n(i)|] \le c\Delta_{n,i}^{\beta(2-\alpha)+1}$; the same reasoning applies to the conditional version, that is

$$\sum_{k=1}^4 \mathbb{E}_i[|I_k^n(i)|] \le R(\Delta_{n,i}^{\beta(2-\alpha)+1}, X_{t_i}).$$

Let us start considering $I_1^n(i)$. We know that $\Delta X_i = \Delta X_i^c + \Delta X_i^J$, where we have denoted by ΔX_i^c the continuous part of the increments of the process X. We study

$$I_1^n(i) = I_{1,1}^n + I_{1,2}^n := I_1^n(i) 1_{\left\{|\Delta X_i| \ge 3\Delta_{n,i}^\beta\right\}} + I_1^n(i) 1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta\right\}}, \tag{3.150}$$

having omitted the dependence upon i in $I_{1,1}^n$ and $I_{1,2}^n$ in order to make the notation easier. Concerning $I_{1,1}^n$, we split again on the sets $\left\{|\Delta X_i^J| \ge 2\Delta_{n,i}^\beta\right\}$ and $\left\{|\Delta X_i^J| < 2\Delta_{n,i}^\beta\right\}$. Recalling that $\varphi(\zeta) = 0$ for $|\zeta| \ge 2\Delta_{n,i}^\beta$, we observe that if $|\Delta X_i^J| \ge 2\Delta_{n,i}^\beta$ then $I_{1,1}^n$ is just 0. Otherwise, if $|\Delta X_i^J| < 2\Delta_{n,i}^\beta$, then it means that $|\Delta X_i^c|$ must be more than $\Delta_{n,i}^\beta$, so we can use (3.38). In the sequel the constant c may change value from line to line. Using the bound on $(\Delta X_i^J)^2$ and the boundedness of φ we get

$$\mathbb{E}[|I_{1,1}^n|] \le c\Delta_{n,i}^{2\beta} \mathbb{E}[1_{\left\{|\Delta X_i| \ge 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 2\Delta_{n,i}^\beta\right\}}] \le c\Delta_{n,i}^{2\beta} \mathbb{P}(|\Delta X_i^c| \ge \Delta_{n,i}^\beta) \le c\Delta_{n,i}^{2\beta + (\frac{1}{2} - \beta)r}. \tag{3.151}$$

Hence

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \mathbb{E}[|I_{1,1}^n|] \le c\Delta_{n,i}^{(\frac{1}{2}-\beta)r - 1 + \alpha\beta}, \tag{3.152}$$

that goes to 0 for $n \to \infty$ since for each choice of $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0, 2)$ we can always find r big enough such that the exponent on $\Delta_{n,i}$ is positive.
We now consider $I_{1,2}^n$ on the sets $\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}$ and $\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}$. Using the boundedness of φ we have

$$\mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}}] \le c\mathbb{E}[(\Delta X_i^J)^2 1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}}].$$

We observe that also in this case $|\Delta X_i| < 3\Delta_{n,i}^\beta$ and $|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta$ involve $|\Delta X_i^c| \ge \Delta_{n,i}^\beta$. Moreover $(\Delta X_i^J)^2 \le c(\Delta X_i)^2 + c(\Delta X_i^c)^2 \le c\Delta_{n,i}^{2\beta} + c(\Delta X_i^c)^2$, hence

$$\mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}}] \le c\Delta_{n,i}^{2\beta} \mathbb{P}(|\Delta X_i^c| \ge \Delta_{n,i}^\beta) + c\mathbb{E}[(\Delta X_i^c)^2 1_{\left\{|\Delta X_i^c| \ge \Delta_{n,i}^\beta\right\}}] \le$$

$$\le c\Delta_{n,i}^{2\beta + r(\frac{1}{2}-\beta)} + c\mathbb{E}[(\Delta X_i^c)^4]^{\frac{1}{2}} \mathbb{P}(|\Delta X_i^c| \ge \Delta_{n,i}^\beta)^{\frac{1}{2}} \le c\Delta_{n,i}^{[2\beta + r(\frac{1}{2}-\beta)] \wedge [1 + \frac{r}{2}(\frac{1}{2}-\beta)]}, \tag{3.153}$$

where we have used Cauchy Schwartz inequality, (3.38) and the fourth point of Lemma 25. Therefore we get

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}}] \le c\Delta_{n,i}^{[r(\frac{1}{2}-\beta) - 1 + \alpha\beta] \wedge [\frac{r}{2}(\frac{1}{2}-\beta) - \beta(2-\alpha)]}, \tag{3.154}$$

that converges to 0 for $n \to \infty$ since we can always find $r \geq 1$ such that the exponent $\Delta_{n,i}$ is positive.
In order to conclude the study of $I_1^n(i)$, we study $I_{1,2}^n 1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}$.

$$\mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}] \leq c \left\| \varphi' \right\|_\infty \Delta_{n,i}^{-\beta} \mathbb{E}[(\Delta X_i^J)^2 |\Delta X_i - \Delta X_i^J| 1_{\left\{|\Delta X_i| \leq 3\Delta_{n,i}^\beta, |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}], \tag{3.155}$$

where we have used the smoothness of φ. Using Holder inequality and the fourth point of Lemma 25 it is upper bounded by

$$c\Delta_{n,i}^{-\beta} \mathbb{E}[|\Delta X_i^c|^p]^{\frac{1}{p}} \mathbb{E}[|(\Delta X_i^J)^{2q} 1_{\left\{|\Delta X_i| \leq 3\Delta_{n,i}^\beta, |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{q}} \leq \tag{3.156}$$

$$\leq c\Delta_{n,i}^{\frac{1}{2}-\beta} \mathbb{E}[|(\Delta X_i^J)^{2q} 1_{\left\{|\Delta X_i| \leq 3\Delta_{n,i}^\beta, |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{q}}.$$

Now, since our indicator function $1_{\left\{|\Delta X_i| \leq 3\Delta_{n,i}^\beta, |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}$ is less then $1_{\left\{|\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}$, we can use the first point of Lemma 28. Through the use of the conditional expectation we get

$$\mathbb{E}[|(\Delta X_i^J)^{2q} 1_{\left\{|\Delta X_i| \leq 3\Delta_{n,i}^\beta, |\Delta X_i^J| \leq 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{q}} \leq c\Delta_{n,i}^{\frac{1+\beta(2q-\alpha)}{q}} \mathbb{E}[R(1, X_{t_i})] \leq c\Delta_{n,i}^{\frac{1+\beta(2q-\alpha)}{q}}, \tag{3.157}$$

where in the last inequality we have used the polynomial growth of R and the third point of Lemma 26.
Replacing (3.157) in (3.156) and taking q small (next to 1), we obtain

$$\mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}] \leq c\Delta_{n,i}^{\frac{1}{2}+\beta+1-\alpha\beta-\epsilon}.$$

It follows

$$\frac{\mathbb{E}[|I_{1,2}^n| 1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}]}{\Delta_{n,i}^{\beta(2-\alpha)+1}} \leq c\Delta_{n,i}^{\frac{1}{2}-\beta-\epsilon}, \tag{3.158}$$

that goes to 0 for $n \to \infty$ since we can always find an ϵ as small as the exponent on $\Delta_{n,i}$ is positive, for $\beta \in (0, \frac{1}{2})$.
Let us now consider $I_2^n(i)$.

$$I_2^n(i) = I_2^n(i)\, 1_{\left\{|\Delta X_i^J| \leq 2\Delta_{n,i}^\beta\right\}} + I_2^n(i)\, 1_{\left\{|\Delta X_i^J| > 2\Delta_{n,i}^\beta\right\}} =: I_{2,1}^n + I_{2,2}^n. \tag{3.159}$$

Concerning the first term of (3.159), we have

$$\mathbb{E}[|I_{2,1}^n|] \leq \Delta_{n,i}^{-\beta} \left\| \varphi' \right\|_\infty \mathbb{E}[(\Delta X_i^J)^2 |\Delta X_i^J - \Delta \tilde{X}_i^J| 1_{\left\{|\Delta X_i^J| \leq 2\Delta_{n,i}^\beta\right\}}] \leq$$

$$\leq c\Delta_{n,i}^{-\beta} \mathbb{E}[(\Delta X_i^J)^4 1_{\left\{|\Delta X_i^J| \leq 2\Delta_{n,i}^\beta\right\}}]^{\frac{1}{2}} \mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^2]^{\frac{1}{2}}, \tag{3.160}$$

where we have used the smoothness of φ and Cauchy-Schwartz inequality. Using again the first point of Lemma 28, we have that

$$\mathbb{E}[(\Delta X_i^J)^4 1_{\left\{|\Delta X_i^J| \leq 2\Delta_{n,i}^\beta\right\}}]^{\frac{1}{2}} = \mathbb{E}[\mathbb{E}_i[(\Delta X_i^J)^4 1_{\left\{|\Delta X_i^J| \leq 2\Delta_{n,i}^\beta\right\}}]]^{\frac{1}{2}} \leq \tag{3.161}$$

$$\leq \Delta_{n,i}^{\frac{1+\beta(4-\alpha)}{2}} \mathbb{E}[R(1, X_{t_i})] \leq c\Delta_{n,i}^{\frac{1}{2}+2\beta-\frac{\alpha\beta}{2}},$$

where we have also used the polynomial growth of R and the third point of Lemma 26.
We now introduce a lemma that will be proved later:

Lemma 33. *Suppose that Assumption* 1 *to* 4 *hold. Then*

1. $\forall q \geq 2$ *we have*
$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^q] \leq c\Delta_{n,i}^2, \tag{3.162}$$

2. *for* $q \in [1,2]$ *and* $\alpha < 1$, *we have*
$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^q]^{\frac{1}{q}} \leq c\Delta_{n,i}^{\frac{1}{2}+\frac{1}{q}}. \tag{3.163}$$

Replacing (3.161) and (3.162) in (4.54) we get

$$\mathbb{E}[|I_{2,1}^n|] \leq c\Delta_{n,i}^{-\beta+\frac{1}{2}+2\beta-\frac{\alpha\beta}{2}+1} = c\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}}. \tag{3.164}$$

Hence
$$\frac{\mathbb{E}[|I_{2,1}^n|]}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \leq c\Delta_{n,i}^{\frac{1}{2}-\beta+\frac{\alpha\beta}{2}}, \tag{3.165}$$

that goes to 0 for $n \to \infty$ since the exponent on $\Delta_{n,i}$ is positive for $\beta < \frac{1}{2(1-\frac{\alpha}{2})}$, that is always true with α and β in the intervals choosen.
We now want to show that also $I_{2,2}^n$ is $o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1})$. We split $I_{2,2}^n$ on the sets $\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}$ and $\left\{|\Delta \tilde{X}_i^J| > 2\Delta_{n,i}^{\beta}\right\}$. We observe that, by the definition of φ, $I_{2,2}^n$ is null on the second set. Adding and subtracting $\Delta \tilde{X}_i^J$ in $I_{2,2}^n 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}}$ we have

$$\mathbb{E}[|I_{2,2}^n| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}}] \leq$$
$$\leq c\mathbb{E}[(\Delta X_i^J - \Delta \tilde{X}_i^J)^2 |\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i^J) - \varphi_{\Delta_{n,i}^{\beta}}(\Delta \tilde{X}_i^J)| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}, |\Delta X_i^J| > 2\Delta_{n,i}^{\beta}\right\}}] +$$
$$+ c\mathbb{E}[(\Delta \tilde{X}_i^J)^2 |\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i^J) - \varphi_{\Delta_{n,i}^{\beta}}(\Delta \tilde{X}_i^J)| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}}]. \tag{3.166}$$

On the second term of (3.166) we can act exactly as we have done in $I_{2,1}^n$, with $\Delta \tilde{X}_i^J$ instead of ΔX_i^J (and so using (3.32) instead of (3.31)). We get

$$\mathbb{E}[(\Delta \tilde{X}_i^J)^2 |\varphi_{\Delta_{n,i}^{\beta}}(\Delta X_i^J) - \varphi_{\Delta_{n,i}^{\beta}}(\Delta \tilde{X}_i^J)| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}}] \leq c\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}}. \tag{3.167}$$

Concerning the first term of (3.166), by the definition of φ we know it is

$$\mathbb{E}[(\Delta X_i^J - \Delta \tilde{X}_i^J)^2 | - \varphi_{\Delta_{n,i}^{\beta}}(\Delta \tilde{X}_i^J)| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}, |\Delta X_i^J| > 2\Delta_{n,i}^{\beta}\right\}}] \leq \tag{3.168}$$
$$\leq c\mathbb{E}[(\Delta X_i^J - \Delta \tilde{X}_i^J)^2] \leq c\Delta_{n,i}^2,$$

where in the last inequality we have used (3.162). Using (3.166) - (3.168) it follows

$$\mathbb{E}[|I_{2,2}^n|] = \mathbb{E}[|I_{2,2}^n| 1_{\left\{|\Delta \tilde{X}_i^J| \leq 2\Delta_{n,i}^{\beta}\right\}}] \leq c\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}} + c\Delta_{n,i}^2 = c\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}}, \tag{3.169}$$

considering that $\Delta_{n,i}^2$ is negligible compared to $\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}}$ since $\beta < \frac{1}{2(1-\frac{\alpha}{2})}$. Hence

$$\frac{\mathbb{E}[|I_{2,2}^n|]}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \leq c\Delta_{n,i}^{\frac{1}{2}-\beta+\frac{\alpha\beta}{2}}, \tag{3.170}$$

that goes to 0 for $n \to \infty$.
Concerning $I_3^n(i)$, we have

$$\mathbb{E}[|I_3^n(i)|] \le c\mathbb{E}[(\Delta X_i^J - \Delta \tilde{X}_i^J)^2] \le c\Delta_{n,i}^2, \tag{3.171}$$

where the last inequality follows from (3.162). Hence $I_3^n(i) = o_{L^1}(\Delta_{n,i}^{\beta(2-\alpha)+1})$, indeed

$$\frac{\mathbb{E}[|I_3^n(i)|]}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \le c\Delta_{n,i}^{1-2\beta+\alpha\beta}, \tag{3.172}$$

that goes to 0 for $n \to \infty$ considering that the exponent on $\Delta_{n,i}$ is positive for $\beta < \frac{1}{2-\alpha}$, condition that is always satisfied for $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0, 2)$.
Let us now consider $I_4^n(i)$. Using Cauchy-Schwartz inequality it is

$$\mathbb{E}[|I_4^n(i)|] \le c\mathbb{E}[(\Delta X_i^J - \Delta \tilde{X}_i^J)^2]^{\frac{1}{2}}\mathbb{E}[(\Delta \tilde{X}_i^J)^2\varphi^2_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)]^{\frac{1}{2}} \le \tag{3.173}$$

$$\le c\Delta_{n,i}\Delta_{n,i}^{\frac{1}{2}+\frac{\beta}{2}(2-\alpha)} = c\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}},$$

where we have used (3.162) and the first point of Lemma 28. It follows

$$\frac{\mathbb{E}[|I_4^n(i)|]}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \le c\Delta_{n,i}^{\frac{1}{2}-\beta+\frac{\alpha\beta}{2}}, \tag{3.174}$$

that goes to 0 for $n \to \infty$ since the exponent on $\Delta_{n,i}$ is more than 0 if $\beta < \frac{1}{2(1-\frac{\alpha}{2})}$, that is always true. Using (3.149), (3.152), (3.154), (3.158), (3.165), (3.170), (4.34) and (4.37) we obtain (3.27).

In order to prove (3.28), we use again reformulation (3.149). Replacing it in the left hand side of (3.28) it turns out that our goal is to show that

$$\frac{1}{n}\sum_{i=0}^{n-1}(\sum_{k=1}^{4} I_k^n(i))\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.175}$$

Using a conditional on $\mathcal{F}_{t_i}$ version of (3.159), (3.164) and (3.169) we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\mathbb{E}_i[|I_2^n(i)f(X_{t_i})|]\frac{1}{\Delta_{n,i}} \le \frac{1}{n}\sum_{i=0}^{n-1}R(\Delta_n^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}-1-\epsilon}, X_{t_i}) = \frac{1}{n}\sum_{i=0}^{n-1}R(\Delta_n^{\frac{1}{2}+\beta-\frac{\alpha\beta}{2}-\epsilon}, X_{t_i}).$$

Since $\beta(1-\frac{\alpha}{2})$ is always more than zero and, $\forall \tilde{\epsilon} > 0$ we can always find ϵ smaller than it, we get

$$\frac{1}{n}\sum_{i=0}^{n-1}I_2^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}) = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.176}$$

From a conditional version of (3.171) we get that $\frac{1}{n}\sum_{i=0}^{n-1}I_3^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}}$ is upper bounded in conditional norm 1 by $\frac{1}{n}\sum_{i=0}^{n-1}R(\Delta_n^{2-1-\epsilon}, X_{t_i}) = \frac{1}{n}\sum_{i=0}^{n-1}R(\Delta_n^{1-\epsilon}, X_{t_i})$ and so

$$\frac{1}{n}\sum_{i=0}^{n-1}I_3^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.177}$$

Using a conditional version of (3.173) we get that $\frac{1}{n}\sum_{i=0}^{n-1} I_4^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}}$ is upper bounded in conditional norm 1 by

$$\frac{1}{n}\sum_{i=0}^{n-1} R(\Delta_n^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}-1-\epsilon}, X_{t_i}) = \frac{1}{n}\sum_{i=0}^{n-1} R(\Delta_n^{\frac{1}{2}+\beta-\frac{\alpha\beta}{2}-\epsilon}, X_{t_i}),$$

hence

$$\frac{1}{n}\sum_{i=0}^{n-1} I_4^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.178}$$

Concerning $I_1^n(i)$, we consider $I_{1,1}^n(i)$ and $I_{1,2}^n(i)$ as defined in (3.150). Using a conditional version of (3.151) on $I_{1,1}^n(i)$ it follows that $\frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\frac{1}{n}\sum_{i=0}^{n-1} I_{1,1}^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}}$ is upper bounded in norm 1 by

$$\frac{1}{n}\sum_{i=0}^{n-1} R(\Delta_n^{(\frac{1}{2}-\beta)r+2\beta-1-\frac{1}{2}+\tilde{\epsilon}}, X_{t_i}) = \frac{1}{n}\sum_{i=0}^{n-1} R(\Delta_n^{(\frac{1}{2}-\beta)r+2\beta-\frac{3}{2}+\tilde{\epsilon}}, X_{t_i}),$$

that goes to zero because we can find r big enough such that the exponent on Δ_n is positive, hence

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{1,1}^n(i)\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}) = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.179}$$

Acting as we did in the proof of (3.27), we consider $I_{1,2}^n(i)$ on the sets $\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}$ and $\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}$. Again, from (3.153) and the arbitrariness of $r > 0$ it follows

$$\frac{1}{n}\sum_{i=0}^{n-1} I_{1,2}^n(i)1_{\left\{|\Delta X_i^J|\ge 4\Delta_{n,i}^\beta\right\}}\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.180}$$

When $|\Delta X_i^J| < 4\Delta_{n,i}^\beta$ we act in a different way, considering the development up to second order of $\varphi_{\Delta_{n,i}^\beta}$, centered in ΔX_i^J:

$$I_{1,2}^n(i)1_{\left\{|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}} = [(\Delta X_i^J)^2\Delta X_i^c\varphi'_{\Delta_{n,i}^\beta}(\Delta X_i^J)\Delta_{n,i}^{-\beta}+$$

$$+(\Delta X_i^J)^2(\Delta X_i^c)^2\varphi''_{\Delta_{n,i}^\beta}(X_u)\Delta_{n,i}^{-2\beta}]1_{\left\{|\Delta X_i|\le 3\Delta_{n,i}^\beta,\,|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}} =$$

$$=: \hat{I}_1^n(i)1_{\left\{|\Delta X_i|\le 3\Delta_{n,i}^\beta,\,|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}} + \hat{I}_2^n(i)1_{\left\{|\Delta X_i|\le 3\Delta_{n,i}^\beta,\,|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}},$$

where $X_u \in [\Delta X_i^J, \Delta X_i]$. Now, acting like we did in (3.155), (3.156) and (3.157), taking q next to 1 we get

$$\mathbb{E}_i[|\hat{I}_2^n(i)1_{\left\{|\Delta X_i|\le 3\Delta_{n,i}^\beta,\,|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}}|] \le R(\Delta_n^{1+\beta(2-\alpha)-\epsilon+1-2\beta}, X_{t_i}) = R(\Delta_n^{2-\alpha\beta-\epsilon}, X_{t_i}).$$

Since for each $\tilde{\epsilon} > 0$ we can find an ϵ such that $\tilde{\epsilon} - \epsilon > 0$ it follows, taking the conditional expectation

$$\frac{1}{n}\sum_{i=0}^{n-1} \hat{I}_2^n(i)1_{\left\{|\Delta X_i|\le 3\Delta_{n,i}^\beta,\,|\Delta X_i^J|<4\Delta_{n,i}^\beta\right\}}\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{1-\alpha\beta-\tilde{\epsilon}}) = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.181}$$

Concerning $\hat{I}_1^n(i) 1_{\left\{|\Delta X_i| \le 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4 \Delta_{n,i}^\beta\right\}}$, we no longer consider the indicator function because it is

$$(\Delta X_i^J)^2 \Delta X_i^c \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i^J)\Delta_{n,i}^{-\beta} +$$
$$+ (\Delta X_i^J)^2 \Delta X_i^c \varphi'_{\Delta_{n,i}^\beta}(\Delta X_i^J)\Delta_{n,i}^{-\beta}(1_{\left\{|\Delta X_i| \le 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4 \Delta_{n,i}^\beta\right\}} - 1)$$

and the second term here above is different from zero only on a set smaller that $\left\{|\Delta X_i| \ge 3\Delta_{n,i}^\beta\right\}$ or $\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}$, on which we have already proved the result (see the study of $I_{1,1}^n(i)$ in (3.179) and $I_{1,2}^n(i)$ in (3.180)). We want to show that

$$\frac{1}{n}\sum_{i=0}^{n-1} \hat{I}_1^n(i) \frac{f(X_{t_i})}{\Delta_{n,i}} = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.182}$$

We start from the reformulation

$$\hat{I}_1^n(i) = \Delta X_i^c \Delta_{n,i}^{-\beta}[(\Delta X_i^J)^2(\varphi'_{\Delta_{n,i}^\beta}(\Delta X_i^J) - \varphi'_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)) + (\Delta X_i^J - \Delta\tilde{X}_i^J)^2 \varphi'_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J) +$$
$$+ 2\Delta\tilde{X}_i^J(\Delta X_i^J - \Delta\tilde{X}_i^J)\varphi'_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J) + (\Delta\tilde{X}_i^J)^2\varphi'_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)] = \sum_{j=1}^4 \hat{I}_{1,j}^n(i).$$

and we observe that, after have used Holder inequality and have remarked that $\varphi'_{\Delta_{n,i}^\beta}$ acts like $\varphi_{\Delta_{n,i}^\beta}$, we can act on $\hat{I}_{1,1}^n$ as we did on I_2^n, on $\hat{I}_{1,2}^n$ as on I_3^n and on $\hat{I}_{1,3}^n$ as on I_4^n. So we get, using also Holder inequality and the fourth point of Lemma 25,

$$\mathbb{E}_i[|\hat{I}_{1,1}^n(i) + \hat{I}_{1,2}^n(i) + \hat{I}_{1,3}^n(i)|] \le R(\Delta_{n,i}^{\frac{1}{2}-\beta}, X_{t_i})(\mathbb{E}_i[|I_2^n(i)|^q]^{\frac{1}{q}} + \mathbb{E}[|I_3^n(i)|^q]^{\frac{1}{q}} + \mathbb{E}[|I_4^n(i)|^q]^{\frac{1}{q}}). \tag{3.183}$$

Now, taking q next to 1, we need the following lemma that we will prove later:

Lemma 34. *Suppose that Assumption* 1 *to* 4 *hold. Then,* $\forall \epsilon > 0$,

$$\mathbb{E}_i[|I_2^n(i)|^{1+\epsilon} + |I_3^n(i)|^{1+\epsilon} + |I_4^n(i)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le R(\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}-\epsilon}, X_{t_i}), \tag{3.184}$$

with $I_2^n(i)$, $I_3^n(i)$ *and* $I_4^n(i)$ *as defined in* (3.149).

From (3.183) and (3.184) it follows

$$\frac{1}{n}\sum_{i=0}^{n-1}[\hat{I}_{1,1}^n(i) + \hat{I}_{1,2}^n(i) + \hat{I}_{1,3}^n(i)]\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}) = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.185}$$

On $\frac{1}{n}\sum_{i=0}^{n-1} \hat{I}_{1,4}^n \frac{f(X_{t_i})}{\Delta_{n,i}} =: \sum_{i=0}^{n-1}\zeta_{n,i}$ we want to use Lemma 9 in [36]. By the independence between L and W we get

$$\frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\sum_{i=0}^{n-1}\mathbb{E}_i[\zeta_{n,i}] = \frac{1}{\Delta_n^{\frac{1}{2}-\tilde{\epsilon}}}\frac{1}{n}\sum_{i=0}^{n-1} f(X_{t_i})\Delta_{n,i}^{-1-\beta}\mathbb{E}_i[(\Delta\tilde{X}_i^J)^2\varphi'_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)]\mathbb{E}_i[\Delta X_i^c] = 0 \tag{3.186}$$

and

$$\frac{\Delta_n^{1-2(\frac{1}{2}-\tilde{\epsilon})}}{n\Delta_n}\frac{1}{n}\sum_{i=0}^{n-1} f^2(X_{t_i})\Delta_{n,i}^{-2-2\beta}\mathbb{E}_i[(\Delta\tilde{X}_i^J)^4\varphi'^2_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)]\mathbb{E}_i[(\Delta X_i^c)^2] \le \tag{3.187}$$
$$\le c\Delta_n^{1-1+2\tilde{\epsilon}-2-2\beta+1+1+\beta(4-\alpha)} = c\Delta_n^{2\tilde{\epsilon}+2\beta-\alpha\beta},$$

where we have also used the fourth point of Lemma 25, the fact that $\frac{1}{n\Delta_n}$ is bounded and the first point of Lemma 28. Using (3.186) and (3.187) we have

$$\frac{1}{n}\sum_{i=0}^{n-1}\hat{I}_{1,4}^n\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{\mathbb{P}}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})})$$

that, joint with (3.185) and the fact that the convergence in norm 1 implies the convergence in probability, give us (3.182). Using also (3.176) - (3.181) we get (3.175) and so (3.28).

In order to prove (3.29), we reformulate $\Delta X_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)$ as we have already done in (3.149) getting

$$(\int_{t_i}^{t_{i+1}}a(X_s)dW_s)\Delta X_i^J\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) =$$

$$= (\int_{t_i}^{t_{i+1}}a(X_s)dW_s)(\Delta X_i^J)[\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)]+$$

$$+(\int_{t_i}^{t_{i+1}}a(X_s)dW_s)(\Delta X_i^J)[\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)-\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)]+ \tag{3.188}$$

$$+(\int_{t_i}^{t_{i+1}}a(X_s)dW_s)(\Delta X_i^J-\Delta \tilde{X}_i^J)\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)+$$

$$+(\int_{t_i}^{t_{i+1}}a(X_s)dW_s)(\Delta \tilde{X}_i^J)\varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J) =: \sum_{j=1}^{4}\tilde{I}_j^n(i).$$

Comparing (3.188) with (3.29) it turns out that our goal is to prove that

$$\frac{1}{\Delta_{n,i}^{\beta(2-\alpha)+1}}\sum_{j=1}^{3}\mathbb{E}[|\tilde{I}_j^n(i)|]\to 0,$$

for $n\to\infty$ (again, acting as we do in the sequel it is also possible to show that $\sum_{j=1}^3\mathbb{E}_i[|\tilde{I}_j^n(i)|]\le R(\Delta_{n,i}^{\beta(2-\alpha)+1}, X_{t_i}))$. Let us start considering $\tilde{I}_1^n(i)$. Using Holder inequality, its expected value is upper bounded by

$$\mathbb{E}[|\int_{t_i}^{t_{i+1}}a(X_s)dW_s|^{p_1}]^{\frac{1}{p_1}}\mathbb{E}[|\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2}]^{\frac{1}{p_2}}. \tag{3.189}$$

We now act on $\mathbb{E}[|\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2}]^{\frac{1}{p_2}}$ as we did in the study of $I_1^n(i)$:

$$|\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2} =$$

$$= |\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2}1_{\{|\Delta X_i|\ge 3\Delta_{n,i}^\beta\}}+$$

$$+|\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i)-\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2}1_{\{|\Delta X_i|< 3\Delta_{n,i}^\beta\}} =: \tilde{I}_{1,1}^n+\tilde{I}_{1,2}^n.$$

Concerning $\tilde{I}_{1,1}^n$, if $|\Delta X_i^J|\ge 2\Delta_{n,i}^\beta$ it is just 0, otherwise we can act exactly as we have done on $I_{1,1}^n$, taking $p_2=2$. Hence, $\forall r\ge 1$,

$$\mathbb{E}[|\tilde{I}_{1,1}^n|]^{\frac{1}{2}}\le (c\Delta_{n,i}^{2\beta+r(\frac{1}{2}-\beta)})^{\frac{1}{2}} = c\Delta_{n,i}^{\beta+\frac{r}{2}(\frac{1}{2}-\beta)}. \tag{3.190}$$

Let us now consider $\tilde{I}^n_{1,2}$. If $|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta$, we act again like we did on $I^n_{1,2}$, taking $p_2 = 2$. It yields again

$$\mathbb{E}[|\tilde{I}^n_{1,2}|1_{\left\{|\Delta X_i^J| \ge 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{2}} \le c\Delta_{n,i}^{\beta + \frac{r}{2}(\frac{1}{2}-\beta)}. \tag{3.191}$$

If $|\Delta X_i^J| < 4\Delta_{n,i}^\beta$ we use the smoothness of φ and Holder inequality getting

$$\mathbb{E}[|\tilde{I}^n_{1,2}|1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}] \le \Delta_{n,i}^{-\beta}\mathbb{E}[|\Delta X_i^J|^{p_2}|\varphi'(\zeta)|^{p_2}|\Delta X_i^c|^{p_2}1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{p_2}} \le$$

$$\le \Delta_{n,i}^{-\beta}\mathbb{E}[|\Delta X_i^c|^{p_2 p}]^{\frac{1}{p_2 p}}\mathbb{E}[|\varphi'(\zeta)|^{p_2 q}|\Delta X_i^J|^{p_2 q}1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{p_2 q}}, \tag{3.192}$$

with ζ a point between ΔX_i^J and ΔX_i.
Now we observe that, if $|\Delta X_i^c| \ge \frac{\Delta_{n,i}^\beta}{4}$, then taking $p_2\, q = 1 + \epsilon$ we have

$$\mathbb{E}[|\varphi'(\zeta)|^{1+\epsilon}|\Delta X_i^J|^{1+\epsilon}1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4\Delta_{n,i}^\beta, |\Delta X_i^c| \ge \frac{\Delta_{n,i}^\beta}{4}\right\}}]^{\frac{1}{1+\epsilon}} \le c\Delta_{n,i}^{\beta + r(\frac{1}{2}-\beta)\frac{1}{1+\epsilon}}$$

where we have used the bound on $|\Delta X_i^J|$ given by the indicator function, the boundedness of φ' and (3.38). Otherwise, by the definition of φ, we know that $|\varphi'(\zeta)| \ne 0$ only if $|\zeta| \in (\Delta_{n,i}^\beta, 2\Delta_{n,i}^\beta)$. Then $\Delta_{n,i}^\beta \le |\zeta| \le |\Delta X_i| + |\Delta X_i^J| \le 2|\Delta X_i^J| + |\Delta X_i^c| \le 2|\Delta X_i^J| + \frac{\Delta_{n,i}^\beta}{4}$, hence $|\Delta X_i^J| \ge \frac{3}{8}\Delta_{n,i}^\beta \ge \frac{\Delta_{n,i}^\beta}{4}$ and so we can say it is

$$\mathbb{E}[|\varphi'(\zeta)|^{1+\epsilon}|\Delta X_i^J|^{1+\epsilon}1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4\Delta_{n,i}^\beta, |\Delta X_i^c| < \frac{\Delta_{n,i}^\beta}{4}\right\}}]^{\frac{1}{1+\epsilon}} \le$$

$$\le c\mathbb{E}[|\Delta X_i^J|^{1+\epsilon}1_{\left\{\frac{\Delta_{n,i}^\beta}{4} \le |\Delta X_i^J| < 4\Delta_{n,i}^\beta,\right\}}].$$

Using the second point of Lemma 4, passing through the conditional expected value we get it is upper bounded by

$$\Delta_{n,i}^{1+\beta(1+\epsilon-\alpha)}\mathbb{E}[R(1, X_{t_i})] \le c\Delta_{n,i}^{1+\beta(1+\epsilon-\alpha)},$$

where in the last inequality we have used the polynomial growth of R and the third point of Lemma 26. Hence

$$\mathbb{E}[|\varphi'(\zeta)|^{1+\epsilon}|\Delta X_i^J|^{1+\epsilon}1_{\left\{|\Delta X_i| < 3\Delta_{n,i}^\beta, |\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{1+\epsilon}} \le \tag{3.193}$$

$$\le c\Delta_{n,i}^{[\beta + r(\frac{1}{2}-\beta)-\epsilon]\wedge[1+\beta(1+\epsilon-\alpha)]\frac{1}{1+\epsilon}} = c\Delta_{n,i}^{[1+\beta(1+\epsilon-\alpha)]\frac{1}{1+\epsilon}}.$$

The last equality follows from the fact that, for each choice of $\beta \in (0, \frac{1}{2})$ and $\alpha \in (0,2)$, we can always find $r \ge 1$ and $\epsilon > 0$ such that $\beta + r(\frac{1}{2}-\beta) - \epsilon > 1 + \beta(1+\epsilon-\alpha)$. Replacing (3.193) in (3.192) and using the fourth point of Lemma 25 we have that

$$\mathbb{E}[|\tilde{I}^n_{1,2}|1_{\left\{|\Delta X_i^J| < 4\Delta_{n,i}^\beta\right\}}]^{\frac{1}{p_2}} \le c\Delta_{n,i}^{[\frac{1}{2}-\beta+1+\beta(1+\epsilon-\alpha)]\frac{1}{p_2}} = c\Delta_{n,i}^{(\frac{3}{2}-\alpha\beta-\epsilon)\frac{1}{p_2}} = c\Delta_{n,i}^{\frac{3}{2}-\alpha\beta-\epsilon}, \tag{3.194}$$

the last equality follows from the choice of both p_2 and q next to 1. Using (3.190), (3.191) and (3.194) we get

$$\mathbb{E}[|\Delta X_i^J|^{p_2}|\varphi_{\Delta_{n,i}^\beta}(\Delta X_i) - \varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J)|^{p_2}]^{\frac{1}{p_2}} \le c\Delta_{n,i}^{[\beta+\frac{r}{2}(\frac{1}{2}-\beta)]\wedge[\frac{3}{2}-\alpha\beta-\epsilon]} = c\Delta_{n,i}^{\frac{3}{2}-\alpha\beta-\epsilon}. \tag{3.195}$$

Replacing (3.37) and (3.195) in (3.189) it follows

$$\mathbb{E}[|\tilde{I}_1^n(i)|] \le c\Delta_{n,i}^{2-\alpha\beta-\epsilon}, \tag{3.196}$$

hence

$$\frac{\mathbb{E}[|\tilde{I}_1^n(i)|]}{\Delta_{n,i}^{1+\beta(2-\alpha)}} \le c\Delta_{n,i}^{1-2\beta-\epsilon}. \tag{3.197}$$

Since we can always find an $\epsilon > 0$ such that $1 - 2\beta - \epsilon > 0$, the expected value above goes to 0 for $n \to \infty$.
Concerning $\tilde{I}_2^n(i)$, we split again on $\tilde{I}_{2,1}^n := \tilde{I}_2^n(i)1_{\left\{|\Delta X_i^J| \le 2\Delta_{n,i}^\beta\right\}}$ and $\tilde{I}_{2,2}^n := \tilde{I}_2^n(i)1_{\left\{|\Delta X_i^J| > 2\Delta_{n,i}^\beta\right\}}$.

$$\mathbb{E}[|\tilde{I}_{2,1}^n|] = \mathbb{E}[|\tilde{I}_2^n(i)|1_{\left\{|\Delta X_i^J| \le 2\Delta_{n,i}^\beta\right\}}] \le$$

$$\le c\Delta_{n,i}^{-\beta}\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s||\Delta X_i^J||\Delta X_i^J - \Delta\tilde{X}_i^J|1_{\left\{|\Delta X_i^J| \le 2\Delta_{n,i}^\beta\right\}}] \le$$

$$\le c\Delta_{n,i}^{-\beta}\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^2|\Delta X_i^J|^2 1_{\left\{|\Delta X_i^J| \le 2\Delta_{n,i}^\beta\right\}}]^{\frac{1}{2}}\mathbb{E}[|\Delta X_i^J - \Delta\tilde{X}_i^J|^2]^{\frac{1}{2}} \le$$

$$\le c\Delta_{n,i}^{1-\beta}\mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^{2p}]^{\frac{1}{2p}}\mathbb{E}[|\Delta X_i^J|^{2q} 1_{\left\{|\Delta X_i^J| \le 2\Delta_{n,i}^\beta\right\}}]^{\frac{1}{2q}},$$

where we have used Cauchy-Schwartz inequality, (3.162) and Holder inequality. Now we take p big and q next to 1, using (3.37) and the first point of Lemma 28 we get

$$\mathbb{E}[|\tilde{I}_{2,1}^n|] \le c\Delta_{n,i}^{1-\beta+\frac{1}{2}+\frac{1}{2}+\frac{\beta}{2}(2-\alpha)-\epsilon} \tag{3.198}$$

and so

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}\mathbb{E}[|\tilde{I}_{2,1}^n|] \le \Delta_{n,i}^{1-2\beta+\frac{\alpha\beta}{2}-\epsilon}. \tag{3.199}$$

It goes to 0 for $n \to \infty$ because we can always find an $\epsilon > 0$ such that the exponent in Δ_{n_i} is positive. Let us now consider $\tilde{I}_{2,2}^n = \tilde{I}_{2,2}^n 1_{\left\{|\Delta\tilde{X}_i^J| \le 2\Delta_{n,i}^\beta\right\}} + \tilde{I}_{2,2}^n 1_{\left\{|\Delta\tilde{X}_i^J| > 2\Delta_{n,i}^\beta\right\}}$.
From the definition of φ, $\tilde{I}_{2,2}^n 1_{\left\{|\Delta\tilde{X}_i^J| > 2\Delta_{n,i}^\beta\right\}} = 0$.

$$\mathbb{E}[|\tilde{I}_{2,2}^n|1_{\left\{|\Delta\tilde{X}_i^J| \le 2\Delta_{n,i}^\beta\right\}}] =$$

$$= \mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s||\Delta\tilde{X}_i^J||\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J) - \varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)|1_{\left\{|\Delta\tilde{X}_i^J| \le 2\Delta_{n,i}^\beta, |\Delta X_i^J| > 2\Delta_{n,i}^\beta\right\}}] +$$

$$+ \mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s||\Delta X_i^J - \Delta\tilde{X}_i^J||\varphi_{\Delta_{n,i}^\beta}(\Delta X_i^J) - \varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)|1_{\left\{|\Delta\tilde{X}_i^J| \le 2\Delta_{n,i}^\beta, |\Delta X_i^J| > 2\Delta_{n,i}^\beta\right\}}] \le$$

$$\le c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon} + \mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s||\Delta X_i^J - \Delta\tilde{X}_i^J|| - \varphi_{\Delta_{n,i}^\beta}(\Delta\tilde{X}_i^J)|],$$

where we have acted exactly like we did in $\tilde{I}_{2,1}^n$, using that $\Delta\tilde{X}_i^J$ is less then $2\Delta_{n,i}^\beta$. We have also used that, by the definition of φ, evaluated in ΔX_i^J it is zero. Now we use Holder inequality, (3.37) and the boundedness of φ to get

$$\mathbb{E}[|\tilde{I}_{2,2}^n|] \le c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon} + \mathbb{E}[|\int_{t_i}^{t_{i+1}} a(X_s)dW_s|^p]^{\frac{1}{p}}\mathbb{E}[|\Delta X_i^J - \Delta\tilde{X}_i^J|^q]^{\frac{1}{q}} \le$$

$$\leq c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon} + c\Delta_{n,i}^{\frac{1}{2}}\mathbb{E}[|\Delta X_i^J - \Delta\tilde{X}_i^J|^q]^{\frac{1}{q}}.$$

Now, if $\alpha < 1$ we use (3.163), with $q = 1 + \epsilon$, getting

$$\mathbb{E}[|\tilde{I}_{2,2}^n|] \leq c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon} + c\Delta_{n,i}^{\frac{1}{2}+\frac{1}{2}+\frac{1}{1+\epsilon}} = c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon}. \tag{3.200}$$

Therefore, for $\alpha < 1$, we have

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}\mathbb{E}[|I_{2,2}^n|] \leq c\Delta_{n,i}^{1-2\beta+\frac{\alpha\beta}{2}-\epsilon}. \tag{3.201}$$

We can find an $\epsilon > 0$ such that the exponent on $\Delta_{n,i}$ is positive hence, if $\alpha < 1$, then $I_{2,2}^n = o_{L^1}(\Delta_{n,i}^{1+\beta(2-\alpha)})$. Otherwise, if $\alpha \geq 1$, we use (3.162) having taken $q = 2$. We get

$$\mathbb{E}[|\tilde{I}_{2,2}^n|] \leq c\Delta_{n,i}^{2-\frac{\alpha\beta}{2}-\epsilon} + c\Delta_{n,i}^{\frac{1}{2}+1} = c\Delta_{n,i}^{\frac{3}{2}}.$$

It follows that, for $\alpha \geq 1$, it is

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}\mathbb{E}[|I_{2,2}^n|] \leq c\Delta_{n,i}^{\frac{1}{2}-\beta(2-\alpha)}. \tag{3.202}$$

We observe that the exponent on $\Delta_{n,i}$ is more than 0 if $\beta < \frac{1}{2}\frac{1}{(2-\alpha)}$, that is always true for $\beta \in (0, \frac{1}{2})$ and $\alpha \in [1, 2)$.
To conclude, we use on $\tilde{I}_3(i)$ Holder inequality, (3.37), the boundedness of φ and then we act as we did on $\tilde{I}_{2,2}^n$, using (3.163) or (3.162), depending on whether or not α is less than 1. In the case $\alpha < 1$ we get

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}\mathbb{E}[|\tilde{I}_3^n(i)|] \leq \frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}c\Delta_{n,i}^{\frac{1}{2}+\frac{1}{2}+\frac{1}{1+\epsilon}} = c\Delta_{n,i}^{1-\beta(2-\alpha)-\epsilon}, \tag{3.203}$$

that goes to 0 for $n \to \infty$ since we can always find $\epsilon > 0$ such that the exponent on $\Delta_{n,i}$ is positive. Otherwise it follows

$$\frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}\mathbb{E}[|\tilde{I}_3^n(i)|] \leq \frac{1}{\Delta_{n,i}^{1+\beta(2-\alpha)}}c\Delta_{n,i}^{\frac{3}{2}} = c\Delta_{n,i}^{\frac{1}{2}-\beta(2-\alpha)}. \tag{3.204}$$

The exponent on $\Delta_{n,i}$ is positive if $\beta < \frac{1}{2}\frac{1}{(2-\alpha)}$, that is always true since we are in the case $\alpha \geq 1$. Hence $\tilde{I}_3^n(i) = o_{L^1}(\Delta_{n,i}^{1+\beta(2-\alpha)})$.
From (3.197) - (3.204) and the reformulation (3.188), it follows (3.29).

Replacing reformulation (3.188) in the left hand side of (3.30), it turns out that the theorem is proved if

$$\frac{1}{n}\sum_{i=0}^{n-1}(\sum_{k=1}^{3}\tilde{I}_k^n(i))\frac{f(X_{t_i})}{\Delta_{n,i}} = o_{L^1}(\Delta_n^{(\frac{1}{2}-\tilde{\epsilon})\wedge(1-\alpha\beta-\tilde{\epsilon})}). \tag{3.205}$$

Using a conditional version of equations (3.196), (3.198), (3.200), (3.203) and (3.204) (adding in the last two $\beta(2-\alpha)$ in the exponent of $\Delta_{n,i}$) we easily get (3.205) and so (3.30). □

3.8.2 Proof of Lemma 27

Proof. In this proof, we emphasize that the sampling scheme $(t_i)_{i=0,\dots,n}$depends on n, by noting $t_i = T_{n,i}$, and we have $T_{n,j} = \sum_{i=0}^{j-1} \Delta_{n,i}$. We define $X_{n,j} := \frac{1}{T_{n,j}} \sum_{i=0}^{j-1} \int_{t_i}^{t_{i+1}} h(X_s)ds$ and we observe that

$$T_{n,j+1}X_{n,j+1} - T_{n,j}X_{n,j} = \int_0^{T_{n,j+1}} h(X_s)ds - \int_0^{T_{n,j}} h(X_s)ds = \int_{T_{n,j}}^{T_{n,j+1}} h(X_s)ds =$$

$$= \int_{T_{n,j}}^{T_{n,j+1}} [h(X_s) - h(X_{T_{n,j}})]ds + \Delta_{n,j} h(X_{T_{n,j}}).$$

Hence

$$S_n := \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta} h(X_{T_{n,i}}) = \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} \Delta_{n,i} h(X_{T_{n,i}}) =$$

$$= \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} [T_{n,i+1}X_{n,i+1} - T_{n,i}X_{n,i}] + \tag{3.206}$$

$$+ \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} \Big(\int_{T_{n,i}}^{T_{n,i+1}} [h(X_{T_{n,i}}) - h(X_s)]ds\Big).$$

Now, concerning the second term of (3.206), we have that its norm 1 is upper bounded by

$$\frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} \Big(\int_{T_{n,i}}^{T_{n,i+1}} \|h'\|_\infty \mathbb{E}[|X_{T_{n,i}} - X_s|]ds\Big) \le$$

$$\le \frac{c}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} \int_{T_{n,i}}^{T_{n,i+1}} \|h'\|_\infty |s - T_{n,i}|ds \le \frac{c}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta+\frac{1}{2}} \le c\Delta_n^{\frac{1}{2}},$$

where we have used the regularity of h, Cauchy-Schwartz inequality, the first point of Lemma 25 and the fact that $\Delta_{n,i} \le \Delta_n$. Therefore the second term of (3.206) converges to zero in norm 1, that implies the convergence to zero in probability. Concerning the first term of (3.206), it is

$$\tilde{S}_n := \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} \sum_{i=0}^{n-1} \Delta_{n,i}^{\delta-1} [T_{n,i+1}X_{n,i+1} - T_{n,i}X_{n,i}] =$$

$$= \frac{\sum_{i=1}^{n-1} T_{n,i}X_{n,i}(\Delta_{n,i-1}^{\delta-1} - \Delta_{n,i}^{\delta-1})}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} + \frac{\Delta_{n,n-1}^{\delta-1} T_{n,n}X_{n,n}}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}}.$$

We define

$$a_{n,j} := \frac{1}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}} T_{n,j}(\Delta_{n,j-1}^{\delta-1} - \Delta_{n,j}^{\delta-1}) \qquad \text{for } j \le n-2,$$

$$a_{n,n-1} = \frac{\Delta_{n,n-1}^{\delta-1} T_{n,n}}{\sum_{i=0}^{n-1} \Delta_{n,i}^{\delta}}.$$

Let us start assuming the following two conditions, that we will prove later:

$$\exists c > 0 : \sup_{n \ge 1} \sum_{i=0}^{n-1} |a_{n,i}| < c < \infty, \tag{3.207}$$

$$\sum_{i=0}^{n-1} a_{n,i} = 1. \tag{3.208}$$

We now write

$$\tilde{S}_n = \sum_{i=0}^{n-1} a_{n,i} X_{n,i} = X + \sum_{i=0}^{n-1} a_{n,i}(X_{n,i} - X),$$

with $X := \int_{\mathbb{R}} h(x)\pi(dx)$. In the last equality here above we have used (3.208). In order to show that

$$\sum_{i=0}^{n-1} a_{n,i}(X_{n,i} - X) \xrightarrow{\mathbb{P}} 0, \tag{3.209}$$

we first prove that, $\forall i \in \{0, ..., n-1\}$, $T_{n,i} = O(\frac{i}{n} T_{n,n})$. Indeed, we clearly have

$$n \min_k \Delta_{n,k} \le T_{n,n} = \sum_{j=0}^{n-1} \Delta_{n,j} \le n \max_k \Delta_{n,k}.$$

Using the first point of Assumption S2 it follows

$$\frac{T_{n,n}}{nc_2} \le \frac{1}{c_2} \max_k \Delta_{n,k} \le \min_k \Delta_{n,k} \le \frac{T_{n,n}}{n} \qquad \text{and so}$$

$$\frac{T_{n,n}}{nc_2} \le \min_k \Delta_{n,k} \le \frac{T_{n,n}}{n}, \qquad \frac{T_{n,n}}{n} \le \max_k \Delta_{n,k} \le \frac{c_2 T_{n,n}}{n}.$$

Hence

$$\frac{i\, T_{n,n}}{nc_2} \le T_{n,i} = \sum_{j=0}^{i-1} \Delta_{n,j} \le \frac{i\, c_2\, T_{n,n}}{n}. \tag{3.210}$$

Now, using ergodic theorem, we know that $\forall \epsilon > 0$ $\exists T_\epsilon > 0$ such that, $\forall T \ge T_\epsilon$,

$$\left|\frac{1}{T}\int_0^T h(X_s)ds - \int_{\mathbb{R}} h(x)\pi(dx)\right| < \epsilon. \tag{3.211}$$

By the equation (3.210), we choose $\eta > 0$, $\eta < 1$ such that, $\forall i \ge \eta n$, $T_{n,i} \ge T_\epsilon$. We can see $\sum_{i=0}^{n-1} a_{n,i}(X_{n,i} - X)$ as $\sum_{i=0}^{\lfloor \eta n \rfloor} a_{n,i}(X_{n,i} - X) + \sum_{i=\lfloor \eta n \rfloor + 1}^{n-1} a_{n,i}(X_{n,i} - X)$. Using (3.207) and (3.211) we get

$$\sum_{i=\lfloor \eta n \rfloor + 1}^{n-1} |a_{n,i}||X_{n,i} - X| \le \epsilon c. \tag{3.212}$$

Concerning $\sum_{i=0}^{\lfloor \eta n \rfloor} a_{n,i}(X_{n,i} - X)$, we use that $|X_{n,i} - X|$ is bounded and that, by its definition, $T_{n,i}$ is upper bounded by $T_{n,\eta n}$. Therefore, using also that $\delta > 0$ and $\delta - 1 < 0$, we have

$$\sum_{i=0}^{\lfloor \eta n \rfloor} |a_{n,i}||X_{n,i} - X| \le c \sum_{i=0}^{\lfloor \eta n \rfloor} \frac{T_{n,\eta n} |\Delta_{n,j-1}^{\delta-1} - \Delta_{n,j}^{\delta-1}|}{\sum_{j=0}^{n-1} \Delta_{n,j}^{\delta}} \le c \frac{n\eta \max_k \Delta_{n,k}}{n(\min_k \Delta_{n,k})^{\delta}} \sum_{i=0}^{\lfloor \eta n \rfloor} |\Delta_{n,j-1}^{\delta-1} - \Delta_{n,j}^{\delta-1}|. \tag{3.213}$$

We use the first and the second point of Assumption S2, getting

$$\sum_{i=0}^{\lfloor \eta n \rfloor} |a_{n,i}||X_{n,i} - X| \le c\eta. \tag{3.214}$$

From (3.212) and (3.214) and the arbitrariness of both ϵ and η it follows that (3.209) holds almost surely and so in probability. If we show (3.207) and (3.208), the lemma is therefore proved. Concerning (3.207), we observe it is enough to study the behavior of $\frac{1}{\sum_{i=0}^{n-1}\Delta_{n,i}^{\delta}}\sum_{i=1}^{n-2}T_{n,i}(\Delta_{n,i-1}^{\delta-1}-\Delta_{n,i}^{\delta-1})$. Indeed if it converges then

$$\sup_{n\geq 1}\sum_{i=1}^{n-1}|a_{n,i}|\leq \sup_{n\geq 1}\sum_{i=1}^{n-2}|a_{n,i}|+\sup_{n\geq 1}|a_{n,n-1}|\leq \sup_{n\geq 1}\sum_{i=1}^{n-2}|a_{n,i}|+c_2<\infty.$$

We focus on $\sum_{i=1}^{n-2}|a_{n,i}|$ and we act like we did in (3.213), using this time that $T_{n,i}\leq T_{n,n}$. We get

$$\sup_{n\geq 1}\sum_{i=1}^{n-2}|a_{n,i}|\leq \sup_{n\geq 1} c\frac{(n-1)\max_k\Delta_{n,k}}{n(\min_k\Delta_{n,k})^{\delta}}\sum_{i=0}^{n-2}|\Delta_{n,i-1}^{\delta-1}-\Delta_{n,i}^{\delta-1}|.$$

Again, using the first and the second point of Assumption S2, we get it is bounded by a constant.
To conclude, we observe that $T_{n,i}=T_{n,i-1}+\Delta_{n,i-1}$ and so it is enough to compute $\sum_{i=1}^{n-1}a_{n,i}$ to get it is equal to 1. □

3.8.3 Proof of Lemma 29

Proof. By the definition of $d(\zeta_n)$, as in law we have that $S_1^\alpha=-S_1^\alpha$, we get $d(\zeta_n)=d(|\zeta_n|)$ and thus we can assume that $\zeta_n>0$. Using a change of variable we obtain

$$d(\zeta_n)=\mathbb{E}[(S_1^\alpha)^2\varphi(S_1^\alpha\zeta_n)]=\int_{\mathbb{R}}z^2\varphi(z\zeta_n)f_\alpha(z)dz=(\zeta_n)^{-3}\int_{\mathbb{R}}u^2\varphi(u)f_\alpha(\frac{u}{\zeta_n})du. \tag{3.215}$$

We want to use an asymptotic expansion of the density (see Theorem 7.22 in [55], with $d=1$ and $\sigma=1$) which states that, if z is big enough, then a development up to order N of $f_\alpha(z)$ is

$$\frac{c_\alpha}{|z|^{1+\alpha}}+\frac{1}{\pi}\frac{1}{|z|}\sum_{k=2}^{N}\frac{a_k}{k!}(|z|^{-\alpha})^k+o_{\mathbb{P}}(|z|^{-\alpha N}), \tag{3.216}$$

for some coefficients a_k. We therefore take $M>0$ big enough such that, for $\frac{u}{\zeta_n}>M$, we can use (3.216). Hence the right hand side of (3.215) can be seen as

$$(\zeta_n)^{-3}\int_{|u|\leq\zeta_n M}u^2\varphi(u)f_\alpha(\frac{u}{\zeta_n})du+(\zeta_n)^{-3}\int_{|u|>\zeta_n M}u^2\varphi(u)f_\alpha(\frac{u}{\zeta_n})du=:I_1^n+I_2^n. \tag{3.217}$$

We have that, $\forall\hat{\epsilon}>0$, $I_1^n=o_{\mathbb{P}}(\zeta_n^{-\hat{\epsilon}})$. Indeed, using that φ and f_α are both bounded, we get

$$\frac{I_1^n}{\zeta_n^{-\hat{\epsilon}}}\leq\zeta_n^{-3+\hat{\epsilon}}\int_{|u|\leq\zeta_n M}u^2du\leq c\zeta_n^{\hat{\epsilon}}, \tag{3.218}$$

that goes to zero because we have assumed that $\zeta_n\to 0$. I_2^n is

$$(\zeta_n)^{-3}\int_{|u|>\zeta_n M}u^2\varphi(u)c_\alpha(\zeta_n)^{1+\alpha}|u|^{-1-\alpha}du+ \tag{3.219}$$

$$+(\zeta_n)^{-3}\int_{|u|>\zeta_n M}u^2\varphi(u)[f_\alpha(\frac{u}{\zeta_n})-\frac{c_\alpha}{|u|^{1+\alpha}}|\zeta_n|^{1+\alpha}]du.$$

The first term of 3.219 can be seen as

$$(\zeta_n)^{\alpha-2}c_\alpha\int_{\mathbb{R}}|u|^{1-\alpha}\varphi(u)du-(\zeta_n)^{\alpha-2}c_\alpha\int_{|u|\le\zeta_nM}|u|^{1-\alpha}\varphi(u)du=$$

$$=(\zeta_n)^{\alpha-2}c_\alpha\int_{\mathbb{R}}|u|^{1-\alpha}\varphi(u)du+o_{\mathbb{P}}((\zeta_n)^{-\hat{\epsilon}}).$$

Indeed, using that φ is bounded, we have

$$\frac{1}{(\zeta_n)^{-\hat{\epsilon}}}|(\zeta_n)^{\alpha-2}c_\alpha\int_{|u|\le\zeta_nM}|u|^{1-\alpha}\varphi(u)du|\le c(\zeta_n)^{\hat{\epsilon}+\alpha-2}\int_{|u|\le\zeta_nM}|u|^{1-\alpha}du\le c(\zeta_n)^{\hat{\epsilon}}, \tag{3.220}$$

that goes to zero for $n\to\infty$.
Replacing (3.218), (3.219) and (3.220) in (3.217) and comparing it with (3.97), it turns out that our goal is to show that the second term of (3.219) is $o_{\mathbb{P}}(\zeta_n^{(-\hat{\epsilon})\wedge(2\alpha-2-\hat{\epsilon})})$. Using on it (3.216) with $N=2$, which implies $|f_\alpha(z)-\frac{c_\alpha}{|z|^{1+\alpha}}|\le\frac{c}{|z|^{1+2\alpha}}$ for $|z|>M$ and some $c>0$, we can upper bound it with $c(\zeta_n)^{2\alpha-2}\int_{|u|\le\zeta_nM}|u|^{1-2\alpha}du$. By the definition of φ we have

$$\int_{|u|>\zeta_nM}|u|^{1-2\alpha}\varphi(u)du=\int_{-2}^{-\zeta_nM}(-u)^{1-2\alpha}\varphi(u)du+\int_2^{\zeta_nM}u^{1-2\alpha}\varphi(u)du\le c+c(\zeta_n)^{2-2\alpha}. \tag{3.221}$$

Therefore we get that the second term of (3.219) is upper bounded by

$$c\zeta_n^{2\alpha-2}+c.$$

The first term here above is clearly $o_{\mathbb{P}}(\zeta_n^{2\alpha-2-\hat{\epsilon}})$ while the second is $o_{\mathbb{P}}(\zeta_n^{-\hat{\epsilon}})$, hence the sum is $o_{\mathbb{P}}(\zeta_n^{(-\hat{\epsilon})\wedge(2\alpha-2-\hat{\epsilon})})$. The lemma is therefore proved. □

3.8.4 Proof of Lemma 33

Proof. We observe that, $\forall\alpha\in[0,2]$, we have

$$\mathbb{E}[|\Delta X_i^J-\Delta\tilde{X}_i^J|^2]=\mathbb{E}[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}[\gamma(X_{s-})-\gamma(X_{t_i})]z\tilde{\mu}(ds,dz))^2]=$$

$$=\mathbb{E}[\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}[\gamma(X_{s-})-\gamma(X_{t_i})]^2|z|^2\bar{\mu}(ds,dz)]\le$$

$$\le c\int_{t_i}^{t_{i+1}}\mathbb{E}[|X_s-X_{t_i}|^2]ds\int_{\mathbb{R}}|z|^2F(z)dz\le c\int_{t_i}^{t_{i+1}}\Delta_{n,i}ds\le c\Delta_{n,i}^2, \tag{3.222}$$

where we have used Ito isometry, the regularity of γ and the first point of Lemma 25.
We have in this way proved (3.162) and showed that (3.163) holds with $q=2$. For $q>2$, using Kunita inequality and acting like we did here above we get

$$\mathbb{E}[|\Delta X_i^J-\Delta\tilde{X}_i^J|^q]\le\mathbb{E}[\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}[\gamma(X_{s-})-\gamma(X_{t_i})]^q|z|^q\bar{\mu}(ds,dz)]+$$

$$+\mathbb{E}[(\int_{t_i}^{t_{i+1}}\int_{\mathbb{R}}[\gamma(X_{s-})-\gamma(X_{t_i})]^2|z|^2\bar{\mu}(ds,dz))^{\frac{q}{2}}]\le c\int_{t_i}^{t_{i+1}}\mathbb{E}[|X_s-X_{t_i}|^q]ds+$$

$$+\mathbb{E}[(\int_{t_i}^{t_{i+1}}|X_s-X_{t_i}|^2ds)^{\frac{q}{2}}]\le c\Delta_{n,i}^2+c\mathbb{E}[\Delta_{n,i}^{\frac{q}{2}-1}\int_{t_i}^{t_{i+1}}|X_s-X_{t_i}|^qds]=$$

$$= c\Delta_{n,i}^2 + c\Delta_{n,i}^{\frac{q}{2}-1} \le c\Delta_{n,i}^2,$$

where we have also used Jensen inequality.
In order to prove (3.163) we observe that, if $\alpha < 1$, then we have

$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|] \le \mathbb{E}[|\int_{t_i}^{t_{i+1}} \int_{|z|\ge 2\Delta_{n,i}^\beta} [\gamma(X_{s-}) - \gamma(X_{t_i})] z \tilde{\mu}(ds, dz)|] + \tag{3.223}$$

$$+\mathbb{E}[|\int_{t_i}^{t_{i+1}} \int_{|z|\le 2\Delta_{n,i}^\beta} [\gamma(X_{s-}) - \gamma(X_{t_i})] z \tilde{\mu}(ds, dz)|].$$

The first term in the right hand side of (3.223) is upper bounded by

$$\|\gamma'\|_\infty \mathbb{E}[\int_{t_i}^{t_{i+1}} \int_{|z|\ge 2\Delta_{n,i}^\beta} |X_{s-} - X_{t_i}||z|F(z)dzds] \le$$

$$\le c\int_{t_i}^{t_{i+1}} \int_{|z|\ge 2\Delta_{n,i}^\beta} \mathbb{E}[|X_{s-} - X_{t_i}|^2]^{\frac{1}{2}} ds|z|F(z)dz \le$$

$$\le c\int_{t_i}^{t_{i+1}} \Delta_{n,i}^{\frac{1}{2}} (\int_{|z|\ge 2\Delta_{n,i}^\beta} |z|F(z)dz)ds \le c\Delta_{n,i}^{\frac{3}{2}}, \tag{3.224}$$

where we have used the compensation formula, the regularity of γ, Cauchy-Schwartz inequality in order to use the first point of Lemma 25 and the boundedness of the integral for $|z| \ge 2\Delta_{n,i}^\beta$. Moreover, acting in the same way, the second term in the right hand side of (3.223) is upper bounded by

$$\|\gamma'\|_\infty \mathbb{E}[\int_{t_i}^{t_{i+1}} \int_{|z|\le 2\Delta_{n,i}^\beta} |X_{s-} - X_{t_i}||z|F(z)dzds] \le \tag{3.225}$$

$$\le c\int_{t_i}^{t_{i+1}} \Delta_{n,i}^{\frac{1}{2}} (\int_{|z|\ge 2\Delta_{n,i}^\beta} |z|^{-\alpha} dz)ds \le c\Delta_{n,i}^{\frac{3}{2}+\beta(1-\alpha)},$$

using again compensation formula, the regularity of γ and Cauchy-Schwartz inequality in order to use the first point of Lemma 25. We have also used the third point of Assumption 4 and computed the integral on z. Using (3.223) - (3.225) we get

$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|] \le c\Delta_{n,i}^{\frac{3}{2}\wedge[\frac{3}{2}+\beta(1-\alpha)]} = c\Delta_{n,i}^{\frac{3}{2}}, \tag{3.226}$$

since $\alpha < 1$ and so $(1-\alpha) > 0$. We now use interpolation theorem (see below Theorem 1.7 in Chapter 4 of [12]) getting

$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^q]^{\frac{1}{q}} \le \mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|]^\theta (\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^2]^{\frac{1}{2}})^{1-\theta},$$

with $\frac{1}{q} = \theta + \frac{1-\theta}{2}$, hence $\theta = \frac{2}{q} - 1$. Using (3.222) and (3.226) it follows

$$\mathbb{E}[|\Delta X_i^J - \Delta \tilde{X}_i^J|^q]^{\frac{1}{q}} \le c\Delta_{n,i}^{\frac{3}{2}\theta}\Delta_{n,i}^{1-\theta} = c\Delta_{n,i}^{\frac{1}{2}\theta+1} = c\Delta_{n,i}^{\frac{1}{q}+\frac{1}{2}},$$

where we have also replaced θ. □

3.8.5 Proof of Lemma 34

Proof. We want to use a conditional version of the interpolation theorem, therefore we have to estimate the norm 2 of $I_2^n(i)$, $I_3^n(i)$ and $I_4^n(i)$. Observing that φ is a bounded function and using Kunita inequality we get

$$\mathbb{E}_i[|I_2^n(i)|^2] \le \mathbb{E}_i[|\Delta X_i^J|^4] \le c\mathbb{E}_i[\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} |\gamma(X_{s-})|^4 |z|^4 \bar{\mu}(ds,dz)] +$$

$$+ c\mathbb{E}_i[(\int_{t_i}^{t_{i+1}} \int_{\mathbb{R}} |\gamma(X_{s-})|^2 |z|^2 \bar{\mu}(ds,dz))^2] \le$$

$$\le c(\int_{\mathbb{R}} |z|^4 F(z)dz)\mathbb{E}_i[\int_{t_i}^{t_{i+1}} |\gamma(X_{s-})|^4 ds] + c\mathbb{E}_i[(\int_{\mathbb{R}} |z|^2 F(z)dz)^2 (\int_{t_i}^{t_{i+1}} |\gamma(X_{s-})|^2 ds)^2] \le$$

$$\le R(\Delta_{n,i}, X_{t_i}) + R(\Delta_{n,i}^2, X_{t_i}) = R(\Delta_{n,i}, X_{t_i}), \tag{3.227}$$

where in the last inequality we have also used the polynomial growth of γ and the third point of Lemma 25.

Concerning the norm 2 of $I_3^n(i)$, we use the conditional version of the first point of Lemma 33 for $q = 2$ to get

$$\mathbb{E}_i[|I_3^n(i)|^2] \le \mathbb{E}_i[|\Delta X_i^J - \Delta \tilde{X}_i^J|^4] \le R(\Delta_{n,i}^2, X_{t_i}). \tag{3.228}$$

We now consider $I_4^n(i)$. Using Cauchy-Schwartz inequality and a conditional version of both the first point of Lemma 33 for $q = 2$ and (3.32) in Lemma 28, where φ acts like the indicator function, we have

$$\mathbb{E}_i[|I_4^n(i)|^2]^{\frac{1}{2}} \le c\mathbb{E}_i[|\Delta X_i^J - \Delta \tilde{X}_i^J|^4]^{\frac{1}{2}} \mathbb{E}_i[|\Delta \tilde{X}_i^J \varphi_{\Delta_{n,i}^\beta}(\Delta \tilde{X}_i^J)|^4]^{\frac{1}{2}} \le R(\Delta_{n,i}^{\frac{3}{2}+\frac{\beta}{2}(4-\alpha)}, X_{t_i}). \tag{3.229}$$

Using interpolation theorem it follows, $\forall j \in \{2,3,4\}$,

$$\mathbb{E}_i[|I_j^n(i)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le \mathbb{E}_i[|I_j^n(i)|]^\theta (\mathbb{E}_i[|I_j^n(i)|^2]^{\frac{1}{2}})^{1-\theta}, \tag{3.230}$$

with θ such that $\frac{1}{1+\epsilon} = \theta + \frac{1-\theta}{2}$, hence $\theta = \frac{2}{1+\epsilon} - 1 = 1 - \frac{2\epsilon}{1+\epsilon}$.

From a conditional version of (3.159), (3.164), (3.169) and equations (3.227) and (3.230) it follows

$$\mathbb{E}_i[|I_2^n(i)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le R(\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}}, X_{t_i})^\theta R(\Delta_{n,i}^{\frac{1}{2}}, X_{t_i})^{1-\theta} =$$

$$= R(\Delta_{n,i}^{(\frac{3}{2}+\beta-\frac{\alpha\beta}{2})(1-\frac{2\epsilon}{1+\epsilon})+\frac{\epsilon}{1+\epsilon}}, X_{t_i}) = R(\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}-\frac{\epsilon}{1+\epsilon}(2+2\beta-\alpha\beta)}, X_{t_i}). \tag{3.231}$$

Since $2 + 2\beta - \alpha\beta$ is always more than zero we can just see the exponent on $\Delta_{n,i}$ as $\frac{3}{2} + \beta - \frac{\alpha\beta}{2} - \epsilon$.

From a conditional version of (3.171), (3.228) and (3.230) it follows

$$\mathbb{E}_i[|I_3^n(i)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le R(\Delta_{n,i}^2, X_{t_i})^\theta R(\Delta_{n,i}, X_{t_i})^{1-\theta} = R(\Delta_{n,i}^{1+\theta}, X_{t_i}) = R(\Delta_{n,i}^{2-\frac{2\epsilon}{1+\epsilon}}, X_{t_i}). \tag{3.232}$$

In the same way, using a conditional version of (3.173), (3.229) and (3.230) it follows

$$\mathbb{E}_i[|I_4^n(i)|^{1+\epsilon}]^{\frac{1}{1+\epsilon}} \le R(\Delta_{n,i}^{(\frac{3}{2}+\beta-\frac{\alpha\beta}{2})(1-\frac{2\epsilon}{1+\epsilon})+\frac{2\epsilon}{1+\epsilon}(\frac{3}{2}+2\beta-\frac{\alpha\beta}{2})}, X_{t_i}) = R(\Delta_{n,i}^{\frac{3}{2}+\beta-\frac{\alpha\beta}{2}+\frac{2\beta\epsilon}{1+\epsilon}}, X_{t_i}). \tag{3.233}$$

The result (3.184) is a consequence of (3.231), (3.232), (3.233) and that 2 is always more than $\frac{3}{2} + \beta - \frac{\alpha\beta}{2}$. □

Part III

Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes.

Introduction

The third part of the thesis deals with the adaptive estimation of the invariant measure. We consider the solution X of the multi dimensional stochastic differential equation with jumps (1), with unique invariant probability measure and associated density. We assume that a continuous record $X^T = (X_t)_{0 \le t \le T}$ of the observations is available.
Dalalyan and Reiss in [25] and Strauch in [90] characterize the minimax rate for the estimation of the invariant law of a continuous d-dimensional diffusion in the isotropic and anisotropic cases, respectively. Such a rate depends on the dimension d and on the smoothness of the invariant measure.

We extend the previously mentioned works by obtaining, in presence of jumps, some estimators whose rate is the same it was in the continuous case for $d \geq 2$ and a rate that depends on the jump index α for $d = 1$.
We propose moreover a bandwidth selection procedure for our Kernel estimator, based on the method introduced by Goldenshluger and Lepski in [39], which leads us to non-parametric adaptive estimator of the invariant density of the multivariate diffusion with jumps. The chapter is based on the paper "Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes" [6], under revision for Journal of Statistical Planning and Inference.

Chapter 4

Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes.

Abstract :

We consider the solution $X = (X_t)_{t\geq 0}$ of a multivariate stochastic differential equation with Levy-type jumps and with unique invariant probability measure with density μ. We assume that a continuous record of observations $X^T = (X_t)_{0\leq t\leq T}$ is available.
In the case without jumps, Dalalyan and Reiss [25] and Strauch [90] have found convergence rates of invariant density estimators, under respectively isotropic and anisotropic Hölder smoothness constraints, which are considerably faster than those known from standard multivariate density estimation.
We extend the previous works by obtaining, in presence of jumps, some estimators which have the same convergence rates they had in the case without jumps for $d \geq 2$ and a rate which depends on the degree of the jumps in the one-dimensional setting.
We propose moreover a data driven bandwidth selection procedure based on the Goldenshluger and Lepski method [39] which leads us to an adaptive non-parametric kernel estimator of the stationary density μ of the jump diffusion X.

Keys words : Adaptive bandwidth selection, anisotropic density estimation, ergodic diffusion with jumps, Lévy driven SDE

4.1 Introduction

Diffusion phenomena arise from a Markovian stochastic modeling and as a solution of SDEs with or without jumps in many areas of applied mathematics. Their investigation concerns different mathematical branches and therefore research interest in questions such as existence and regularity of solutions of stochastic differential equations has constantly grown over the past years.
The study of the statistical properties of diffusion models has emerged since such models are widely used for applications in finance and biology. Diffusion processes with jumps, in particular, have been used in neuroscience for instance in [26] while in finance they have been introduced to model the dynamic of asset prices [56], [70], exchange rates [11], or volatility processes [8].

In this chapter, we aim at estimating adaptively the invariant density μ associated to the process $(X_t)_{t\geq 0}$, solution of the following multivariate stochastic differential equation with Levy-type jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d\setminus\{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds, dz), \tag{4.1}$$

where W is a d-dimensional Brownian motion and $\bar{\mu}$ a compensated Poisson random measure with a possible infinite jump activity. We assume that a continuous record of observations $X^T = (X_t)_{0\leq t\leq T}$ is available.

Practical concerns raise new questions such as the dependence of statistical features on the observation scheme: it is, for the applications, a subject of interest to consider basic questions in different observation scenarios. From a theoretical point of view, it is however also of substantial interest to work under the assumption that a continuous record of the diffusion considered is available.
In this framework, it belongs to the folklore of the statistics for stochastic processes without jumps that the invariant density can be estimated under standard nonparametric assumptions with a parametric rate (cfr Chapter 4.2 in [58]). The proof relies on the existence of diffusion local time and its properties and so such a result is restricted to the one - dimensional setting.
In the mono -dimensional case Schmisser considers, in [83], a strictly stationary and β mixing process $(X_t)_{t\geq 0}$ observed at discrete times $t = 0, \Delta_n, ..., n\Delta_n$. The author estimates the successive derivatives $\mu^{(j)}$ of the stationary density either on a compact set or on $\mathbb{R}$ thanks to a penalized least square method. If the derivative $\mu^{(j)}$ belongs to the Besov space $\mathcal{B}^{\alpha}_{2,\infty}$, then the L^2 risk of the estimator converges with rate $(n\Delta_n)^{\frac{-2\alpha}{2\alpha+2j+1}}$, and the procedure does not require the knowledge of α. When $j = 0$, the invariant density associated to the process $(X_t)_{t\geq 0}$ is estimated with a convergence rate which is the same found by Comte and Merlevède in [21] and [22].

Regarding the literature on statistical properties of multidimensional diffusion processes in the continuous case, an important reference is given by Dalalyan and Reiss in [25], where they show an asymptotic statistical equivalence for inference on the drift in the multidimensional diffusion case. As a by-product of the study they prove, under isotropic Hölder smoothness constraints, convergence rates of invariant density estimators for pointwise estimation which are faster than those known from standard multivariate density estimation. Their result relies on upper bounds on the variance of additive diffusion functionals, proven by an application of the spectral gap inequality in combination with a bound on the transition density of the process. Still in the continuous case, in a recent paper, Strauch [90] has extended their work

by building adaptive estimators in the multidimensional diffusion case which achieve fast rates of convergence over anisotropic Hölder balls. In both [25] and [90] the drift function is supposed to be in the form $b = -\nabla V$, where the function V is referred to as potential.
The notion of anisotropy plays an important role. Indeed, the smoothness properties of elements of a function space may depend on the chosen direction of $\mathbb{R}^d$.
The Russian school considered anisotropic spaces from the beginning of the theory of function spaces in 1950-1960s (in [76] the author takes account of the developments). However, results on minimax rates of convergence in classical statistical models were rare for a lot of time.
The question of optimal bandwidth selection based on iid observations for density estimation with respect to sup - norm risk was not completely solved until the pretty recent developments gathered in [60]. The methodology detailed in Goldenshluger and Lepski [39] inspired the data-driven selection procedure of the bandwidth of the kernel estimator proposed by many authors such as Strauch in [90] and Comte, Prieur and Samson in [23] and provides the starting point for the study of our adaptive procedure as well.

In presence of jumps, we are only aware of a few works which takes place in the non parametric framework. For example, Schmisser investigates in [84] the non parametric adaptive estimation of the coefficients of a jumps diffusion process and in [34], together with Funke, the non parametric adaptive estimation of the drift of an integrated jump diffusion process.

In this chapter, we provide a non-parametric estimator of the invariant density μ with a fully data-driven procedure of the bandwidth. We propose to estimate the invariant density μ by means of a kernel estimator, we therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$. A natural estimator of μ at $x \in \mathbb{R}^d$ in the anisotropic context is given by

$$\hat{\mu}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m}) du,$$

where $h = (h_1, ..., h_d)$ is a multi - index bandwidth, which will be chosen through the data-driven selection procedure. We first prove some bounds on the transition semigroup and on the transition density that will be useful to find sharp upper bounds on the variance of integral functionals of the diffusion X. Through them, we find the following convergence rates for the pointwise estimation of the invariant density of our diffusion with jumps:

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \begin{cases} \frac{(\log T)^{(2-\frac{(1+\alpha)}{2}) \vee 1}}{T} & \text{for } d = 1, \\ \frac{\log T}{T} & \text{for } d = 2, \\ T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} & \text{for } d \geq 3, \end{cases}$$

where $\alpha \in (0, 2)$ is the degree of jumps activity of the Lévy process and $\bar{\beta}$ is the harmonic mean smoothness of the invariant density over the d different dimensions. We remark that the rate we find for $d \geq 3$ is the same Strauch found in [90] in absence of jumps, which is also the rate gathered in [25] up to replacing the mean smoothness with β, the common smoothness over the d dimensions.
The case $d = 1$ evidences the main difference between what happens with and without jumps. Indeed, if in the continuous case the optimal convergence rate was

$\frac{1}{T}$, now the rate we found is between $\frac{\log T}{T}$ and $\frac{(\log T)^{\frac{3}{2}}}{T}$. It is worth noting here that such a convergence rate is not necessarily the optimal one in the jumps framework. As a matter of fact in the continuous case different approaches, as the diffusion local time, have been used to get the rate $\frac{1}{T}$; we do not exclude the possibility that also in presence of jumps the implementation of other methods could lead to a convergence rate faster than the one presented here above for the mono-dimensional setting.
To complete the comparison to the continuous framework, we recall that in both [25] and [90] the convergence rate found in the case $d = 2$ was $\frac{(\log T)^4}{T}$ and so the convergence of the estimator seems being faster in presence of jumps than without them. The reason why it happens is that, to find the convergence rate, the transition density $(p_t)_{t\in\mathbb{R}^+}$ is needed to be upper bounded. If in [25] the authors assume to have $p_t(x,y) \le c(t^{-\frac{d}{2}} + t^{\frac{3d}{2}})$ and in [90] Nash and Poincaré inequalities lead Strauch to a bound analogous to the one presented in [25]; Lemma 35 below provides us a different bound which guides us to a different rate. However, in absence of the term $t^{\frac{3d}{2}}$ in the assumption before, which is the case for example considering a bounded drift, also in the continuous setting the convergence rate turns out being, as in the jump -diffusion case, equal to $\frac{\log T}{T}$.
It is moreover worth noting here that, if in [25] and [90] they needed to assume the existence of the transition density and a bound on it, we derive them through Lemma 35: all the assumptions we need are directly on the model (4.1).
We no longer need to assume that the drift is of the form $b = -\nabla V$ (where $V \in \mathcal{C}^2$ is referred to as potential) as it was in both [25] and [90].

After having provided the rates of convergence of the estimators we finally propose, in the case $d \ge 3$, a fully data-driven selection procedure of the bandwidth of the kernel estimator, inspired by the methodology detailed in Goldenshluger and Lespki [39]. The method has the decisive advantage of being anisotropic: the bandwidths selected in each direction are in general different, which is coherent with the possibly different regularities with respect to each variable. Finally, we prove that for the selected optimal bandwidth the following estimation holds:

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \le c_1 \inf_{h\in\mathcal{H}_T} (B(h) + V(h)) + c_1 e^{-c_2(\log T)^2}, \tag{4.2}$$

where we have denoted as $\|\cdot\|_A$ the L^2 norm on A, a compact subset of $\mathbb{R}^d$ and as $\mathcal{H}_T$ the set of candidate bandwidths; $B(h)$ is a bias term and $V(h)$ an estimate of the variance bound. We remark that the estimator leads to an automatic trade - off between the bias and the variance: the second term on the right hand side of (4.2) is indeed negligible compared to the first one.
Moreover, as the rate optimal choice $h(T)$ belongs to the set of candidate bandwidths $\mathcal{H}_T$, (4.2) turns out being

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \le c_1 T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2(\log T)^2},$$

where $\bar{\beta}$ is the mean smoothness of the invariant density.

The chapter is organised as follows. We give in Section 2 the assumptions on the process X. In section 4.3 we define the anisotropic Hölder balls and we construct our estimator. Section 4.4 is devoted to the statements of our main results; which will be proven in the two following sections. In particular, we show how we get the convergence rates for the invariant density estimation in Section 4.5 while in Section 4.6 we prove the estimator we find through our bandwidth selection procedure is adaptive. Some technical results are moreover presented in the Appendix.

4.2 Model Assumptions

We consider the question of nonparametric estimation of the invariant density of a d-dimensional diffusion process X, assuming that a continuous record $X^T = \{X_t, 0 \leq t \leq T\}$ up to time T is observed. This diffusion is given as a strong solution of the following stochastic differential equations with jumps:

$$X_t = X_0 + \int_0^t b(X_s)ds + \int_0^t a(X_s)dW_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} \gamma(X_{s^-})z\tilde{\mu}(ds,dz), \quad t \in [0,T], \tag{4.3}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $a : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$; $W = (W_t, t \geq 0)$ is a d-dimensional Brownian motion and μ is a Poisson random measure on $[0,T] \times \mathbb{R}^d$ associated to the Lévy process $L = (L_t)_{t \in [0,T]}$, with $L_t := \int_0^t \int_{\mathbb{R}^d} z\tilde{\mu}(ds,dz)$. The compensated measure is $\tilde{\mu} = \mu - \bar{\mu}$; we suppose that the compensator has the following form: $\bar{\mu}(dt,dz) := F(dz)dt$, where conditions on the Levy measure F will be given later.
The initial condition X_0, W and L are independent.

In what follows, we suppose the following assumptions hold:

A1: *The functions $b(x)$, $\gamma(x)$ and $a(x)$ are globally Lipschitz and, for some $c \geq 1$,*

$$c^{-1}\mathbb{I}_{d\times d} \leq a(x) \leq c\mathbb{I}_{d\times d},$$

where $\mathbb{I}_{d\times d}$ denotes the $d \times d$ identity matrix.
Denoting with $|.|$ and $< .,. >$ respectively the Euclidian norm and the scalar product in $\mathbb{R}^d$, we suppose moreover that there exists a constant $c > 0$ such that, $\forall x \in \mathbb{R}^d$, $|b(x)| \leq c$.

Under Assumption 1 the equation (4.3) admits a unique non-explosive càdlàg adapted solution possessing the strong Markov property, cf [7] (Theorems 6.2.9. and 6.4.6.).

A2 (Drift condition) :
There exist $\tilde{C} > 0$ and $\tilde{\rho} > 0$ such that $< x, b(x) > \leq -\tilde{C}|x|$, $\forall x : |x| \geq \tilde{\rho}$.

We furthermore need the following assumptions on the jumps:

A3 (Jumps) : *1.The Lévy measure F is absolutely continuous with respect to the Lebesgue measure and we denote $F(z) = \frac{F(dz)}{dz}$.*
2. We suppose that there exist $c > 0$ such that for all $z \in \mathbb{R}^d$, $F(z) \leq \frac{c}{|z|^{d+\alpha}}$, with $\alpha \in (0,2)$ and that $supp(F) = \mathbb{R}^d$.
3. The jump coefficient γ is upper bounded, i.e. $\sup_{x\in\mathbb{R}^d} |\gamma(x)| := \gamma_{max} < \infty$. We suppose moreover that, $\forall x \in \mathbb{R}^d$, $Det(\gamma(x)) \neq 0$.
4. If $\alpha = 1$, we require for any $0 < r < R < \infty$ $\int_{r<|z|<R} zF(z)dz = 0$.
5. There exists $\epsilon > 0$ and a constant $c > 0$ such that $\int_{\mathbb{R}^d} |z|^2 e^{\epsilon|z|} F(z)dz \leq c$.

As we will see in Lemma 36 below, Assumption 2 ensures, together with the last points of Assumption 3, the existence of a Lyapunov function. The process X admits therefore a unique invariant distribution π and the ergodic theorem holds.
We assume the invariant probability measure π of X being absolutely continuous

with respect to the Lebesgue measure and from now on we will denote its density as μ: $d\pi = \mu dx$.
For any set $S \subset \mathbb{R}^d$ we define $\mu(S) := \int_S \mu(x)dx$ and, by abuse of notation, we will write $\mu(f) := \mathbb{E}[f(X_0)] = \int_{\mathbb{R}^d} f(x)\mu(x)dx$ for functions $f : \mathbb{R}^d \to \mathbb{R}$.
We define moreover $L^2(\mu) := \left\{ f : \mathbb{R}^d \to \mathbb{R} : \int_{\mathbb{R}^d} |f(x)|^2 \mu(x)dx < \infty \right\}$ and
$L^1(\mu) := \left\{ f : \mathbb{R}^d \to \mathbb{R} : \int_{\mathbb{R}^d} |f(x)| \mu(x)dx < \infty \right\}$.
For each $g \in L^1(\mu)$ we denote as $\|g\|_{L^1(\mu)} := \mu(|g|)$ the L^1 norm with respect to μ on $\mathbb{R}^d$.
The transition semigroup of the process X on $L^1(\mu)$ is $P_t f(x) := \mathbb{E}[f(X_t)|X_0 = x]$.
The transition density is denoted by p_t and it is such that $P_t f(x) = \int_{\mathbb{R}^d} f(y) p_t(x, y)dy$; we will see in Lemma 35 that it exists.

The process X is called β - mixing if $\beta_X(t) = o(1)$ for $t \to \infty$ and exponentially β - mixing if there exists a constant $\gamma > 0$ such that $\beta_X(t) = O(e^{-\gamma t})$ for $t \to \infty$, where β_X is the β - mixing coefficient of the process X as defined in Section 1.3.2 of [27]. We recall that, for a Markov process X with transition semigroup $(P_t)_{t \in \mathbb{R}^+}$ and $\mathcal{L}(X_0) = \eta$, the β - mixing coefficient of X is given by

$$\beta_X(t) := \sup_{s \in \mathbb{R}^+} \int_{\mathbb{R}^d} \|P_t(x, .) - \eta P_{s+t}(x, .)\| \, \eta P_s(dx, .), \tag{4.4}$$

where $\eta P_t = \mathcal{L}(X_t)$ and $\|\lambda\|$ stands for the total variation norm of a signed measure λ.
For the exponential mixing property of general multidimensional diffusions, the reader may consult Theorem 3 of Kusuoka and Yoshida [57] for the α - mixing; Meyn and Tweedie [72], Stramer and Tweedie [89] and Veretennikov [93] for the β - mixing. The mixing property for general diffusions with jumps has been investigated by Masuda in [67].
Now we mention the notion of exponential ergodicity in the sense of [72].

Definition 14. *We say that X is exponentially ergodic if it admits a unique invariant distribution π and additionally if there exist positive constants c and ρ for which, for each f centered under μ,*

$$\|P_t f\|_{L^1(\mu)} \leq c e^{-\rho t} \|f\|_\infty .$$

We will see in Lemma 36 that both the exponential ergodicity and the exponential β - mixing can be derived from our assumptions.

In Lemmas 36 and 35 below we will prove some bounds on the transition semigroup and on the transition density that will be useful to establish tight upper bounds on the variance

$$Var(\int_0^T f(X_s)ds), \qquad f \in L^2(\mu)$$

of integral functionals of the diffusion X.
Bounds of this type were proven before, in [25] (cf. their Proposition 1), by combining estimates based on the spectral gap inequality and on upper bounds on the

transition densities of X. Through them they prove, under isotropic Hölder smoothness constraints, convergence rates of invariant density estimators for pointwise estimation which are considerably faster than those known from standard multivariate density estimation.
We replace the spectral gap inequality with a control from L^1 to L^∞ given by the exponential ergodicity. Moreover, contrary to [25], we don't need to assume that such controls hold true since we get them as consequence of Lemma 35 and 36 below, having required some assumptions only directly on the model (4.3).
In the next section we will construct adaptive estimators for the density in the multidimensional diffusion case with jumps, which achieve 'fast' rates of convergence over anisotropic Hölder balls.

4.3 Construction of the estimator

In several cases, the regularity of some function $g : \mathbb{R}^d \to \mathbb{R}$ depends on the direction in $\mathbb{R}^d$ chosen. We thus work under the following anisotropic smoothness constraints.

Definition 15. *Let $\beta = (\beta_1, ..., \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, ..., \mathcal{L}_d)$, $\mathcal{L}_i > 0$. A function $g : \mathbb{R}^d \to \mathbb{R}$ is said to belong to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ of functions if, for all $i \in \{1, ..., d\}$,*

$$\left\| D_i^k g \right\|_\infty \leq \mathcal{L}_i \qquad \forall k = 0, 1, ..., \lfloor \beta_i \rfloor,$$

$$\left\| D_i^{\lfloor \beta_i \rfloor} g(. + te_i) - D_i^{\lfloor \beta_i \rfloor} g(.) \right\|_\infty \leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \qquad \forall t \in \mathbb{R},$$

for $D_i^k g$ denoting the k-th order partial derivative of g with respect to the i-th component, $\lfloor \beta_i \rfloor$ denoting the largest integer strictly smaller than β_i and $e_1, ..., e_d$ denoting the canonical basis in $\mathbb{R}^d$.

From now on we deal with the estimation of the density μ belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$.

Given the observation X^T of a diffusion X, solution of (4.3), we propose to estimate the invariant density μ by means of a kernel estimator. To estimate some $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ we therefore introduce some kernel function $K : \mathbb{R} \to \mathbb{R}$ satisfying

$$\int_{\mathbb{R}} K(x)dx = 1, \quad \|K\|_\infty < \infty, \quad \text{supp}(K) \subset [-1, 1], \quad \int_{\mathbb{R}} K(x)x^l dx = 0,$$

for all $l \in \{0, ..., M\}$ with $M \geq \max_i \beta_i$.
Denoting by X_t^j, $j \in \{1, ..., d\}$ the j-th component of X_t, $t \geq 0$, a natural estimator of μ at $x = (x_1, ..., x_d)^T \in \mathbb{R}^d$ in the anisotropic context is given by

$$\hat{\mu}_{h,T}(x) = \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m}) du. \tag{4.5}$$

As we will see in Section 4.4.2, a main question concerns the choice of the multi-index bandwidth $h = (h_1, ..., h_d)^T$.

4.4 Main results

4.4.1 Convergence rates for invariant density estimation

We want to investigate on the convergence rates for invariant density estimation. In order to determine the asymptotic behaviour of our estimator for $T \to \infty$, we study the variance of general additive functionals of X in d dimension. To do so, we need some properties as the exponential ergodicity of the process and a bound on the transition density. Such properties will be derived from our assumptions through the following lemmas, that we will prove in the appendix.
The following bounds on the transition density and on the transition semigroup hold true.

Lemma 35. *Suppose that A1 - A3 hold. Then, for $T \geq 0$, there exists a transition density $p_t(x, y)$ for which for any $t \in [0, T]$ there are a $c_0 > 0$ and a $\lambda_0 > 0$ such that, for any pair of points $x, y \in \mathbb{R}^d$, we have*

$$p_t(x, y) \leq c_0 (t^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{t}} + \frac{t}{(t^{\frac{1}{2}} + |y-x|)^{d+\alpha}}).$$

Lemma 36. *Suppose that A1 - A3 hold. Then the process X is exponentially ergodic and exponentially β - mixing.*

On the basis of the two previous lemmas we can prove the following bound on the variance, which is the heart of the study on the convergence rate.

Proposition 23. *Suppose that A1 - A3 hold and let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded, measurable function with support $\mathcal{S}$ satisfying $|\mathcal{S}| < 1$. Then, there exists a constant C independent of f such that*

- $Var(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |\mathcal{S}|^2 (1 + (\log(\frac{1}{|\mathcal{S}|}))^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{|\mathcal{S}|}))$ *for $d = 1$,*
- $Var(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |\mathcal{S}|^2 (1 + \log(\frac{1}{|\mathcal{S}|}))$ *for $d = 2$,*
- $Var(\int_0^T f(X_t)dt) \leq CT \|f\|_\infty^2 |\mathcal{S}|^{1+\frac{2}{d}}$ *for $d \geq 3$.*

From the bias - variance decomposition in the anisotropic case (see Proposition 1 in [20]) we get the following bound

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^d h_l^{2\beta_l} + T^{-2} Var(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_t^m}{h_m})dt).$$

We want to bound the variance here above using Proposition 23 on the function $f(y) := \frac{1}{\prod_{l=1}^d h_l} \prod_{m=1}^d K(\frac{x_m - y_m}{h_m})$. As it will be explained in the proof of Proposition 24 in Section 4.5, for $d \geq 3$ it is

$$Var(\frac{1}{\prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K(\frac{x_m - X_t^m}{h_m})dt) \leq cT (\prod_{l=1}^d h_l)^{\frac{2}{d}-1},$$

which leads us to the following convergence rate.

Proposition 24. *Suppose that A1 - A3 hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in* (4.5) *satisfies, for $d \geq 3$, the following risk estimates:*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \sum_{l=1}^{d} h_l^{2\beta_l} + T^{-1}(\prod_{l=1}^{d} h_l)^{\frac{2}{d}-1}. \tag{4.6}$$

Defined $\frac{1}{\bar{\beta}} := \frac{1}{d}\sum_{l=1}^{d} \frac{1}{\beta_l}$, the rate optimal choice $h_l = h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}}$ yields the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim T^{-\frac{2\bar{\beta}}{2\bar{\beta}+d-2}}.$$

We underline that, in the continuous case, the convergence rate found by Strauch in [90] for the estimation of the invariant density μ belonging to the anisotropic Hölder class $\mathcal{H}_d(\beta + 1, \mathcal{L})$ is $T^{-\frac{2(\overline{\beta+1})}{2(\overline{\beta+1})+d-2}}$, for $d \geq 3$.
In Proposition 24 we estimate μ over anisotropic Hölder class $\mathcal{H}_d(\beta, \mathcal{L})$ and we therefore extend [90] to the jumps - diffusion case: the convergence rate we obtain is the same it was in the case without jumps, which is also analogous to the rate first obtained by Dalalyan and Reiss in [25] for the estimation of the invariant density μ over isotropic Hölder class $\mathcal{H}_d(\beta+1, \mathcal{L})$, up to replacing the mean smoothness $\overline{\beta + 1}$ with $\beta + 1$, the common smoothness over the d different dimensions.

For $d = 1$ and $d = 2$, the bound on the variance changes. Therefore, the rate optimal choice h will be different as well, as explained in following two propositions.

Proposition 25. *Suppose that A1 - A3 hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in* (4.5) *satisfies, for $d = 1$, the following risk estimates:*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h^{2\beta} + \frac{1}{T}(1 + (\log(\frac{1}{h}))^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{h})). \tag{4.7}$$

The rate optimal choice for h yields to the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{(\log T)^{(2-\frac{(1+\alpha)}{2})\vee 1}}{T}.$$

It is worth remarking that, in Proposition 25, it is stated the main difference between the case with and without jumps. Indeed, if in the continuous case the convergence rate was $\frac{1}{T}$, now it depends on the degree of the jumps α and it is between $\frac{\log T}{T}$ and $\frac{(\log T)^{\frac{3}{2}}}{T}$.
We need to say that the convergence rate we have found here above for the estimation of the invariant density of a stochastic differential equations with jumps in the one dimensional setting is not necessarily the optimal one. In the continuous case other methods have been explored for such an estimation when $d = 1$, as the use of diffusion local time to get the optimal rate $\frac{1}{T}$. We do not rule out the possibility to get a sharper bound through the exploitation of other approaches also for the jumps case, finding therefore a convergence rate faster than the one presented in the previous proposition.

Proposition 26. *Suppose that A1 - A3 hold. If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then the estimator given in* (4.5) *satisfies, for* $d = 2$, *the following risk estimates:*

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim h_1^{2\beta_1} + h_2^{2\beta_2} + \frac{1}{T}(1 + \log(\frac{1}{h_1 h_2})). \tag{4.8}$$

The rate optimal choice for h yields to the convergence rate

$$\mathbb{E}[|\hat{\mu}_{h,T}(x) - \mu(x)|^2] \lesssim \frac{\log T}{T}.$$

Comparing our result with the convergence rate obtained in the continuous case over isotropic Hölder class $\mathcal{H}_d(\beta + 1, \mathcal{L})$ in [25] and anisotropic Hölder class $\mathcal{H}_d(\beta + 1, \mathcal{L})$ in [90], which is $\frac{(\log T)^4}{T}$ in both works, one can observe that the convergence rate seems being faster in presence of jumps.
The reason why it happens is that in [25] they assume the transition density to be upper bounded by $C(t^{-\frac{d}{2}} + t^{\frac{3d}{2}})$, which is a bound different from the one we get from Lemma 35.
If the term $t^{\frac{3d}{2}}$ would have been absent in their assumption, e. g. for bounded drift, then the convergence rate in the continuous case could have been improved to $\frac{\log T}{T}$, which is also what we get in the jump- diffusion case.
In [90], Nash and Poincaré inequalities lead the author to an upper bound on the transition density which is analogous to the one found in [25] (see Remark 2.4 of [90]).

From the pointwise estimation of the invariant density gathered in the three previous propositions we move to the estimation on $L^2(A)$, where A is a compact set of $\mathbb{R}^d$.
In the sequel, for $A \subset \mathbb{R}^d$ compact and for $g \in L^2(A)$, $\|g\|_A^2 := \int_A |g(x)|^2 dx$ denotes the L^2 norm with respect to Lebesgue on A.
As a consequence of Propositions 24, 25 and 26 and the fact that the constants which turn out in the proofs do not depend on x, the following corollary holds true:

Corollary 2. *If $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$, then for the rate optimal choice for h = h(T) provided in Propositions 24, 25 and 26 we have the following risk estimates:*

$$\mathbb{E}[\|\hat{\mu}_{h,T} - \mu\|_A^2] \lesssim V_d(T) := \begin{cases} \frac{(\log T)^{(2 - \frac{(1+\alpha)}{2}) \vee 1}}{T} & \text{for } d = 1 \\ \frac{\log T}{T} & \text{for } d = 2 \\ T^{-\frac{2\bar{\beta}}{2\bar{\beta} + d - 2}} & \text{for } d \geq 3. \end{cases} \tag{4.9}$$

The proof of Corollary 2 will be given in Section 4.5.

4.4.2 Adaptive procedure

The question of density estimation belongs to the canonical framework of nonparametric statistics.
As detailed in Propositions 25 and 26, both the bandwidth and the upper bound on the rate of convergence appearing on the right hand side of (4.7) and (4.8) do not depend on the unknown smoothness of the invariant density μ and so there is

no gain in implementing a data-driven bandwidth selection procedure for density estimation in the framework of continuous observations of a one or two dimensional diffusion process with jumps. Hence, throughout the sequel we restrict to the case $d \geq 3$.
It is clear from the previous section that for $d \geq 3$, instead, the proposed bandwidth choice depends on the regularity of the density μ, which is unknown. This is why we study a data-driven bandwidth selection device.
We emphasize that the d selected bandwidths are different, and this anisotropy property is important in our setting: the regularity in each direction can be various. The bandwidth selection procedure has to be able to provide such different choices for h_1, h_2, ... , h_d.
To select h adequately, we propose the following method, inspired from Goldenshluger and Lespki [39].
We define the set of candidate bandwidths $\mathcal{H}_T$ as

$$\mathcal{H}_T \subset \left\{ h = (h_1, ..., h_d)^T \in (0,1]^d : \frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^{d} h_l \leq (\frac{1}{\log T})^{\frac{3d}{d-2}} \right\}, \qquad (4.10)$$

The conditions on $\prod_{l=1}^d h_l$ we have just given are needed to use Talagrand inequality, on the basis of which we show our adaptive result.
We suppose moreover that the growth of $|\mathcal{H}_T|$ is at most polynomial in T, which is there exists $c > 0$ for which $|\mathcal{H}_T| \leq cT^c$.
An example of $\mathcal{H}_T$ is the following set of candidate bandwidths:

$$\mathcal{H}_T := \Big\{ h = (h_1, ..., h_d)^T \in (0,1]^d : \; h_i = \frac{1}{k_i} \text{ with } k_i \in \mathbb{N}, \qquad (4.11)$$

$$\frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^{d} \frac{1}{k_l} \leq (\frac{1}{\log T})^{\frac{3d}{d-2}} \Big\} \qquad (4.12)$$

In correspondence of the variation of $h \in \mathcal{H}_T$, we have the following family of estimators, defined as in (4.5)

$$\mathcal{F}(\mathcal{H}_T) := \left\{ \hat{\mu}_h(x) := \frac{1}{T} \int_0^T \mathbb{K}_h(X_u - x) du : \quad x \in \mathbb{R}^d, \quad h \in \mathcal{H}_T \right\}$$

where, for $y \in \mathbb{R}^d$, it is

$$\mathbb{K}_h(y) := \prod_{l=1}^{d} \frac{1}{h_l} \prod_{m=1}^{d} K(\frac{y_m}{h_m}). \qquad (4.13)$$

We aim at selecting an estimator from the family $\mathcal{F}(\mathcal{H}_T)$ in a completely data-driven way, based only on the observation of the continuous trajectory of the process X solution of (4.3).
We now turn to describing the selection procedure from $\mathcal{F}(\mathcal{H}_T)$, which is based on auxiliary estimators relying on the convolution operator. According to our records, it was introduced in [61] as a device to circumvent the lack of ordering among a set of estimators in anisotropic case, where the increase of the variance of an estimator does not imply a decrease of its bias.
For any bandwidths $h = (h_1, ..., h_d)^T$, $\eta = (\eta_1, ..., \eta_d)^T \in \mathcal{H}_T$ and $x \in \mathbb{R}^d$, we define

$$\mathbb{K}_h * \mathbb{K}_\eta(x) := \prod_{j=1}^{d} (K_{h_j} * K_{\eta_j})(x_j) = \prod_{j=1}^{d} \int_{\mathbb{R}} K_{h_j}(u - x_j) K_{\eta_j}(u) du.$$

We moreover define the kernel estimators

$$\hat{\mu}_{h,\eta}(x) := \frac{1}{T}\int_0^T (\mathbb{K}_h * \mathbb{K}_\eta)(X_u - x)du, \quad x \in \mathbb{R}^d.$$

We remark that for how we have defined the kernel estimators, since the convolution is commutative, it is $\hat{\mu}_{h,\eta} = \hat{\mu}_{\eta,h}$.
The proposed selection procedure relies on comparing the differences $\hat{\mu}_{h,\eta} - \hat{\mu}_\eta$.
We define

$$A(h) := \sup_{\eta \in \mathcal{H}_T} (\|\hat{\mu}_{h,\eta} - \hat{\mu}_\eta\|_A^2 - V(\eta))_+, \tag{4.14}$$

with

$$V(h) := \frac{k}{T}(\prod_{l=1}^d h_l)^{\frac{2}{d}-1},$$

where k is a numerical constant which is large. In particular, it is sufficient to choose it bigger than the constants $2k_0^*$ and $2k_0$ which appear in Lemma 38. Even if k is not explicit, it can be calibrated by simulations as done for example in Section 5 of [23] through the implementation of a method inspired by Lacour, Massart and Rivoirard in [59].
Heuristically, $A(h)$ is an estimate of the squared bias and $V(h)$ of the variance bound. It is worth noticing that the penalty term $V(h)$ which is used here comes from Proposition 23 for the function f being the Kernel function.
Thus, the selection is done by setting

$$\tilde{h} := \arg\min_{h \in \mathcal{H}_T} (A(h) + V(h)). \tag{4.15}$$

We introduce the following notation: $\mu_h := \mathbb{K}_h * \mu$, which is the function that is estimated without bias by $\hat{\mu}_h$, i.e. $\mathbb{E}[\hat{\mu}_h(x)] = \mu_h(x)$. Moreover we define $\mu_{h,\eta} := \mathbb{K}_h * \mathbb{K}_\eta * \mu$ and a bias term $B(h) := \|\mu_h - \mu\|_{\tilde{A}}^2$, where we have denoted as $\|.\|_{\tilde{A}}$ the L^2 - norm on $\tilde{A}$, a compact set in $\mathbb{R}^d$ which is such that $\tilde{A} := \left\{\zeta \in \mathbb{R}^d : d(\zeta, A) \leq 2\sqrt{d}\right\}$.
The following result holds.

Theorem 15. *Suppose that assumptions A1 - A3 hold. Then, we have*

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1 \inf_{h \in \mathcal{H}_T} (B(h) + V(h)) + c_1 e^{-c_2(\log T)^2},$$

for c_1 and c_2 positive constants.

The bound stated in Theorem 15 shows that the estimator leads to an automatic trade-off between the bias $\|\mu_h - \mu\|_{\tilde{A}}^2$ and the variance $V(h)$, up to a multiplicative constant c_1. The last term is indeed negligible. The proof of Theorem 15 is postponed to Section 4.6.

We recall that Proposition 24 provides us the rate optimal choice $h(T)$ for $d \geq 3$, which is $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}}$.
Using such a bandwidth we will prove in Section 4.6 the following theorem.

Theorem 16. *Suppose that assumptions A1 - A3 hold and let $\mathcal{H}_T$ be defined by* (4.12)*. Then, we have*

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \leq c_1(\frac{1}{T})^{\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2(\log T)^2},$$

for c_1 and c_2 positive constants.

Underlining once again that the second term in the right hand side of the equation here above is negligible compared to the first, we have that the risk estimates we get using the bandwidth provided by our selection procedure converges to zero fast. In particular, its convergence rate coincides to the optimal one provided by both [25] and [90] in the case without jumps.

4.5 Proof convergence rates for invariant density estimation

In this section we prove Propositions 24, 25 and 26, which gives us the convergence rate for the estimation of the invariant density $\mu \in \mathcal{H}_d(\beta, \mathcal{L})$ in the three different situation: $d = 1$, $d = 2$ and $d \geq 3$.
We emphasize that all the constants will appear in the proofs do not depend on the point x considered.
We start showing the bound on the variance gathered in Proposition 23.

4.5.1 Proof of Proposition 23

Proof. We consider first of all the case $d \geq 3$. We define the function $f_c := f - \mu(f)$. From the symmetry and the stationarity we have

$$Var(\int_0^T f(X_s)ds) = 2\int_0^T \int_0^s \mathbb{E}[f_c(X_s)f_c(X_t)]dtds = 2\int_0^T \int_0^s \mathbb{E}[f_c(X_0)f_c(X_{s-t})]dtds$$

Applying the change if variable $u := s - t$, using Fubini and computing the integral we have that the quantity here above is equal to $2\int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du$. Let now $0 < \delta < D \leq T$, where the specific choice of δ and D will be given later. The idea is to deal with the integral here above in different way for u which is in different intervals. For this reason we see $\int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du$ as

$$\int_0^\delta (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du + \int_\delta^D (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du + \tag{4.16}$$

$$+\int_D^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du.$$

We now observe that

$$\int_0^\delta (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du = \int_0^\delta (T-u)(\mathbb{E}[f(X_0)f(X_u)] - (\mu(f))^2)du \leq \tag{4.17}$$

$$\leq cT\int_0^\delta | < P_u f, f >_\mu |du,$$

where we have denoted as $< .,. >_\mu$ the scalar product deriving by the norm with respect to the measure μ, for which $< g, h >_\mu := \int_{\mathbb{R}^d} g(x)h(x)\mu(x)dx$, each $g, h \in L^2(\mu)$. In the last inequality we have moreover used that $(\mu(f))^2$ is always more than 0.
Now we use Cauchy-Schwartz inequality and the fact that $P_u f$ is a contraction map in $L^2(\mu)$ to get

$$\int_0^\delta | < P_u f, f >_\mu |du \leq \int_0^\delta \sqrt{\|P_u f\|_\mu^2 \, \|f\|_\mu^2} du \leq \int_0^\delta \sqrt{\|f\|_\mu^4} du \leq \|f\|_\infty^2 \, \mu(\mathcal{S})\delta, \tag{4.18}$$

where in the last inequality we have used the estimation

$$\|f\|_\mu^2 = \int_{\mathcal{S}} |f(x)|^2 \mu(x) dx \leq \|f\|_\infty^2 \, \mu(\mathcal{S}).$$

Concerning the second integral in (4.16), we remark that (4.17) still holds on $[\delta, D]$. We then estimate it through the definition of transition semigroup. It is

$$\int_\delta^D | < P_u f, f >_\mu | du \leq \int_\delta^D \int_{\mathbb{R}^d} |f(x)| \int_{\mathbb{R}^d} |f(y)| p_u(x,y) dy \mu(x) dx du. \tag{4.19}$$

We want to use the bound on the transition density given in Lemma 35 which holds for $t \in [0, T]$ but it is not uniform in t big. Nevertheless, for $t \geq 1$, we have

$$p_t(x,y) = \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x, \zeta) p_{\frac{1}{2}}(\zeta, y) d\zeta \leq$$

$$\leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x,\zeta)(e^{-\lambda_0 (y-\zeta)^2 \frac{1}{2}} + \frac{1}{(\sqrt{\frac{1}{2}} + |y - \zeta|)^{d+\alpha}}) d\zeta \leq c \int_{\mathbb{R}^d} p_{t-\frac{1}{2}}(x,\zeta) d\zeta \leq c,$$

where the constant c changes from line to line. The right hand side of (4.19) is therefore upper bounded by

$$\int_\delta^D \int_{\mathbb{R}^d} |f(x)| c \int_{\mathbb{R}^d} |f(y)| (u^{-\frac{d}{2}} e^{-\lambda_0 \frac{|y-x|^2}{u}} + \frac{u}{(u^{\frac{1}{2}} + |y-x|)^{d+\alpha}} + 1) dy \, \mu(x) dx \, du \leq$$

$$\leq \int_\delta^D \int_{\mathcal{S}} |f(x)| c \int_{\mathcal{S}} |f(y)| (u^{-\frac{d}{2}} + u^{1-\frac{(d+\alpha)}{2}} + 1) dy \, \mu(x) dx \, du \leq$$

$$\leq c \, \|f\|_\infty^2 \, \mu(\mathcal{S}) |\mathcal{S}| \int_\delta^D (u^{-\frac{d}{2}} + u^{1-\frac{(d+\alpha)}{2}} + 1) du,$$

where we have bounded in both integrals the absolute value of f with its infinity norm.

Now we want to calculate the integral with respect to the variable u. We observe that, since $d \geq 3$, $1 - \frac{d}{2} < 0$. The exponent of the second term in the integral here above, after having integrated, is $2 - \frac{d+\alpha}{2}$. It is more than zero if $d < 4 - \alpha$, which is possible only if $\alpha \in (0,1)$ and $d = 3$, less then zero otherwise.

Therefore, we have to consider the two different possibilities, according to the fact that the exponent would be positive or negative. It follows

$$\int_\delta^D | < P_u f, f >_\mu | du \leq c \, \|f\|_\infty^2 \, \mu_b(\mathcal{S}) |\mathcal{S}| (\delta^{1-\frac{d}{2}} + \delta^{2-\frac{d+\alpha}{2}} 1_{\{d \geq 4-\alpha\}} + D^{2-\frac{d+\alpha}{2}} 1_{\{d < 4-\alpha\}} + D). \tag{4.20}$$

We are now left to estimate the third integral of (4.16). From Lemma 36 it follows it is upper bounded by

$$cT \int_D^T | < P_u f_c, f_c >_\mu | du \leq cT \int_D^T \sqrt{\|P_u f_c\|_{L^1(\mu)}^2 \, \|f_c\|_\infty^2} du \leq \tag{4.21}$$

$$\leq cT \int_D^T \sqrt{(e^{-\rho u} \, \|f_c\|_\infty)^2 \, \|f_c\|_\infty^2} du.$$

We recall it is $f_c(x) = f(x) - \mu(f)$, where $\mu(f) = \int_{\mathcal{S}} f(x) \mu(x) dx$.
Therefore we have

$$|f_c(x)| \leq |f(x)| + |\mu(f)| \leq |f(x)| + \|f\|_\infty \, \mu(\mathcal{S}) \tag{4.22}$$

and so

$$\|f_c\|_\infty \le \|f\|_\infty + \|f\|_\infty \mu(\mathcal{S}) \le c \|f\|_\infty ,$$

where in the last inequality we have used the following estimation

$$|\mu(\mathcal{S})| \le \|\mu\|_\infty |\mathcal{S}| \le c|\mathcal{S}| \tag{4.23}$$

and the fact that $|\mathcal{S}| < 1$.
Therefore we get

$$cT\int_D^T | < P_u f_c, f_c >_\mu |du \le c\,T \|f\|_\infty^2 \int_D^T e^{-\rho u} du \le c\,T \|f\|_\infty^2 e^{-\rho D}. \tag{4.24}$$

Replacing (4.18), (4.20) and (4.24) in (4.16) we have that

$$| \int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du| \le c\,T \|f\|_\infty^2 |\mathcal{S}|(\delta + |\mathcal{S}|\delta^{1-\frac{d}{2}} + |\mathcal{S}|\delta^{2-\frac{d+\alpha}{2}} 1_{\{d \ge 4-\alpha\}} + \tag{4.25}$$
$$+ |\mathcal{S}|D^{2-\frac{d+\alpha}{2}} 1_{\{d<4-\alpha\}} + |\mathcal{S}|D) + c\,T \|f\|_\infty^2 e^{-\rho D}.$$

We now want to choose δ and D for which the estimation here above is as sharp as possible. Recalling that the exponent on δ are less than zero in the second and the third terms of the right hand side of (4.25), we have that for a small choice of δ would correspond the smallness of the first term while the second and the third would be big, the opposite would hold for a big δ. In the same way, the behaviour of the last two terms of the right hand side of (4.25) relies on the choice D. Aiming at balancing them, we define $\delta := |\mathcal{S}|^{\frac{2}{d}}$ and $D := [\max(-\frac{2}{\rho}\log(|\mathcal{S}|), 1)] \wedge T$. Replacing them in (4.25) if $T > (-\frac{2}{\rho}\log(|\mathcal{S}|))$ we get

$$| \int_0^T (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du| \le$$

$$\le c\,T \|f\|_\infty^2 (|\mathcal{S}|^{1+\frac{2}{d}} + |\mathcal{S}|^{1+\frac{4-\alpha}{d}} + |\mathcal{S}|^2 (log|\mathcal{S}|)^{2-\frac{d+\alpha}{2}} + |\mathcal{S}|^2 log|\mathcal{S}| + |\mathcal{S}|^2),$$

which give us the result we wanted remarking that both 2 and $1 + \frac{4-\alpha}{d}$ are always more than $1 + \frac{2}{d}$ for $d \ge 3$ and $\alpha \in (0, 2)$.
Otherwise, if $T \le (-\frac{2}{\rho}\log(|\mathcal{S}|))$, by the definition of D we obtain $D = T$. We still have $|\mathcal{S}|^2 D^{2-\frac{d+\alpha}{2}} 1_{\{d<4-\alpha\}} \le c|\mathcal{S}|^2(log|\mathcal{S}|)^{2-\frac{d+\alpha}{2}}$ and, moreover, the last integral which we dealt with in (4.24) is in this case between T and T and so its contribution is null. Hence, the result still holds true.

We now consider the case $d = 1$. We can act exactly like we did in the case $d \ge 3$, splitting the integral in three parts. Estimations (4.18) and (4.24) still holds while, using again Lemma 35 on the interval $[\delta, D]$, we have

$$| \int_\delta^D (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du| \le cT \|f\|_\infty^2 \mu(\mathcal{S})|\mathcal{S}| \int_\delta^D (u^{-\frac{1}{2}} + u^{1-\frac{(1+\alpha)}{2}} + 1)du \le$$

$$\le cT \|f\|_\infty^2 \mu(\mathcal{S})|\mathcal{S}|(D^{\frac{1}{2}} + D^{2-\frac{(1+\alpha)}{2}} + D),$$

where we have used that now, integrating, both the exponent we get are positive. In total in the case $d = 1$, using also (4.23), we therefore have

$$Var(\int_0^T f(X_s)ds) \le cT \|f\|_\infty^2 (|\mathcal{S}|\delta + |\mathcal{S}|^2(D + D^{2-\frac{(1+\alpha)}{2}}) + e^{-\rho D}).$$

As we have already done, we want to make the estimation here above as sharp as possible. This time there isn't any constraint on the smallness of δ and so we can choose directly $\delta := 0$. Regarding D we observe that, if $\alpha > 1$, then $2 - \frac{(1+\alpha)}{2} < 1$; otherwise $2 - \frac{(1+\alpha)}{2} > 1$. In each case we have the same trade off we had in the case $d \geq 3$ and so we keep taking $D := [\max(-\frac{2}{\rho}\log(|\mathcal{S}|), 1)] \wedge T$; it follows $Var(\int_0^T f(X_s)ds) \leq cT \left\| f \right\|_\infty^2 |\mathcal{S}|^2(1 + (\log|\mathcal{S}|)^{2-\frac{(1+\alpha)}{2}} + \log(\frac{1}{|\mathcal{S}|}))$.

In the case $d = 2$ estimations (4.18) and (4.24) keep holding true. The bound on the transition density gathered in Lemma 35 for $d = 2$ yields

$$|\int_\delta^D (T-u)\mathbb{E}[f_c(X_0)f_c(X_u)]du| \leq cT \left\| f \right\|_\infty^2 \mu(\mathcal{S})|\mathcal{S}| \int_\delta^D (u^{-1} + u^{1-\frac{(2+\alpha)}{2}} + 1)du \leq$$

$$\leq cT \left\| f \right\|_\infty^2 \mu(\mathcal{S})|\mathcal{S}|(\log(\frac{D}{\delta}) + D^{2-\frac{(2+\alpha)}{2}} + D),$$

having remarked that $2 - \frac{(2+\alpha)}{2} = 1 - \frac{\alpha}{2} > 0$ because $\alpha \in (0,2)$. This entails, using also (4.23),

$$Var(\int_0^T f(X_s)ds) \leq cT \left\| f \right\|_\infty^2 (|\mathcal{S}|\delta + |\mathcal{S}|^2(\log(\frac{D}{\delta}) + D^{2-\frac{(2+\alpha)}{2}} + D) + e^{-\rho D}).$$

Aiming at balancing the terms, we choose again
$\delta := |\mathcal{S}|$ and $D := [\max(-\frac{2}{\rho}\log(|\mathcal{S}|), 1)] \wedge T$.
It follows that $\log(\frac{D}{\delta}) \leq c|\log|\log|\mathcal{S}||| + c|\log(\frac{1}{|\mathcal{S}|})|$ and so, observing that $\log|\log(|\mathcal{S}|)|$ is negligible compared to $\log(\frac{1}{|\mathcal{S}|})$, the bound on the variance becomes

$$Var(\int_0^T f(X_s)ds) \leq cT \left\| f \right\|_\infty^2 |\mathcal{S}|^2(1 + \log(\frac{1}{|\mathcal{S}|})),$$

where we have also used that $2 - \frac{(2+\alpha)}{2}$ is always less than 1 and so $(\log(\frac{1}{|\mathcal{S}|}))^{2-\frac{(2+\alpha)}{2}} < \log(\frac{1}{|\mathcal{S}|})$.
The proposition is therefore proved. □

4.5.2 Proof of Proposition 24

Proof. Estimation (4.6) is a straightforward consequence of the bias - variance decomposition and Proposition 23 applied to $f(y) := \frac{1}{\prod_{l=1}^d h_l}\prod_{m=1}^d K(\frac{x_m - y_m}{h_m})$, whose support $\mathcal{S}$ is such that $|\mathcal{S}| \leq c\prod_{l=1}^d h_l$ and which is by construction such that $\left\| f \right\|_\infty \leq c(\prod_{l=1}^d h_l)^{-1}$.
To find the optimal choice of h we define $h_l(T) := (\frac{1}{T})^{a_l}$ for $l \in \{1, ..., d\}$ and we look for a_1, ... a_d such that the upper bound of the mean-squared error in the right hand side of (4.6) would result as small as possible.
Replacing the definition of $h_l(T)$ in the bias - variance decomposition, it means searching for $a_1, ..., a_d$ for which we get the balance and so we have to resolve the following system:

$$\begin{cases} \beta_i a_i = \beta_{i+1}a_{i+1} & \forall i \in \{1, ..., d-1\} \\ 2\beta_d a_d = 1 + (\frac{2}{d} - 1)\sum_{l=1}^d a_l. \end{cases}$$

We observe that, as a consequence of the first $d-1$ equations, we can write a_l as $\frac{\beta_d}{\beta_l} a_d$ for each $l \in \{1, ..., d-1\}$. Therefore, the last equation becomes

$$2\beta_d a_d = 1 + (\frac{2}{d} - 1)\beta_d a_d \sum_{l=1}^{d} \frac{1}{\beta_l}.$$

Defining $\frac{1}{\bar{\beta}} := \frac{1}{d} \sum_{l=1}^{d} \frac{1}{\beta_l}$, it follows $2\beta_d a_d = 1 + (\frac{2}{d} - 1)\beta_d a_d \frac{d}{\bar{\beta}}$, which yields

$$a_d = \frac{\bar{\beta}}{\beta_d (2\bar{\beta} + (d-2))}$$

and, replacing it in the system, we have

$$a_l = \frac{\bar{\beta}}{\beta_l (2\bar{\beta} + (d-2))} \quad \forall l \in \{1, ..., d\}.$$

Taking in the right hand side of (4.6) the rate optimal choice $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l (2\bar{\beta} + (d-2))}}$ we get the convergence rate wanted. □

4.5.3 Proof of Proposition 25

Proof. The upper bound of the mean-squared error follows from the decomposition bias - variance and from Proposition 23, recalling that for $f(X_t) := \frac{1}{h} K(\frac{x - X_t}{h})$ we have $\|f\|_\infty \le c h^{-1}$ and its support $\mathcal{S}$ is such that $|\mathcal{S}| \le ch$.
Now, aiming at balancing the terms, we take $h := (\frac{1}{T})^a$; getting the mean-squared error is upper bounded by

$$(\frac{1}{T})^{2a\beta} + \frac{1}{T} + \frac{(a \log T)^{2 - \frac{(1+\alpha)}{2}}}{T} + \frac{a \log T}{T}.$$

If a gets bigger clearly h gets smaller; it is enough to take a such that $2a\beta > 1$ to obtain the first two terms here above are negligible compared to the last ones, which gives us the convergence rate $\frac{(\log T)^{2 - \frac{(1+\alpha)}{2}}}{T}$ for $\alpha \le 1$, $\frac{\log T}{T}$ for $\alpha > 1$. □

4.5.4 Proof of Proposition 26

Proof. Again, (4.8) follows naturally from the bias - variance decomposition and Proposition 23.
Regarding the convergence rate, we take again $h_l := (\frac{1}{T})^{a_l}$ for $l = 1, 2$.
It follows $\log(\frac{1}{h_1 h_2}) = a_1 \log T + a_2 \log T$ and so the mean-squared error is upper bounded by

$$(\frac{1}{T})^{2a_1\beta_1} + (\frac{1}{T})^{2a_2\beta_2} + \frac{c \log T}{T}.$$

Taking a_1 and a_2 big enough to make the first two terms here above negligible compared to the third, we get the convergence rate $\frac{\log T}{T}$. □

4.5.5 Proof of Corollary 2

Proof. It is a straightforward consequence of Propositions 24, 25 and 26 after having remarked that the constants which turn out in all the previous propositions do not depend on the point x considered. Indeed, such propositions yield

$$\mathbb{E}[\|\hat{\mu}_{h,T} - \mu\|_A^2] = \mathbb{E}[\int_A |\hat{\mu}_{h,T}(x) - \mu(x)|^2 dx] \leq c \int_A cV_d(T) dx \leq c|A|V_d(T).$$

□

4.6 Proof of the adaptive procedure

The heart of the proof of Theorem 15 consist of finding an upper bound for the expected value of $A(h)$, which is gathered in the following proposition.

Proposition 27. *Suppose that assumptions A1 - A3 hold. Then,* $\forall h \in \mathcal{H}_T$,

$$\mathbb{E}[A(h)] \leq c_1 B(h) + c_1 e^{-c_2 (\log T)^2}.$$

Proposition 27 will be proven after the proofs of Theorems 15 and 16.

4.6.1 Proof of Theorem 15.

Proof. From triangular inequality it follows $\forall h \in \mathcal{H}_T$

$$\|\hat{\mu}_{\tilde{h}} - \mu\|^2 \leq c(\left\|\hat{\mu}_{\tilde{h}} - \hat{\mu}_{h,\tilde{h}}\right\|^2 + \left\|\hat{\mu}_{h,\tilde{h}} - \hat{\mu}_h\right\|^2 + \|\hat{\mu}_h - \mu\|^2) \tag{4.26}$$

By the definition (4.14) of $A(h)$ it follows that the first and the second term of (4.26) are respectively upper bounded by $A(h) + V(\tilde{h})$ and $A(\tilde{h}) + V(h)$, having also used on the second term that $\hat{\mu}_{h,\tilde{h}} = \hat{\mu}_{\tilde{h},h}$. Then, since $\tilde{h}$ has been defined in (4.15) as the $h \in \mathcal{H}_T$ for which $A(h) + V(h)$ is minimal, we clearly have that $A(\tilde{h}) + V(\tilde{h}) \leq A(h) + V(h)$.
Hence, for any $h \in \mathcal{H}_T$, we get

$$\|\hat{\mu}_{\tilde{h}} - \mu\|^2 \leq c(A(h) + V(h) + \|\hat{\mu}_h - \mu\|^2). \tag{4.27}$$

We want an upper bound for the expected value of the left hand side of the equation (4.27) and so we need to evaluate $\mathbb{E}[\|\hat{\mu}_h - \mu\|^2] = \mathbb{E}[\int_A |\hat{\mu}_h(x) - \mu(x)|^2 dx]$.
From the standard bias variance decomposition, recalling that $\mu_h = \mathbb{K}_h * \mu$ is such that $\mathbb{E}[\hat{\mu}_h(x)] = \mu_h(x)$, we get

$$\mathbb{E}[\int_A |\hat{\mu}_h(x) - \mu(x)|^2 dx] = \int_A |\mu_h(x) - \mu(x)|^2 dx + \int_A \mathbb{E}[|\hat{\mu}_h(x) - \mu_h(x)|^2] dx.$$

Now, we can upper bound the first term of the right hand side here above by enlarging the integration domain getting

$$\int_A |\mu_h(x) - \mu(x)|^2 dx \leq \int_{\tilde{A}} |\mu_h(x) - \mu(x)|^2 dx = B(h). \tag{4.28}$$

Moreover, as consequence of Proposition 23 in the case $d \geq 3$ we obtain, as it was in Proposition 24,

$$\int_A \mathbb{E}[|\hat{\mu}_h(x) - \mu_h(x)|^2]dx = \tag{4.29}$$

$$= \int_A Var(\frac{1}{T\prod_{l=1}^d h_l}\int_0^T \prod_{m=1}^d K(\frac{x_m - X_u^m}{h_m})du)dx \leq |A|\frac{c}{T}(\prod_{l=1}^d h_l)^{\frac{2}{d}-1}.$$

Comparing the upper bound given in (4.29) with the definition of $V(h)$ and using also (4.27) and (4.28) we get, for each $h \in \mathcal{H}_T$,

$$\mathbb{E}[\|\hat{\mu}_h - \mu\|^2] \leq c(B(h) + V(h)) + \mathbb{E}[A(h)].$$

Now from Proposition 27 and the arbitrariness of the bandwidth h we are considering it follows

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|^2] \leq c_1 \inf_{h' \in \mathcal{H}_T}(B(h') + V(h')) + c_1 e^{-c_2(\log T)^2},$$

as we wanted. □

As a consequence of Theorem 15 we get, considering the rate optimal choice $h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}}$ provided by Proposition 24, the estimation gathered in Theorem 16. Its proof relies on the fact that, for how we have found it in Proposition 24, if the rate optimal bandwidth belongs to $\mathcal{H}_T$ then the $\inf_{h\in\mathcal{H}_T}(B(h) + V(h))$ is clearly realized by it.

4.6.2 Proof of Theorem 16

Proof. We observe that, for the rate optimal choice $h(T)$ of the bandwidth, the conditions gathered in the right hand side of (4.10), which are $\frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d h_l \leq (\frac{1}{\log T})^{\frac{3d}{d-2}}$, hold true. Indeed,

$$\prod_{l=1}^d h_l(T) = (\frac{1}{T})^{\frac{\bar{\beta}}{2\bar{\beta}+d-2}\sum_{l=1}^d \frac{1}{\beta_l}} = (\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}}.$$

the upper bound condition is therefore

$$(\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}} \leq (\frac{1}{\log T})^{\frac{3d}{d+2}},$$

which is true if and only if $\log T \leq T^{\frac{d-2}{3(2\bar{\beta}+d-2)}}$. Now we observe that $\frac{d-2}{3(2\bar{\beta}+d-2)} > 0$ for $\bar{\beta} > 1 - \frac{d}{2}$, which is always true since $d \geq 3$. In particular we can write $\frac{d-2}{3(2\bar{\beta}+d-2)} =: \gamma \in (0,1)$ and, given that eventually for T going to ∞ it is $\log T \leq T^\gamma$, we have $\prod_{l=1}^d h_l(T) \leq (\frac{1}{\log T})^{\frac{3d}{d+2}}$.
In the same way it is

$$(\frac{1}{T})^{\frac{d}{2\bar{\beta}+d-2}} \geq \frac{(\log T)^{2d}}{T^{\frac{d}{3}}}.$$

For the same reasoning as here above it is true if $(\frac{1}{3} - \frac{1}{2\bar{\beta}+d-2})\frac{1}{2} =: \gamma$ is positive.
Being $\bar{\beta} > 1$ and $d \geq 3$, it turns out $\gamma > 0$, as we wanted.
Up to consider $\tilde{h}_l(T) := \frac{1}{\lfloor T^{\frac{\bar{\beta}}{\beta_l(2\bar{\beta}+d-2)}} \rfloor}$ instead of $h_l(T)$, which is asymptotically equivalent and which leads to the same convergence rate, we have that the rate optimal

choice belongs to the set of candidate bandwidths $\mathcal{H}_T$ proposed in (4.12).
Having now $h(T) \in \mathcal{H}_T$, for how we have found the rate optimal choice in Proposition 24, the $\inf_{h \in \mathcal{H}_T}(B(h) + V(h))$ is clearly realized by it and so the bound stated in Theorem 15 is actually (see also Corollary 2)

$$\mathbb{E}[\|\hat{\mu}_{\tilde{h}} - \mu\|_A^2] \le c_1(\frac{1}{T})^{\frac{2\bar{\beta}}{2\bar{\beta}+d-2}} + c_1 e^{-c_2(\log T)^2},$$

as we wanted.

□

We have showed Theorem 15 using, as main tool, the bound on $\mathbb{E}[A(h)]$ stated in Proposition 27. Its proof, as we will see in the next section, relies on the use of Berbee's coupling method as in Viennet [94] and on a version of Talagrand inequality given in Klein and Rio [53].

4.6.3 Proof of Proposition 27

Proof. To find a bound for $\mathbb{E}[A(h)]$, for each $h \in \mathcal{H}_T$ we want to use Talagrand inequality, stated on random variables which are independent. Therefore, we start introducing some blocks which are mutually independent. We do it through the use of Berbee's coupling method as done in Viennet [94], Proposition 5.1 and its proof p. 484.
We assume that $T = 2p_T q_T$, with p_T integer and q_T a real to be chosen. We split the initial process $X = (X_t)_{t \in [0,T]}$ in $2p_T$ processes of a length q_T: for each $j \in \{1, ..., p_T\}$ we consider
$X^{j,1} := (X_t)_{t \in [2(j-1)q_T, (2j-1)q_T]}$ and $X^{j,2} := (X_t)_{t \in [(2j-1)q_T, 2jq_T]}$.
Then, there exist a process $(X_t^*)_{t \in [0,T]}$ satisfying the following properties:

1. For $j \in \{1, ..., p_T\}$ the processes $X^{j,1}$ and $X^{*\,j,1} := (X_t^*)_{t \in [2(j-1)q_T, (2j-1)q_T]}$ have the same distribution and so have the processes $X^{j,2}$ and $X^{*\,j,2} := (X_t^*)_{t \in [(2j-1)q_T, 2jq_T]}$.
2. For $j \in \{1, ..., p_T\}$, $\mathbb{P}(X^{j,1} \neq X^{*\,j,1}) \le \beta_X(q_T)$ and $\mathbb{P}(X^{j,2} \neq X^{*\,j,2}) \le \beta_X(q_T)$, where β_X is the β - mixing coefficient of the process X as in (4.4).
3. For each $k \in \{1, 2\}$, $X^{*\,1,k}, ..., X^{*\,p_T,k}$ are independent.

We denote by $\hat{\mu}_h^*$ the estimator computed using X_t^* instead of X_t and we write $\hat{\mu}_h^* = \frac{1}{2}(\hat{\mu}_h^{*(1)} + \hat{\mu}_h^{*(2)})$ to separate the part coming from $X^{*\,\cdot,1}$ (super -index (1)) and those coming from $X^{*\,\cdot,2}$ (super -index (2)), having $\hat{\mu}_h^{*(1)} := \frac{1}{p_T q_T}\sum_{j=1}^{p_T}\int_{2(j-1)q_T}^{(2j-1)q_T} \mathbb{K}_h(X_u^* - x)du$.
In a natural way we define moreover $\hat{\mu}_{h,\eta}^* := \mathbb{K}_\eta * \hat{\mu}_h^*$, which can be written again as $\frac{1}{2}(\hat{\mu}_{h,\eta}^{*(1)} + \hat{\mu}_{h,\eta}^{*(2)})$, to separate the contribution of $X^{*\,\cdot,1}$ and $X^{*\,\cdot,2}$.
With this background we can evaluate $\mathbb{E}[A(h)]$. We recall that, as defined in (4.14),

$$A(h) := \sup_{\eta \in \mathcal{H}_T}(\|\hat{\mu}_{h,\eta} - \hat{\mu}_\eta\|^2 - V(\eta))_+.$$

Now we can see $\hat{\mu}_{h,\eta} - \hat{\mu}_\eta$ as sum of different terms which we deal with singularly:

$$\hat{\mu}_{h,\eta} - \hat{\mu}_\eta := (\hat{\mu}_{h,\eta} - \hat{\mu}_{h,\eta}^*) + (\hat{\mu}_{h,\eta}^* - \mu_{h,\eta}) + (\mu_{h,\eta} - \mu_\eta) + (\mu_\eta - \hat{\mu}_\eta^*) + (\hat{\mu}_\eta^* - \hat{\mu}_\eta).$$

As a consequence of the triangular inequality and of the definition of $A(h)$ the following estimation holds true:

$$A(h) \leq \sup_{\eta \in \mathcal{H}_T} [\left\| \hat{\mu}_{h,\eta} - \hat{\mu}^*_{h,\eta} \right\|^2 + (\left\| \hat{\mu}^*_{h,\eta} - \mu_{h,\eta} \right\|^2 - \frac{V(\eta)}{2})_+ +$$

$$+ \left\| \mu_{h,\eta} - \mu_\eta \right\|^2 + (\left\| \mu_\eta - \hat{\mu}^*_\eta \right\|^2 - \frac{V(\eta)}{2})_+ + \left\| \hat{\mu}^*_\eta - \hat{\mu}_\eta \right\|^2] = \sup_{\eta \in \mathcal{H}_T} [\sum_{j=1}^5 I_j^{h,\eta}]$$

We start considering $I_5^{h,\eta}$. We define the set

$$\Omega^* := \{ X_t = X_t^* \quad \forall t \in [0,T] \} .$$

As a consequence of the second property of the process X^* and of the β - mixing with exponential decay showed in Lemma 36 we get, recalling that $2p_T q_t = T$ (with q_T and p_T to be chosen),

$$\mathbb{P}(\Omega^{*c}) \leq 2p_T \beta_X(q_T) \leq c \frac{T}{q_T} e^{-\gamma q_T}. \tag{4.30}$$

From the definition of $\hat{\mu}^*_h$ and Jensen inequality it is

$$\left\| \hat{\mu}^*_\eta - \hat{\mu}_\eta \right\|^2 = \int_A (\frac{1}{T} \int_0^T \mathbb{K}_\eta(X_t - x) - \mathbb{K}_\eta(X_t^* - x) dt)^2 dx 1_{\Omega^{*c}} \leq$$

$$\leq c \int_A \frac{1}{T^2} T (\int_0^T \mathbb{K}_\eta^2(X_t - x) + \mathbb{K}_\eta^2(X_t^* - x) dt) dx 1_{\Omega^{*c}} \leq c \left\| \mathbb{K}_\eta \right\|_\infty^2 1_{\Omega^{*c}} \tag{4.31}$$

By the definition (4.13) we get that, $\forall h \in \mathcal{H}_T$, $\left\| \mathbb{K}_h \right\|_\infty \leq (\prod_{l=1}^d h_l)^{-1}$.
We recall that, from how we have defined in (4.10) the set $\mathcal{H}_T$, we have that $\forall h \in \mathcal{H}_T$ $\prod_{l=1}^d h_l > \frac{(\log T)^{2d}}{T^{\frac{d}{3}}}$ and so

$$\left\| \mathbb{K}_\eta \right\|_\infty^2 < \frac{1}{(\prod_{l=1}^d \eta_l)^2} \leq \frac{T^{\frac{2d}{3}}}{(\log T)^{4d}}.$$

Replacing this bound in (4.31) it follows

$$\sup_{\eta \in \mathcal{H}_T} \left\| \hat{\mu}^*_\eta - \hat{\mu}_\eta \right\|^2 \leq c \frac{T^{\frac{2d}{3}}}{(\log T)^{4d}} 1_{\Omega^{*c}}.$$

We take its expectation and we use (4.30), getting a term which depends on q_T, a real to be chosen. From the arbitrariness of q_T we get a convergence to zero as fast as we want, for T going to ∞. Indeed, taking $q_T := (\log T)^2$ yields $\forall h, \eta \in \mathcal{H}_T$,

$$\mathbb{E}[|\sup_{\eta \in \mathcal{H}_T} I_5^{h,\eta}|] = \mathbb{E}[|\sup_{\eta \in \mathcal{H}_T} \left\| \hat{\mu}^*_\eta - \hat{\mu}_\eta \right\|^2 |] \leq c \frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma (\log T)^2}. \tag{4.32}$$

Regarding $\sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta}$, we estimate it through (4.32) and the following lemma, which will be proven in the appendix.

Lemma 37. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded, measurable function with support $\mathcal{S}$ satisfying*
$diam(\mathcal{S}) \le 2\sqrt{d}$; $\tilde{A}$ a compact set in $\mathbb{R}^d$ such that $A \subset \tilde{A}$ and $\tilde{A} = \left\{ \zeta : d(\zeta, A) \le 2\sqrt{d} \right\}$ and g a function in $L^2(\tilde{A})$. Then,

$$\|f * g\|_A \le \|f\|_{1,\mathbb{R}^d} \|g\|_{2,\tilde{A}},$$

where we have denoted as $\|.\|_A$ the usual L^2 norm on A, $\|.\|_{1,\mathbb{R}^d}$ the L^1 norm on $\mathbb{R}^d$ and $\|.\|_{2,\tilde{A}}$ the L^2 norm on $\tilde{A}$.

We recall that $\hat{\mu}_{h,\eta} = \mathbb{K}_\eta * \hat{\mu}_h$ and $\hat{\mu}^*_{h,\eta} = \mathbb{K}_\eta * \hat{\mu}^*_h$. Therefore, remarking that $diam(K) \le 2$ and so by the definition of $\mathbb{K}_\eta$ it is $diam(\mathbb{K}_\eta) \le 2\sqrt{d}$, we can use Lemma 37, which yields

$$\sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta} = \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta * (\hat{\mu}_h - \hat{\mu}^*_h)\|^2 \le \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta\|^2_{1,\mathbb{R}^d} \|\hat{\mu}_h - \hat{\mu}^*_h\|^2_{\tilde{A}}.$$

Taking the expected value, using that $\|\mathbb{K}_\eta\|_{1,\mathbb{R}^d} \le c \ \forall \eta \in \mathcal{H}_T$ and the equation (4.32), remarking that the dependence on the integration set considered is hidden in the constant c in which this time will appear $|\tilde{A}|$ instead of $|A|$ we get

$$\mathbb{E}[| \sup_{\eta \in \mathcal{H}_T} I_1^{h,\eta}|] \le c \frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma(\log T)^2}. \tag{4.33}$$

We still use Lemma 37 to study $\sup_{\eta \in \mathcal{H}_T} I_3^{h,\eta}$, recalling that $\mu_{h,\eta} = \mathbb{K}_\eta * \mu_h$ and $\mu_\eta = \mathbb{K}_\eta * \mu_b$. It yields

$$\sup_{\eta \in \mathcal{H}_T} I_3^{h,\eta} = \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta * (\mu_h - \mu)\|^2 \le \sup_{\eta \in \mathcal{H}_T} \|\mathbb{K}_\eta\|^2_{1,\mathbb{R}^d} \|\mu_h - \mu\|^2_{\tilde{A}} \le c \|\mu_h - \mu\|^2_{\tilde{A}} = cB(h). \tag{4.34}$$

We are left to study $I_2^{h,\eta}$ and $I_4^{h,\eta}$, for which we need the following lemma that will be showed right after the proof of this proposition.

Lemma 38. *For $i = 1, 2$, there exist some positive constants c_1^*, c_2^*, c_3^* and a constant k_0^* such that, for any $\bar{k} \ge k_0^*$,*

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\left\|\mu_\eta - \hat{\mu}_\eta^{*(i)}\right\|^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \le c_1^* T^{c_2^*} e^{-c_3^*(\log T)^2}. \tag{4.35}$$

Moreover there exist k_0 such that, for any $\bar{k} \ge k_0$,

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\left\|\mu_{h,\eta} - \hat{\mu}_{h,\eta}^{*(i)}\right\|^2 - \frac{\bar{k}}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \le c_1^* T^{c_2^*} e^{-c_3^*(\log T)^2}. \tag{4.36}$$

Concerning $I_4^{h,\eta}$ observe it is $\mu_\eta - \hat{\mu}^*_\eta = \frac{1}{2}(2\mu_\eta - \hat{\mu}^{*(1)}_\eta - \hat{\mu}^{*(2)}_\eta)$. Hence, from triangular inequality and the definition of positive part function, we get

$$I_4^{h,\eta} \le c(\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|^2 - \frac{V(\eta)}{2})_+ + c(\left\|\mu_\eta - \hat{\mu}_\eta^{*(2)}\right\|^2 - \frac{V(\eta)}{2})_+.$$

From (4.35), for a k in the definition of $V(\eta)$ big enough, for which we have $\frac{k}{2} > (k_0^* \vee k_0)$, we get

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_4^{h,\eta}] \leq c_1 T^{c_2} e^{-c_3 (\log T)^2}. \tag{4.37}$$

In the same way, remarking that

$$I_2^{h,\eta} \leq c(\left\|\hat{\mu}_{h,\eta}^{*(1)} - \mu_{h,\eta}\right\|^2 - \frac{V(\eta)}{2})_+ + c(\left\|\hat{\mu}_{h,\eta}^{*(2)} - \mu_{h,\eta}\right\|^2 - \frac{V(\eta)}{2})_+$$

and using (4.36) it follows

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} I_2^{h,\eta}] \leq c_1 T^{c_2} e^{-c_3 (\log T)^2}. \tag{4.38}$$

From (4.32) , (4.33), (4.34), (4.37) and (4.38) we obtain, for any $h \in \mathcal{H}_T$,

$$\mathbb{E}[A(h)] \leq c\frac{T^{\frac{2d}{3}+1}}{(\log T)^{4d+2}} e^{-\gamma (\log T)^2} + cB(h) + c_1 T^{c_2} e^{-c_3 (\log T)^2} \leq cB(h) + c_1 e^{-c_2 (\log T)^2},$$

as we wanted. □

To conclude the proof of the adaptive procedure we need to show Lemma 38, which core is the use of the Talagrand inequality. First of all, we recall the following version of the Talagrand inequality, which has been stated as Lemma 2 in [23] and which is a straightforward consequence of the Talagarand inequality given in Klein and Rio [53].

Lemma 39. *Let $T_1, ..., T_n$ be independent random variables with values in some Polish space $\mathcal{X}$, $\mathcal{R}$ a countable class of measurable functions from $\mathcal{X}$ into $[-1,1]^p$ and $v_p(r) := \frac{1}{p}\sum_{j=1}^p [r(T_j) - \mathbb{E}[r(T_j)]]$. Then,*

$$\mathbb{E}[(\sup_{r \in \mathcal{R}} |v_p(r)|^2 - 2H^2)_+] \leq c(\frac{v}{p} e^{-c\frac{pH^2}{v}} + \frac{M^2}{p^2} e^{-c\frac{pH}{M}}), \tag{4.39}$$

with c a universal constant and where

$$\sup_{r \in \mathcal{R}} \|r\|_\infty \leq M, \quad \mathbb{E}_b[\sup_{r \in \mathcal{R}} |v_p(r)|] \leq H, \quad \sup_{r \in \mathcal{R}} \frac{1}{p} \sum_{j=1}^p Var(r(T_j)) \leq v.$$

4.6.4 Proof of Lemma 38

Proof. Since the two cases $i = 1$ and $i = 2$ are similar, we study only one of them. We start proving (4.35), the proof of inequality (4.36) follows the same line. We first observe it is

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_A^2 - \frac{\bar{k}}{T}(\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq \sum_{\eta \in \mathcal{H}_T} \mathbb{E}[(\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_A^2 - \frac{\bar{k}}{T}(\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+]. \tag{4.40}$$

Our goal is now to find a bound for the right hand side of the inequality here above using the version of the Talagrand inequality gathered in Lemma 39. To do it, we need to introduce some notation.

We observe that $\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_A^2 = \sup_{r, \|r\|=1} < \mu_\eta - \hat{\mu}_\eta^{*(1)}, r >^2$, and the supremum can be considered over a countable dense set of function r such that $\|r\| = 1$; let us denote this set by $\mathcal{B}(1)$.
We define

$$T_j(z) := \frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z)dt; \quad r(T_j) := \int_A T_j(z) r(z) dz.$$

Thus

$$v_{p_T}(r) =< \hat{\mu}_\eta^{*(1)} - \mu_\eta, r >= \frac{1}{p_T} \sum_{j=1}^{p_T} [r(T_j) - \mathbb{E}[r(T_j)]]$$

is a centered empirical process with independent variables

$$\psi_r(X^{*j,1}) := r(T_j) - \mathbb{E}[r(T_j)] =$$

$$= \int_A (\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} [\mathbb{K}_\eta(X_t^{*j,1} - z) - \mathbb{E}[\mathbb{K}_\eta(X_t^{*j,1} - z)]]dt) r(z) dz,$$

to which we want to apply Talagrand inequality (4.39). Therefore, we have to compute M, H and v as defined in Lemma 39. We start by the calculation of M. For any $r \in \mathcal{B}(1)$ it is, using the definition of r and Cauchy - Schwartz inequality,

$$|\int_A T_j(z) r(z) dz| \le (\int_A T_j^2(z) dz)^{\frac{1}{2}}.$$

Now from the definition of T and Jensen inequality it follows

$$\int_A T_j^2(z) dz \le \int_A \frac{1}{q_T^2} q_T \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta^2(X_t^{*j,1} - z) dt dz \le$$

$$\le c \|K_\eta\|_\infty^2 |\mathcal{S}| \le c(\prod_{l=1}^d \eta_l)^{-2} (\prod_{l=1}^d \eta_l) = c(\prod_{l=1}^d \eta_l)^{-1},$$

where we have also used that the support of K_η is on $\mathcal{S}$ which size is $\prod_{l=1}^d \eta_l$. Hence,

$$|\int_A T_j(z) r(z) dz| \le c(\prod_{l=1}^d \eta_l)^{-\frac{1}{2}} =: M. \tag{4.41}$$

Regarding the computation of H, from the definition of $v_{p_T}(r)$ and the fact that the random variables $\psi_r^{*j,1}$ are centered and independents it follows

$$\mathbb{E}[\sup_{r \in \mathcal{B}(1)} |v_{p_T}(r)|^2] = \mathbb{E}_b[\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_A^2] =$$

$$= \int_A Var(\frac{1}{p_T} \sum_{j=1}^{p_T} \frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} (\mathbb{K}_\eta(X_t^{*j,1} - z) - \mathbb{E}[\mathbb{K}_\eta(X_t^{*j,1} - z)]) dt) dz =$$

$$= \int_A \frac{1}{p_T} Var(\frac{1}{q_T} \int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z) dt) dz \le c|A| \frac{1}{p_T} \frac{1}{q_T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d} - 1},$$

where in the last inequality we have used the estimation for the variance in the case $d \ge 3$ gathered in Proposition 23, considering that taking the Kernel function as f we have that its support $\mathcal{S}$ is such that $|\mathcal{S}| \le c(\prod_{l=1}^d \eta_l)$. It yields

$$\mathbb{E}[\sup_{r \in \mathcal{B}(1)} |v_{p_T}(r)|^2] \le \frac{c}{T} (\prod_{l=1}^d \eta_l)^{\frac{2}{d} - 1} =: H^2 \tag{4.42}$$

In order to use Lemma 39, we are left to compute v.
We observe it is

$$\frac{1}{p_T}\sum_{j=1}^{p_T} Var(\int_A \frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1}-z)dt\, r(z)dz) =$$

$$= \frac{1}{p_T}\sum_{j=1}^{p_T} Var(\frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} (\mathbb{K}_\eta * r)(X_t^{*j,1})dt).$$

We want to prove a tight upper bound for the variance of the integral functional $\frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1})dt$ of the diffusion X^* , where we have denoted $f_\eta := \mathbb{K}_\eta * r$. Following the proof we have given of Proposition 23 we have,

$$Var(\frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1})dt) \le \frac{c}{q_T^2}\int_0^{q_T}(q_T-u)\mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})]du =$$

$$= \frac{c}{q_T^2}(\int_0^D (q_T-u)\mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})]du + \int_D^{q_T}(q_T-u)\mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})]du),$$

where we have introduced $f_{\eta,c}(x) := f_\eta(x) - \mu(f_\eta)$ and D a quantity to be chosen in order to balance the contribution of the two integrals here above. We denote moreover as P_t^* the transition semigroup of the process X^* Concerning the integral between 0 and D, we act like we did in (4.17) and we use Cauchy - Schwartz inequality and the fact that P_t^* is a contraction. It follows

$$\frac{c}{q_T^2}\int_0^D (q_T-u)\mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})]du \le \frac{c}{q_T}|\int_0^D < P_t^* f_\eta, f_\eta >_\mu dt| \le$$

$$\le \frac{c}{q_T}\int_0^D (\|P_t^* f_\eta\|_\mu^2 \|f_\eta\|_\mu^2)^{\frac{1}{2}} dt \le \frac{c}{q_T}\int_0^D \|f_\eta\|_\mu^2 dt \le \frac{cD}{q_T}, \tag{4.43}$$

where in the last inequality we have used the fact that $\|\mu\|_\infty \le c$, Young inequality and the definition of the Kernel function and of r in order to say

$$\|f_\eta\|_\mu^2 = \|\mathbb{K}_\eta * r\|_\mu^2 \le c\,\|\mathbb{K}_\eta * r\|_{2,\mathbb{R}^d}^2 \le c\,\|\mathbb{K}_\eta\|_{1,\mathbb{R}^d}^2 \|r\|_{2,\mathbb{R}^d}^2 \le c. \tag{4.44}$$

Regarding the integral between D and q_T, we act like we did in (4.21) using the exponential ergodicity gathered in Lemma 36 to get

$$\frac{c}{q_T^2}\int_D^{q_T}(q_T-u)\mathbb{E}[f_{\eta,c}(X_0^{*j,1}), f_{\eta,c}(X_u^{*j,1})]du \le \tag{4.45}$$

$$\le \frac{c}{q_T}|\int_D^{q_T} < P_t^* f_{\eta,c}, f_{\eta,c} >_\mu dt| \le \frac{c}{q_T}\int_D^{q_T} e^{-\rho t}\|f_{\eta,c}\|_\infty^2 dt.$$

We now recall that $f_{\eta,c}(x) = f_\eta(x) - \mu(f_\eta)$ with $\mu(f_\eta) = \int_{\mathbb{R}^d} f_\eta(x)\mu(x)dx$. Hence, from Cauchy- Schwartz inequality and (4.44), we get $|\mu(f_\eta)| \le c$ and therefore

$$|f_{\eta,c}(x)| \le |f_\eta(x)| + c. \tag{4.46}$$

To estimate the infinity norm of $f_{\eta,c}$ we remark it is, $\forall y \in \mathbb{R}^d$,

$$|f_\eta(y)| = |\mathbb{K}_\eta * r(y)| = |\int_A \mathbb{K}_\eta(y-z)r(z)dz| \le$$

$$\le (\int_A \mathbb{K}_\eta^2(y-z)dz)^{\frac{1}{2}}(\int_A r^2(z)dz)^{\frac{1}{2}} = (\int_{\mathcal{S}} \mathbb{K}_\eta^2(y-z)dz)^{\frac{1}{2}},$$

where we have used Cauchy - Schwartz inequality and the fact that $r \in \mathcal{B}(1)$ and so its 2 - norm is equal to 1 by definition. Moreover, the Kernel function is different from 0 only on its support $\mathcal{S}$, which size is $\prod_{l=1}^d \eta_l$. Hence, recalling also that the infinite norm of $\mathbb{K}_\eta$ is upper bounded by $c(\prod_{l=1}^d \eta_l)^{-1}$, we get

$$|\mathbb{K}_\eta * r(y)| \le c(\|\mathbb{K}_\eta\|_\infty^2 |\mathcal{S}|)^{\frac{1}{2}} \le \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}}.$$

It follows

$$\|f_{\eta,c}\|_\infty \le \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}} + c \le \frac{c}{\sqrt{\prod_{l=1}^d \eta_l}}, \tag{4.47}$$

given that the second term is negligible compared to the first. Replacing (4.47) in (4.45), using also (4.43), we obtain

$$Var(\frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} f_\eta(X_t^{*j,1})dt) \le \frac{c}{q_T}(D + \frac{e^{-\rho D}}{\prod_{l=1}^d \eta_l}).$$

We look for a D for which the first and the second term of the right hand side of the inequality here above have the same magnitude. Therefore, we choose $D := [\max(-\frac{1}{\rho}\log(\prod_{l=1}^d \eta_l), 1)] \wedge q_T$. Replacing such a value we get, if $q_T > -\frac{1}{\rho}\log(\prod_{l=1}^d \eta_l)$,

$$\frac{1}{p_T}\sum_{j=1}^{p_T} Var(\int_A \frac{1}{q_T}\int_{(2j-1)q_T}^{2jq_T} \mathbb{K}_\eta(X_t^{*j,1} - z)dt\, r(z)dz) \le \frac{c}{q_T}(1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|}))$$

Otherwise, if $q_T \le -\frac{1}{\rho}\log(\prod_{l=1}^d \eta_l)$, by the definition of D we have $D = q_T$. We still have the contribution of $\frac{c}{q_T}D$ which is in this case less than $\frac{c}{q_T}(\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))$ and, moreover, the contribution of the integral between D and q_T is now null since we have $D = q_T$. Hence we have

$$v := \frac{c}{q_T}(1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|})). \tag{4.48}$$

We use Lemma 39 on the right hand side of (4.40), recalling that $\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_A^2 = \sup_{r\in\mathcal{B}(1)} < \mu_\eta - \hat{\mu}_\eta^{*(1)}, r >^2 = \sup_{r\in\mathcal{B}(1)} |v_{p_T}(r)|^2$; with M, H and v as found in (4.41), (4.42) and (4.48). It follows

$$\mathbb{E}[\sup_{\eta\in\mathcal{H}_T}(\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|^2 - \frac{\bar{k}}{T}(\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \le$$

$$\le c\sum_{\eta\in\mathcal{H}_T} \frac{(1+\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}{p_T q_T} e^{-c\frac{\frac{p_T}{T}(\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1}}{\frac{1}{q_T}(1+\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}} + \frac{(\prod_{l=1}^d \eta_l)^{-1}}{p_T^2} e^{-c\frac{p_T\frac{1}{\sqrt{T}}(\prod_{l=1}^d \eta_l)^{\frac{1}{d}-\frac{1}{2}}}{(\prod_{l=1}^d \eta_l)^{-\frac{1}{2}}}}.$$

We recall that $2p_T q_T = T$, where q_T is chosen above equation (4.32) as $(\log T)^2$; we can therefore upper bound the right hand side of the equation here above with

$$c\sum_{\eta\in\mathcal{H}_T} \frac{(1+\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}{T} e^{-\frac{c}{(\prod_{l=1}^d \eta_l)^{1-\frac{2}{d}}(1+\log(\frac{1}{|\prod_{l=1}^d \eta_l|}))}} + \frac{(\log T)^4}{(\prod_{l=1}^d \eta_l)T^2} e^{-c\frac{\sqrt{T}}{(\log T)^2}(\prod_{l=1}^d \eta_l)^{\frac{1}{d}}} \le$$

$$\leq (\frac{\log T}{T} e^{-c(\log T)^2} + \frac{T^{\frac{d}{3}-2}}{(\log T)^{2d-2}} e^{-cT^{\frac{1}{6}}})|\mathcal{H}_T|,$$

where in the last inequality we have used that, by the definition (4.10) we have given of $\mathcal{H}_T$, $\forall h \in \mathcal{H}_T$ we have $\frac{(\log T)^{2d}}{T^{\frac{d}{3}}} \leq \prod_{l=1}^d h_l \leq (\frac{1}{\log T})^{\frac{3d}{d-2}}$ and so $(\prod_{l=1}^d \eta_l)^{1-\frac{2}{d}}(1 + \log(\frac{1}{|\prod_{l=1}^d \eta_l|})) \leq c\frac{1}{(\log T)^2}$ and $\frac{\sqrt{T}}{(\log T)^2}(\prod_{l=1}^d \eta_l)^{\frac{1}{d}} \geq cT^{\frac{1}{6}}$; we have therefore upper bounded each element of the sum with a quantity which does not depend on η.
We have moreover assumed that $|\mathcal{H}_T|$ has polynomial growth in T; and so there is a constant $c > 0$ such that

$$\mathbb{E}[\sup_{\eta \in \mathcal{H}_T} (\left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|^2 - \frac{\bar{k}}{T}(\prod_{l=1}^d \eta_l)^{\frac{2}{d}-1})_+] \leq (\frac{\log T}{T} e^{-c(\log T)^2} + \frac{T^{\frac{d}{3}-2}}{(\log T)^{2d-2}} e^{-cT^{\frac{1}{6}}})T^c;$$

inequality (4.35) follows.
As a consequence of Lemma 37 and the definition of Kernel function we moreover have

$$\left\|\mu_{h,\eta} - \hat{\mu}_{h,\eta}^{*(1)}\right\|^2 = \left\|\mathbb{K}_h * (\mu_\eta - \hat{\mu}_\eta^{*(1)})\right\|^2 \leq \|\mathbb{K}_h\|_{1,\mathbb{R}^d}^2 \left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_{2,\tilde{A}}^2 \leq c \left\|\mu_\eta - \hat{\mu}_\eta^{*(1)}\right\|_{2,\tilde{A}}^2.$$

Using that and (4.35) we have just shown we obtain (4.36).

□

4.7 Appendix

4.7.1 Proof of Lemma 35

Proof. Lemma 35 relies on the first point of Theorem 1.1 in [18], which needs the following assumptions on the coefficients a, b, γ and on the jumps.
(H^a) There are $c_1 > 0$ and $\beta \in (0,1)$ such that for all $x, y \in \mathbb{R}^d$, $|a(x) - a(y)| \leq c_1|x-y|^\beta$ and, for some $c_2 \geq 1$, $c_2^{-1}\mathbb{I}_{d\times d} \leq a(x) \leq c_2\mathbb{I}_{d\times d}$.
(H^k) The function $k(x,z) := |z|^{d+\alpha}F(\frac{z}{\gamma(x)})$ is bounded, measurable and, if $\alpha = 1$, for any $0 < r < R < \infty$ it is

$$\int_{r\leq z\leq R} zk(x,z)|z|^{-d-1}dz = 0. \tag{4.49}$$

(H^b) The function b belongs to the Kato class $\mathbb{K}^2$ which is, as defined in [18],

$$\mathbb{K}^\gamma := \left\{ f : \mathbb{R}^d \to \mathbb{R} \text{ satisfies} \lim_{\delta\to 0} \sup_x \int_0^\delta \int_{\mathbb{R}^d} |f(x \pm y)|\eta_{\gamma,\gamma-1}(s,y)dy\, ds = 0 \right\},$$

where we have denoted $f(x \pm y)$ as an abbreviation for $f(x+y) + f(x-y)$ and

$$\eta_{\alpha,\gamma}(t,x) := t^{\frac{\gamma}{2}}(|x| + t^{\frac{1}{2}})^{-d-\alpha}.$$

Through this paper we have assumed that Assumptions A1 - A3 hold.
From A1 it follows H^a since we have asked the boundedness of a and, in the case in which x and y are such that $|x-y| > 1$ we have that $\exists c$ such that $|a(x) - a(y)| \leq |a(x)| + |a(y)| \leq 2c \leq 2c|x-y|^\beta$, for each $\beta \in (0,1)$. When $|x-y| \leq 1$, instead, we have $|a(x) - a(y)| \leq L|x-y| = L|x-y|^{1-\beta}|x-y|^\beta \leq L|x-y|^\beta$, as consequence of

the Lipschitz continuity.
Regarding H^k, we have that k is a bounded function as a consequence of second and third points of A3. Indeed,

$$|k(x,z)| = |z|^{d+\alpha}|F(\frac{z}{\gamma(x)})| \le |z|^{d+\alpha}\frac{|\gamma(x)|}{|z|^{d+\alpha}} \le \gamma_{max} < \infty.$$

Observing moreover that (4.49) holds true on the basis of the fourth point of A3, H^k clearly follows.
We are left to show that $b \in \mathbb{K}^2$. Noticing that (cf also Remark 2.6 in [18])

$$\int_0^\delta \eta_{\alpha,\alpha-1}(s,y)ds = \int_0^\delta \frac{s^{\frac{(\alpha-1)}{2}}}{(|y|+s^{\frac{1}{2}})^{d+\alpha}}ds \le c\frac{(|y|^2\wedge\delta)^{\frac{(1+\alpha)}{2}}}{|y|^{d+\alpha}} = \frac{1}{|y|^{d-1}}(1\wedge\frac{\delta}{|y|^2})^{\frac{(1+\alpha)}{2}},$$

it is enough to show that

$$M_b^2(\delta) := \sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d}|b(x+y)|\frac{1}{|y|^{d-1}}(1\wedge\frac{\delta}{|y|^2})^{\frac{3}{2}}dy \longrightarrow 0 \quad \text{as } \delta\to 0$$

to get $b \in \mathbb{K}^2$.

From Assumption A1 we now b is upper bounded by a constant. Therefore we have

$$\sup_{x\in\mathbb{R}^d}|\int_{\mathbb{R}^d}|b(x+y)|\frac{1}{|y|^{d-1}}(1\wedge\frac{\delta}{|y|^2})^{\frac{3}{2}}dy| \le \tag{4.50}$$

$$\le c\int_{\{y:|y|\le\sqrt{\delta}\}}\frac{1}{|y|^{d-1}}dy + c\int_{\{y:|y|>\sqrt{\delta}\}}\frac{1}{|y|^{d-1}}(\frac{\delta}{|y|^2})^{\frac{3}{2}}dy.$$

We move to polar coordinates system, getting the right hand side of the equation (4.50) here above is upper bounded by

$$\int_0^{\sqrt{\delta}} c d\rho + c\delta^{\frac{3}{2}}\int_{\sqrt{\delta}}^\infty \rho^{-3}d\rho = c\sqrt{\delta} + c\delta^{\frac{3}{2}}\frac{1}{2\delta},$$

which clearly goes to 0 for $\delta \to 0$. It yields that H^b holds true.
It entails we can use Theorem 1.1 of [18]; Lemma 35 follows.

□

4.7.2 Proof of Lemma 36

Proof. The exponential ergodicity and the exponential β - mixing of the process X are showed in Proposition 3.8 and the second point of Theorem 2.2 of [67].
To use them we have to show that Assumptions 1, 2 and 3* stated in [67] hold.
Assumption 1 of [67] is a regularity condition which corresponds to our assumption A1.
We want to show that point b of Assumption 2 of [67] holds, which is the following:
(b) There exists a constant $\Delta > 0$ such that X_Δ admits a density $p_\Delta(x,y)$ with respect to the Lebesgue measure on $\mathbb{R}^d$ for every $x \in \mathbb{R}^d$, and $(x,y) \mapsto p_\Delta(x,y)$ is bounded in $y \in \mathbb{R}^d$ and in $x \in K$ for every compact $K \subset \mathbb{R}^d$. Moreover, for every $x \in \mathbb{R}^d$ and every open ball $U \subset \mathbb{R}^d$ there exists a point $z = z(x,U) \in supp(F)$ such that $\gamma(x)\cdot z \in U$.

We observe that the existence of a bounded density has already been proven in Lemma 35. Moreover, from second and third points of A3, we know that $supp(F) = \mathbb{R}^d$ and that γ is an invertible matrix. Hence, for every $x \in \mathbb{R}^d$ and every open ball $U \subset \mathbb{R}^d$ there exists a point $z = z(x, U) \in \mathbb{R}^d$ such that $\gamma(x) \cdot z \in U$.
To conclude, we have to prove that Assumption 3* holds and so we have to show the existence of a Lyapunov function. We therefore want to provide a function f^* which satisfies the drift condition $Af^* \leq -c_1 f^* + c_2$ for $c_1 > 0$ and $c_2 > 0$. A denotes the generator of the diffusion, which is the sum of the continuous and discrete part

$$A_c f(x) := \frac{1}{2}\sum_{i,j=1}^{d} a_{i,j}^2(x)\partial_{i,j}^2 f(x) + \sum_{i=1}^{d} b_i(x)\partial_i f(x) \qquad \text{and}$$

$$A_d f(x) := \int_{\mathbb{R}^d} (f(x + \gamma(x)\cdot z) - f(x) - \gamma(x)\cdot z \cdot \nabla f(x))F(z)dz,$$

for every function $f : \mathbb{R}^d \to \mathbb{R}$, $f \in \mathcal{C}^2(\mathbb{R}^d)$.
From the fifth point of condition A3 we know there exists $\epsilon > 0$ such that $\int_{\mathbb{R}^d} |z|^2 e^{\epsilon|z|}F(z)dz \leq c$. For such an ϵ we define $f^*(x) := e^{\epsilon|x|}$. We observe it is $\partial_i f^*(x) = \epsilon e^{\epsilon|x|}\frac{x_i}{|x|}$ and

$$\partial_{i,j}^2 f^*(x) = \epsilon\frac{x_i x_j}{|x|^2}e^{\epsilon|x|}(\epsilon - \frac{1}{|x|}) + \epsilon e^{\epsilon|x|}\frac{1}{|x|}1_{j=i}. \tag{4.51}$$

We therefore have, using also the drift condition gathered in Assumption A2, $\forall x : |x| > \tilde{\rho}$

$$|A_c f^*(x)| \leq \frac{1}{2}\epsilon e^{\epsilon|x|}(\epsilon + \frac{2}{|x|})\sum_{i,j=1}^{d}|a_{i,j}^2(x)| + \epsilon e^{\epsilon|x|}\frac{1}{|x|} < x, b(x) > \leq$$

$$\leq c\epsilon e^{\epsilon|x|}(\epsilon + \frac{1}{|x|})\sum_{i,j=1}^{d}|a_{i,j}^2(x)| - \tilde{C}\epsilon e^{\epsilon|x|}. \tag{4.52}$$

Concerning the discrete part of the generator, from intermediate value theorem we have

$$|A_d f^*(x)| \leq \int_{\mathbb{R}^d}\int_0^1 (\gamma(x)\cdot z)^T \cdot H^2 f_{(x+s\gamma(x)\cdot z)} \cdot (\gamma(x)\cdot z)dsdz,$$

where $H^2 f_{(x+s\gamma(x)\cdot z)}$ denotes the hessian matrix of the function f computed in the point $x + s\gamma(x)\cdot z$.
We now split the integral in the right hand side here above, to act differently depending on whether $|z|$ is more or less than $\frac{|x|}{2\|\gamma\|_\infty}$. We therefore get

$$|A_d f^*(x)| \leq \int_{z:|z|\leq\frac{|x|}{2\|\gamma\|_\infty}}\int_0^1 (\gamma(x)\cdot z)^T \cdot H^2 f_{(x+s\gamma(x)\cdot z)} \cdot (\gamma(x)\cdot z)dsdz +$$

$$+ \int_{z:|z|>\frac{|x|}{2\|\gamma\|_\infty}}\int_0^1 (\gamma(x)\cdot z)^T \cdot H^2 f_{(x+s\gamma(x)\cdot z)} \cdot (\gamma(x)\cdot z)dsdz =: I_1 + I_2.$$

Concerning I_1, from (4.51) it follows

$$I_1 \leq c\int_{z:|z|\leq\frac{|x|}{2\|\gamma\|_\infty}}\int_0^1 |z|^2\,\|\gamma\|_\infty^2\,\epsilon e^{\epsilon|x+sz\gamma(x)|}(\epsilon + \frac{1}{|x+sz\gamma(x)|})F(z)dzds \leq$$

$$\leq c\epsilon e^{\epsilon|x|}\int_{z:|z|\leq\frac{|x|}{2\|\gamma\|_\infty}}|z|^2e^{\epsilon|z|\|\gamma\|_\infty}(\epsilon+\frac{1}{|x|-|z|\,\|\gamma\|_\infty})F(z)dz\leq c\epsilon e^{\epsilon|x|}(\epsilon+\frac{1}{|x|}),\quad(4.53)$$

where in the last inequality we have used the fifth point of A3 and the fact that on the integral we are considering it is $|z|\leq\frac{|x|}{2\|\gamma\|_\infty}$ and so $\frac{1}{|x|-|z|\|\gamma\|_\infty}\leq\frac{2}{|x|}$.
We now study the term I_2, which is $I_{2,1}+I_{2,2}:=$

$$\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty},|x+sz\gamma(x)|\leq1\right\}}\int_0^1(\gamma(x)\cdot z)^T\cdot H^2f_{(x+s\gamma(x)\cdot z)}\cdot(\gamma(x)\cdot z)dsdz+$$

$$+\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty},|x+sz\gamma(x)|>1\right\}}\int_0^1(\gamma(x)\cdot z)^T\cdot H^2f_{(x+s\gamma(x)\cdot z)}\cdot(\gamma(x)\cdot z)dsdz.$$

On $I_{2,1}$ we can upper bound the hessian matrix with $c\epsilon$ and so we get

$$I_{2,1}\leq c\epsilon\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty},|x+sz\gamma(x)|\leq1\right\}}|z|^2\,\|\gamma\|_\infty^2\,F(z)dz\leq\quad(4.54)$$

$$\leq c\epsilon\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty}\right\}}|z|^2e^{\epsilon|z|}e^{-\epsilon|z|}F(z)dz\leq c\epsilon e^{-\epsilon\frac{|x|}{2\|\gamma\|_\infty}},$$

where in the last inequality we have used that $|z|>\frac{|x|}{2\|\gamma\|_\infty}$; after that we have enlarged the domain of integration and used the fifth point of Assumption A3 to upper bound the integral with a constant.
On $I_{2,2}$ we still use (4.51), getting

$$I_{2,2}\leq c\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty},|x+sz\gamma(x)|>1\right\}}\int_0^1|z|^2\,\|\gamma\|_\infty^2\,\epsilon e^{\epsilon|x+sz\gamma(x)|}(\epsilon+\frac{1}{|x+sz\gamma(x)|})F(z)dzds\leq$$

$$\leq c(\epsilon+1)\epsilon e^{\epsilon|x|}\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty},|x+sz\gamma(x)|>1\right\}}|z|^2e^{\epsilon|z|\|\gamma\|_\infty}F(z)dz.\quad(4.55)$$

We define $J(x):=\int_{\left\{z:|z|>\frac{|x|}{2\|\gamma\|_\infty}\right\}}|z|^2e^{\epsilon|z|\|\gamma\|_\infty}F(z)dz$. From (4.53) (4.54) and (4.55), observing that the domain of integration of the integral defined as J contains the one in (4.55) and using the boundedness of γ, we get

$$|A_df^*(x)|\leq c\,\epsilon e^{\epsilon|x|}(\epsilon+\frac{1}{|x|}+e^{-c\epsilon|x|}+(\epsilon+1)J(x)).\quad(4.56)$$

From (4.52) and (4.56) we get, using also the boundedness of a and J which follows from A1 the fifth point of A3,

$$|Af^*(x)|\leq\epsilon e^{\epsilon|x|}(c\epsilon-\tilde{C})+c\epsilon e^{\epsilon|x|}(\frac{1}{|x|}+e^{-c\epsilon|x|}+J(x)).\quad(4.57)$$

Since $\epsilon>0$ can be chosen small; $c\epsilon$ is therefore less then $\tilde{C}$ and so the first term here above turns out being negative.
Moreover, the second term on the right hand side of (4.57) is $o(f^*)$. Indeed, $c\epsilon(\frac{1}{|x|}+e^{-c\epsilon|x|})$ clearly goes to 0 for $|x|\to\infty$ and the domain of integration of the integral defined as $J(x)$ is the set of z such that $|z|>\frac{|x|}{2\|\gamma\|_\infty}$. Therefore, for $|x|\to\infty$ the contribution of the integral becomes null.
It follows $Af^*\leq-c_1f^*+o(f^*)$, as we wanted.
We get the drift condition holding true on the proposed function f^* and so the Assumption 3* does. □

4.7.3 Proof of Lemma 37

Proof. We observe that, from the definition of convolution and Cauchy-Schwartz inequality it is

$$|f * g(x)| = \int_{\mathbb{R}^d} |f(x-y)||g(y)|dy = \int_{\mathbb{R}^d} |f(x-y)|^{\frac{1}{2}}|g(y)||f(x-y)|^{\frac{1}{2}}dy \leq$$

$$\leq c(\int_{\mathbb{R}^d} |f(x-y)||g(y)|^2 dy)^{\frac{1}{2}}(\int_{\mathbb{R}^d} |f(x-y)|dy)^{\frac{1}{2}} = c(\int_{\mathbb{R}^d} |f(x-y)||g(y)|^2 dy)^{\frac{1}{2}} \|f\|_{1,\mathbb{R}^d}^{\frac{1}{2}}.$$

Therefore the following bound on the L^2 norm on A holds true:

$$\|f * g\|_A^2 = \int_A |f * g(x)|^2 dx \leq c\,\|f\|_{1,\mathbb{R}^d} \int_A (\int_{\mathbb{R}^d} |f(x-y)||g(y)|^2 dy)dx. \tag{4.58}$$

We now use the fact that $\mathcal{S}$, the support of f, satisfies $diam(\mathcal{S}) < 2\sqrt{d}$ and so, since $x \in A$, it follows $\int_{\tilde{A}^c} |f(x-y)||g(y)|^2 dy = 0$ for $\tilde{A}$ compact set of $\mathbb{R}^d$ such that $\tilde{A} := \left\{\zeta \in \mathbb{R}^d : d(\zeta, A) \leq 2\sqrt{d}\right\}$. Using moreover Fubini's theorem, the right hand side of (4.58) becomes

$$c\,\|f\|_{1,\mathbb{R}^d} \int_A (\int_{\tilde{A}} |f(x-y)||g(y)|^2 dy)dx =$$

$$= c\,\|f\|_{1,\mathbb{R}^d} \int_{\tilde{A}} |g(y)|^2 (\int_A |f(x-y)|dx)dy \leq c\,\|f\|_{1,\mathbb{R}^d}^2 \|g\|_{2,\tilde{A}}^2,$$

where in the last estimation we have enlarged the integration domain of f to $\mathbb{R}^d$. The estimation here above joint with (4.58) gives us

$$\|f * g\|_A^2 \leq c\,\|f\|_{1,\mathbb{R}^d}^2 \|g\|_{2,\tilde{A}}^2.$$

□

Bibliography

[1] Aase, K. K., Guttorp, P. (1987). Estimation in models for security prices. Scandinavian Actuarial Journal, 1987(3-4), 211-224.

[2] Aït-Sahalia, Y. and Yu, J. (2006). Saddlepoint approximations for continuous-time Markov processes, Journal of Econometrics, 134, 507–551

[3] Amorino C, and Gloter A. (2019) Contrast function estimation for the drift parameter of ergodic jump diffusion process. Scand J Statist. 2019;1–68. https://doi.org/10.1111/sjos.12406.

[4] Amorino, C., Gloter, A. (2019). Joint estimation for volatility and drift parameters of ergodic jump diffusion processes via contrast function. arXiv preprint arXiv:1910.11602.

[5] Amorino, C. and Gloter, A. (2019). Unbiased truncated quadratic variation for volatility estimation in jump diffusion processes. arXiv preprint arXiv:1904.10660.

[6] Amorino, C. and Gloter, A., (2020). Invariant density adaptive estimation for ergodic jump diffusion processes over anisotropic classes. arXiv preprint arXiv:2001.07422.

[7] Applebaum, David. Lévy processes and stochastic calculus. Cambridge university press, 2009.

[8] Barndorff-Nielsen, O. E. and Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. J. R. Stat. Soc., Ser. B, Stat. Methodol., 63, 167-241.

[9] Barndorff-Nielsen,O.E.and Shephard,N.(2004). Power and bipower variation with stochastic volatility and jumps. J. Financial Econom. 2 1–48.

[10] Barndorff-Nielsen, O. E., Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps. Stochastic Process. Appl. 116 796–806. MR2218336

[11] Bates, D.S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark. The Review of Financial Studies, 9(1), 69-107.

[12] Bennett, C., Sharpley, R. C. (1988). Interpolation of operators (Vol. 129). Academic press.

[13] Bhattacharya, R. N.: Criteria for recurrence and existence of invariant measures for multidimensional diffusions. Ann. Probab. 6 (4), 541–553 (1978)

[14] Bichteler, K., Gravereaux, J. B. and Jacod, J. Malliavin calculus for processes with jumps, Gordon and Breach Science Publishers, New York, 1987. MR MR1008471 (90h: 60056).

[15] R. M. Blumenthal and R. K. Getoor. Sample functions of stochastic processes with stationary independent increments. J. Math. Mech., 10:493–516, 1961.

[16] Bodo, B. A. and Thomoson, M. E. (1987). A review on stochastic differential equations for applications in hydrology, Stochastic Hydrol. Hydraul., 1, 81–100.

[17] Brownlees CT, Nualart E, Sun Y. On the Estimation of Integrated Volatility in the Presence of Jumps and Microstructure Noise. Available at SSRN 2791342. 2019.

[18] Chen, Z. Q., Hu, E., Xie, L., Zhang, X. (2017). Heat kernels for non-symmetric diffusion operators with jumps. Journal of Differential Equations, 263(10), 6576-6634.

[19] Clément, E. and Gloter, A. (2018). Estimating functions for SDE driven by stable Lévy processes. To appear in Annales de l'Institut Henri Poincaré.

[20] Comte, F., Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. In Annales de l'IHP Probabilités et statistiques (Vol. 49, No. 2, pp. 569-609).

[21] Comte, F., Merlevède, F. (2005). Super optimal rates for nonparametric density estimation via projection estimators. Stochastic processes and their applications, 115(5), 797-826.

[22] Comte, F., Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. ESAIM: Probability and Statistics, 6, 211-238.

[23] Comte, F., Prieur, C., Samson, A. (2017). Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. Stochastic processes and their applications, 127(11), 3689-3718.

[24] Dacunha-Castelle, D., Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. Stochastics: An International Journal of Probability and Stochastic Processes, 19(4), 263-284.

[25] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probab. Theory Relat. Fields, 137(1), 25–47.

[26] Ditlevsen, S. and Greenwood, P. (2013). The Morris–Lecar neuron model embeds a leaky integrate-and-fire model. Journal of Mathematical Biology 67 239-259.

[27] Doukhan, P. (2012). Mixing: properties and examples (Vol. 85). Springer Science and Business Media.

[28] Dufresne, F. and Gerber, H. U. (1991). Risk theory for the compound Poisson process that is perturbed by diffusion, Insurance: Math. Econ., 10, 51–59.

[29] Efromovich, S.: Nonparametric curve estimation. Methods, theory, and applications. Springer Series in Statistics. Springer-Verlag, NewYork, 1999

[30] Embrechts, P. and Shmidli, H. (1994). Ruin estimation for a general insurance risk model, Adv. Appl. Prob., 26, 404–422

[31] Eraker, B., Johannes, M. and Polson N. (2003). The Impact of Jumps in Volatility and Returns. J. Finance, 58(3), 1269-1300.

[32] Feller, W. An Introduction to Probability Theory and its Applications. vol 2 (second edition), Wiley, 1971.

[33] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. Statistics: A Journal of Theoretical and Applied Statistics, 20(4), 547-557.

[34] Funke, B., Schmisser, E. (2018). Adaptive nonparametric drift estimation of an integrated jump diffusion process. ESAIM: Probability and Statistics, 22, 236-260.

[35] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. Statistics, 21(1), 99-116.

[36] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi- dimensional diffusion processes. Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques, 29, 119-151.

[37] Gerber, H. U. (1979). An introduction to mathematical risk theory, S.S. Heubner Foundation Monograph Series, 8. University of Pennsylvania, Wharton School

[38] Gloter, A., Loukianova, D. and Mai, H. (2018). Jump filtering and efficient drift estimation for Lévy-driven SDEs. The Annals of Statistics, 46(4), 1445-1480.

[39] Goldenshluger, A., Lepski, O. (2011). Bandwidth selection in kernel density estimation: oracle inequalities and adaptive minimax optimality. The Annals of Statistics, 39(3), 1608-1632.

[40] Gukhal, C. R. (2001). Analytical valuation of American options on jump-diffusion processes, Math. Finance, 1, 220–242.

[41] Guttrop, P. and Kulperger, R. (1984). Statistical inference for some Volterra population processes in a random environment, Can. J. Statist., 12, 289–302.

[42] Hall, P. and Heyde, C. (1980). Martingale Limit Theory and its Applications,Academic Press, New York.

[43] Hanson, F. B. and Tuckwell, H. C. (1981). Logistic growth with random density independent disasters, Theor. Pop. Biol., 19, 1–18.

[44] Ibragimov, I. A., Has' Minskii, R. Z. (2013). Statistical estimation: asymptotic theory (Vol. 16). Springer Science and Business Media.

[45] Jacod,J.(2008).Asymptotic properties of realized power variations and related functionals of semimartingales. Stochastic Process. Appl. 118 517–559. MR2394762

[46] Jacod, J., and Protter, P. (2011). Discretization of processes (Vol. 67). Springer Science and Business Media.

[47] Jacod, J. and Reiss, M. A remark on the rates of convergence for integrated volatility estimation in the presence of jumps. Annals of stat Volume 42(3), 1131-1144, (2014)

[48] Jacod, J., Shiryaev, A. N. (2003). Limit theorems for stochastic processes, volume 288 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].

[49] Jacod, Jean; Todorov, Viktor. Efficient estimation of integrated volatility in presence of infinite variation jumps. Ann. Statist. 42 (2014), no. 3, 1029–1069. doi:10.1214/14-AOS1213.

[50] Jakobsen, N. and Sørensen, M. (2017). Estimating functions for jump-diffusions, Preprint.

[51] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. Scandinavian Journal of Statistics, 24(2), 211-229.

[52] Kessler, M., Lindner, A., and Sorensen, M. (2012). Statistical methods for stochastic differential equations. Chapman and Hall/CRC.

[53] Klein, T., Rio, E. (2005). Concentration around the mean for maxima of empirical processes. The Annals of Probability, 33(3), 1060-1077.

[54] Kohatsu-Higa, A., Nualart, E. and Tran, N.K., 2017. LAN property for an ergodic diffusion with jumps. Statistics, 51(2), pp.419-454.

[55] Kolokoltsov, V. N. (2011). Markov processes, semigroups, and generators (Vol. 38). Walter de Gruyter.

[56] Kou, S.G. (2002). A Jump-Diffusion Model for Option Pricing. Management Science, 48, 1086-1101.

[57] Kusuoka, S., Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. Probability Theory and Related Fields, 116(4), 457-484.

[58] Kutoyants, Y. A. (2013). Statistical inference for ergodic diffusion processes. Springer Science and Business Media.

[59] Lacour, C., Massart, P., Rivoirard, V. (2017). Estimator selection: a new method with applications to kernel density estimation. Sankhya A, 79(2), 298-335.

[60] Lepski, O. (2013). Multivariate density estimation under sup-norm loss: oracle approach, adaptation and independence structure. The Annals of Statistics, 41(2), 1005-1034.

[61] Lepski, O. V., Levit, B. Y. (1999). Adaptive non-parametric estimation of smooth multivariate functions.

[62] Li, C. and Chen, D. (2016). Estimating jump-diffusions using closed form likelihood expansions. Journal of Econometrics, 195, 51–71

[63] Loéve, M. Probability Theory, Springer, New York, NY, USA, 4th edition, 1977

[64] Mancini, C. (2001). Disentangling the jumps of the diffusion in a geometric jumping Brownian motion. G. Ist. Ital. Attuari LXIV 19–47.

[65] Mancini, C. (2011). The speed of convergence of the threshold estimator of integrated variance. Stochastic processes and their applications, 121(4), 845-855.

[66] Masuda, H. (2004). On multi-dimensional Ornstein-Uhlenbeck processes driven by a general Lévy process, Bernoulli, 10, 97–120.

[67] Masuda, H. (2007). Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. Stochastic processes and their applications, 117(1), 35-56.

[68] Masuda, H. (2013). Convergence of gaussian quasi-likelihood random fields for ergodic lévy driven sde observed at high frequency. Annals of stat., 41(3), 1593-1641.

[69] Merton, R. C. (1973). Theory of rational option pricing. Theory of Valuation, 229-288.

[70] Merton, R.C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3, 125-144

[71] Meyn, S. P. and Tweedie, R. L. (1992). Stability of Markov processes II: Continuous-time processes and sampled chains, Adv. in Appl. Probab., 24, 542–574.

[72] Meyn, S. P., Tweedie, R. L. (1993). Stability of Markovian processes III: Foster–Lyapunov criteria for continuous-time processes. Advances in Applied Probability, 25(3), 518-548.

[73] Mies, F. (2018). State-dependent jump activity estimation for Markovian semimartingales. arXiv preprint arXiv:1811.06351.

[74] Mulinacci, S. (1996). An approximation of American option prices in a jump-diffusion model. Stochastic processes and their applications, 62(1), 1-17.

[75] Mykland, P. A., and Zhang, L. (2006). ANOVA for diffusions and Ito processes. The Annals of Statistics, 34(4), 1931-1963.

[76] Nikolskii, S.M. (1977). Approximation of functions of several variables and imbedding theorems (Russian). Sec. ed., Moskva, Nauka 1977 English translation of the first ed., Berlin 1975

[77] Nualart, David, and Eulalia Nualart. Introduction to Malliavin calculus. Vol. 9. Cambridge University Press, 2018.

[78] B. L. S. Prakasa Rao (1983) Asymptotic theory for non-linear least squares estimator for diffusion processes, Series Statistics, 14:2, 195-209

[79] Prakasa Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. Statistical inference from stochastic processes (Ithaca, NY, 1987).

[80] Protter, P. E. (2004). Stochastic integration and differential equations. 2nd ed. Applications of Mathematics 21. Berlin: Springer.

[81] Rogers, L. C. G.,Williams, D.: Diffusions, Markov processes, and martingales. Vol. 2. Itô calculus.Wiley Series in Probability and Mathematical Statistics. NewYork, 1987

[82] Sato, K. Lévy Processes and Infinitely Divisible Distributions. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 1999.

[83] Schmisser, E. (2013). Nonparametric estimation of the derivatives of the stationary density for stationary processes. ESAIM: Probability and Statistics, 17, 33-69.

[84] Schmisser, E. (2019). Non parametric estimation of the diffusion coefficients of a diffusion with jumps. Stochastic Processes and their Applications, 129(12), 5364-5405.

[85] Scott, L. O. (1997). Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of Fourier inversion methods, Math. Finance, 4, 413–426

[86] Shimizu, Y. (2007). Asymptotic Inference for Stochastic Differential Equations with Jumps from Discrete Observations and Some Practical Approaches (Doctoral dissertation, University of Tokyo).

[87] Shimizu, Y. (2006). M-Estimation for Discretely Observed Ergodic Diffusion Processes with Infinitely many Jumps. Statistical Inference for Stochastic Processes, 9, 179-225.

[88] Shimizu, Y., and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. Statistical Inference for Stochastic Processes, 9(3), 227-277.

[89] Stramer, O., Tweedie, R. L. (1997). Existence and stability of weak solutions to stochastic differential equations with non-smooth coefficients. Statistica Sinica, 577-593.

[90] Strauch, C. (2018). Adaptive invariant density estimation for ergodic diffusions over anisotropic classes. The Annals of Statistics, 46(6B), 3451-3480.

[91] Todorov, Viktor. Jump activity estimation for pure-jump semimartingales via self-normalized statistics. Ann. Statist. 43 (2015), no. 4, 1831–1864. doi:10.1214/15-AOS1327.

[92] Tsybakov, A. B. (2004). Introduction à l'estimation non-paramétrique (Introduction to nonparametric estimation). Mathématiques et Applications (Paris). 41.

[93] Veretennikov, A. Y. (1988). Bounds for the mixing rate in the theory of stochastic equations. Theory of Probability and Its Applications, 32(2), 273-281. ISO 690

[94] Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. Probability theory and related fields, 107(4), 467-492.

[95] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. Journal of Multivariate Analysis, 41, 220-242.

universite PARIS-SACLAY

École doctorale de mathématiques Hadamard (EDMH)

Titre: Correction de biais pour l'estimation de la dérive et de la volatilité d'une diffusion à sauts et estimation non-paramétrique adaptative de la mesure stationnaire

Mots clés: Équations différentielles stochastiques à sauts, haute fréquence, méthode de filtrage des sauts, vitesse de convergence, diffusion ergodique avec sauts, procédure adaptative de sélection de la fenêtre

Résumé: Le sujet de la thèse est l'estimation paramétrique et non-paramétrique dans des modèles de processus à sauts. La thèse est constituée de 3 parties qui regroupent 4 travaux.
La première partie, qui est composée de deux chapitres, traite de l'estimation des paramètres de dérive et volatilité par des méthodes de contraste depuis des observations discrètes, avec pour objectif principal de minimiser les conditions sur le pas d'observation, afin que celui ci puisse par exemple aller arbitrairement lentement vers 0.
La seconde partie de la thèse concerne des développements asymptotiques, et correction de biais, pour l'estimation de la volatilité intégrée.
La troisième partie de la thèse concerne l'estimation adaptative de la mesure stationnaire pour des processus à saut.

Title: Bias correction for drift and volatility estimation of a jump diffusion and non-parametric adaptive estimation of the invariant measure

Keywords: Lévy - driven SDE , high frequency data, threshold estimator, convergence speed, ergodic diffusion with jumps, adaptive bandwidth selection

Abstract: The thesis deals with the parametric and non-parametric inference in jump process models. It consists of 3 parts which gather 4 chapters.
The first part, which contains 2 chapters, focuses on the estimation of the drift and volatility parameters via some contrast function methods starting from a discretely observed process. The main goal is to minimise the conditions on the discretization step so that it can go to 0 arbitrarily slowly.
The second part of the thesis regards some asymptotic developments, and bias correction, for the estimation of the integrated volatility.
The third part of the thesis is about the adaptive estimation of the invariant measure for jump processes.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France