

Asymptotic Inference for Stochastic Differential
Equations with Jumps from Discrete Observations
and Some Practical Approaches

by
YASUTAKA SHIMIZU

Division of Mathematical Science

Graduate School of Engineering Science
Osaka University, Toyonaka, Osaka 560-8531 Japan.
E-mail: yasutaka@sigmath.es.osaka-u.ac.jp

Dissertation
submitted to the University of Tokyo
for the degree of
Doctor of Mathematical Sciences

Osaka, Japan, October 2007

© Yasutaka Shimizu 2007

Acknowledgements

I would like to express my sincere thanks to Professor Nakahiro Yoshida, who was my supervisor while I was a student of the University of Tokyo, for giving me many valuable instructions and advices. It was a great pleasure and privilege in my life that I had received his solicitous education, and that I could get close with the great teacher throughout my student days.

I am also grateful to Professor Yutaka Kano and Associate Professor Masayuki Uchida, who are professors of Osaka university where currently I belong as an assistant professor. They always gave me a lot of help and encouragement both officially and privately.

Contents

1	Introduction	9
1.1	A brief historical review of the asymptotic inference	10
1.2	Overview of this thesis and its background	13
1.3	General notation	16
2	Diffusion processes with jumps	19
2.1	Stochastic differential equations with jumps	19
2.1.1	Solution-processes	19
2.1.2	A jump mechanism	21
2.2	Finite and infinite activity for jump parts	22
2.3	Ito's formula and a differential operator	24
2.4	Ergodic diffusion processes with jumps	27
3	Parametric estimation in finite activity models	31
3.1	Setting of the model	31
3.2	Discussion and conclusions	32
3.2.1	Assumptions and examples	32
3.2.2	Contrast functions and efficient estimators	35
3.2.3	Construction of contrast functions	38
3.3	Moment estimates in the finite activity case	44
3.4	Limit theorems and some remarks	48
3.5	Proof of the main theorem	64
3.5.1	Proof of consistency	64
3.5.2	Proof of asymptotic normality	66
4	Parametric estimation in infinite activity models	79
4.1	Models and assumptions	79
4.2	Selection of data	82

4.3	Estimating functions and asymptotic results	86
4.3.1	Moment type estimating functions	86
4.3.2	Consistency and asymptotic normality	89
4.4	Examples and simulation study	91
4.5	Moment estimates in infinite activity case	99
4.6	Limit theorems	103
4.7	Proofs of the main theorems	111
4.7.1	Proof of consistency	111
4.7.2	Proof of asymptotic normality	112
4.7.3	Proof of Lemma 4.8	113
4.7.4	Proof of Lemma 4.9	115
5	Nonparametric estimation of Lévy densities	121
5.1	Nonparametric framework	121
5.2	Notations and assumptions	123
5.3	Density estimation and the optimal rate	124
5.3.1	Continuously observed case	124
5.3.2	Discretely observed case	126
5.4	Simulation study	128
5.4.1	Simple examples	128
5.4.2	Some troubles in the estimation	129
5.4.3	Selection problem of the filter	130
5.4.4	Data adaptive filter	134
5.5	Conclusion of this chapter	137
5.6	Some moment estimates	139
5.7	Proofs of the main theorems	143
5.7.1	Proof of Theorem 5.1	143
5.7.2	Proof of Theorem 5.2	144
6	Practical inference from finite samples	149
6.1	Asymptotic filter and its problem	149
6.2	A criterion for selecting the filter	152
6.3	A bias correction	155
6.4	Direct plug-in method	158
6.4.1	Plug-in rule	158
6.4.2	Simulation results	159
6.5	Theoretical discussion	164

6.5.1	What is a validity?	164
6.5.2	Mathematical validity	164
6.6	To a practical approach	172
A	Central limit theorems for arrays	175
A.1	Martingale difference arrays	175
A.2	General triangular arrays	178
B	Weak Convergence in C-space	181
B.1	Weak Convergence and tightness	181
B.2	Tightness criteria	184
	Bibliography	187

Chapter 1

Introduction

In all situations that involve uncertainties, we must do decision-makings based on some inference. The inference will be done from the past observations for the corresponding phenomena. We collect various types of data, build a model of the corresponding phenomena, and estimate the unknowns. We again collect new data, find out if the model is suitable or not, and correct the model if we need. Finally we will make a decision through the final model. Such an inference procedure is desired that the larger the sample size becomes, the more accurately the model can predict the future's phenomena. The asymptotic inference is a field to study such inference procedures and properties of estimators under the situation where the sample size tends to infinity.

One may think that it is unrealistic to consider such a situation where the sample size tends to infinity. However, as Ibragimov and Has'minskii [41] says by quoting the word of Gnedenko and Kolmogorov [34],

The epistemological value of *the theory of statistical estimation* revealed only by *asymptotic theories*.

The asymptotic inference, which gives a mathematical validity of the constructed procedures and comparing methods with the another procedure, is one of the most important steps in the prediction.

In this thesis, we are interested in the asymptotic inference for continuous-time stochastic processes that follow some stochastic differential equations with jumps. However, as a prologue, we shall begin to introduce the general history of the asymptotic inference briefly.

1.1 A brief historical review of the asymptotic inference

In the history of the asymptotic inference, it would be well known that several Statistics Giants; Pearson, Fisher, Cramér, Wald, Le Cam, Hájek, and so on have largely contributed to the development of the theory.

The notion of the consistency of estimators, which is the most important and basic notion in the asymptotic theory, is suggested by Fisher [27]. After that, consistent estimators have been sufficiently studied by many authors. It seems that Pearson, K. suggested the method of moments; see Cramér [20], which is the first general method to construct estimators. The maximum likelihood estimators, which is originated with Fisher [27], are discussed in a general framework by Wald [110] and Le Cam [57], Bayesian estimators were also discussed by Le Cam [57], and the study of the asymptotic inference were becoming increasingly popular.

It would be natural that the next interest was to define the *best* estimator. In statistics, there was, from long ago, a notion of comparing an *expected loss (risk)* to discuss the *goodness* of estimators. This idea goes back to Laplace or Gauss, who proposed to minimize the expected absolute deviation or the least-squares loss. It was reintroduced by Wald [108, 111] in a sophisticated form to the statistical scene two decades after them, and the notion of the *asymptotic efficiency*, whose fundamental notion was also probably originated with Fisher [27], comes from the fusion of the decision theory and the asymptotic theory. Moreover the development of the modern mathematical statistics based on the likelihood ratio by Wald, Le Cam, Hájek and other authors improved and completed the notion of the asymptotic efficiency via the concept of the *local asymptotic normality* (LAN) or more generally *local asymptotic mixed normality* (LAMN); see Le Cam [58] and Jeganathan [45]. These are so called an *asymptotic risk minimization procedure*.

Their theories were discussed on the extremely general framework called the *statistical experiment*, which was merely a family of probability measures $\mathcal{E}_\varepsilon = \{P_\theta^{(\varepsilon)}, \theta \in \Theta\}$ on a σ -field of a state space, where $\varepsilon > 0$ and $\theta \in \Theta$ were parameters. Therefore it included not only *static* statistical models as in the traditional large sample theory but also many dynamical models by stochastic processes. Moreover the notion of minimizing of the *asymptotic risk* were applied not only to the parametric framework as above but also to the nonparametric framework. On these historical flow and the theory itself, it is familiar in e.g. Ibragimov and Has'minskii [41], Le Cam [59], and Basawa and Scott [8]. In this way, a general, rigorous and massive system of the asymptotic

inference had been constructed.

Against such a background of modern statistical theories, many authors studied the inference for stochastic processes. It would be a natural flow in the history since the theory of stochastic processes had been developing coincidentally, and were just beginning to be applied to some fields. One of the important and essential tools for investigations of the inference for stochastic processes was the likelihood ratio or the log-likelihood function.

The inference for discrete-time stochastic processes has been investigated earlier by Wald [109]. Billingsley [9], Roussas [83, 84, 85], Prakasa Rao [76], and so on studied the discrete-time Markovian case. For long-memory time series, the case of linear processes that includes important AR, MA and ARMA models was studied earlier by Whittle [112] systematically, the case of non-linear processes that includes recently well-used ARCH and GARCH were by Tjøstheim [104]. Kreiss [51] and Jeganathan [46] discussed the inference for linear and non-linear time series in the framework of LAN and LAMN, respectively. For the further review, see the recent great book Taniguchi and Kakizawa [101] for the asymptotic inference for time-series.

On the other hand, the first systematic treatment of problems in the statistical inference for continuous-time stochastic processes was due to Grenander [36]. In the case of continuous-time processes, the calculation of the likelihood ratio is frequently difficult. He considered a special Gaussian stationary Markov process, and tried to calculate the likelihood ratio by the following schemes: calculate the likelihood ratio with respect to the true distribution for a finite set of time points $0 = t_0 < t_1 < \dots < t_n = T$ and then let the sample size $n \rightarrow \infty$ so that $\max_{1 \leq i \leq n} (t_i - t_{i-1}) \rightarrow 0$. This method can not be applied to general stochastic processes but would give us an important insight for the inference from discrete observations later.

After him, the asymptotic inference for each special type of continuous-time stochastic processes has been studied based on the likelihood theory by several authors. Billingsley [9] studied the Markov processes with the general state space. Athreya and Keiding [3], Feigin [25] considered the Markov branching processes. Brown [18] and Brillinger [17] studied point processes. Basawa and Brockwell [6] discussed gamma and stable processes, and Akritas and Johnson [2] studied general Lévy processes. Moreover Kutoyants [54] studied the LAN properties for the likelihood ratio of diffusion-type processes and point processes after the manner of Ibragimov and Has'minskii [41], and Yoshida [115, 116] discussed the M -estimation for diffusion processes via the LAMN property.

In 1990s, the asymptotic inference for stochastic processes began to be discussed in

the general framework of semimartingales that was a wide class of stochastic processes including point processes, Lévy processes, diffusion processes and also diffusion processes with jumps. The LAMN property discussed by Basawa and Scott [8] in general framework was applied to the class of semimartingales by Luschgy [62], and he also introduced the new concept of the *local asymptotic quadraticity* (LAQ) in Luschgy [63]. Taraskin [102] extended the concept of the LAN for semimartingales to *local asymptotic infinite divisibility* (LAID), and so on. On the other hand, without the LAN or the LAMN theory, the asymptotic likelihood theory based on the notion of various information quantities were introduced by Barndorff-Nielsen and Sørensen [11] and Küchler and Sørensen [52]. Sørensen [98, 99] gives the concrete discussion of the asymptotic inference for continuously observed diffusion processes with jumps using their theories. Basawa and Prakasa Rao [7] and Prakasa Rao [80] give comprehensive explanations on these historical flow and theories.

In these days, the higher-order asymptotic inference for semimartingales via the asymptotic expansion approach is becoming active at the initiative of Yoshida [117, 119, 120], Sakamoto and Yoshida [87, 88], Uchida and Yoshida [107] and so on. Their asymptotic expansion approaches, which were initiated by Yoshida, N., are practically applied to the financial econometrics, and are expected to be powerful tools for the study of the mathematical and applied finance; see Kunitomo and Takahashi [53], Takahashi and Yoshida [103], Masuda and Yoshida [68] and the references therein.

Until a little before, such a statistical theory for semimartingales was possibly only for theorists, and practitioners were not so familiar with that. However we are compelled to wonder the situation has been changed in these days. That is, not only professional mathematicians and statisticians but also practitioners who use statistics in their business are becoming familiar with the term of “stochastic processes”, and statistical inference for them is of major interest even for practitioners. It is nothing else that there has been the kind of emphasis on the modeling by stochastic processes in applications recently.

The remarkable tendency as above is particularly seen in the fields of econometrics or financial engineerings. Since the appearance of two great papers in the field of econometric; Black and Scholes [14] and Merton [70], modern economic theorists are busy to formulate their theories to represent the dynamical change of securities' prices by building dynamic models via semimartingales rather than the classical econometric theory, which is erected on the traditional static or equilibrium structure as classical statistical mechanics. At the same time, practitioners are also busy to learn their new theories and try to use them in their businesses. If it is doubtful, you should go to

the corner of Finance and Marketing in some big book stores. You could see many terminologies on stochastic calculus and mathematical statistics there. In this way, the statistical inference for semimartingales is now becoming a standard tool not only in academic but also even in practice.

However, it is quite another between the great advance of the inference theory for stochastic processes and a correct understanding and application of the theory. Actually, in application, invalid *ad hoc* methods are sometimes used although there is the corresponding theory that is mathematically well-established. One of the reasons would be the mathematical difficulty of the inference theory. However it seems that there would be more fundamental problems. One is that the cooperation among mathematical statisticians, applied statisticians and practitioners is not so smooth. Another is that there still remains gap between the theory and the practice.

Fortunately, it seems that the former problem has been changing in a favorable way recently. Many academic conferences and meetings together with professional mathematicians, statisticians and practitioners in financial institutions are held in the world. Therefore we are going to study further in order to solve the latter problem.

1.2 Overview of this thesis and its background

The purpose of this thesis is the asymptotic inference for the following d -dimensional stochastic differential equations with jumps:

$$X_t = X_0 + \int_0^t a(X_s) ds + \int_0^t b(X_s) dw_s + \int_0^t \int_{\mathbb{R}^d \setminus \{0\}} c(X_{s-}, z) r(ds, dz)$$

that includes some unknown quantities. Especially we are interested in the inference from discrete observations obtained at the time points $t_i^n = ih_n$ ($i = 0, 1, 2, \dots, n$); $\{X_{t_i^n}\}_{i=0}^n$, where h_n is the length of the observational interval, w is a Wiener process, and r is a compensated Poisson random measure; see Chapter 2 for details. If the compensated random measure r vanishes; $r \equiv 0$, then we call X simply a diffusion process.

In the history of the inference for stochastic processes, such types of models have occupied the much attention of many statisticians as widely used models in applications. For example, Modeling prices of securities in financial markets is particularly important; see Merton [70], Aase and Guttorp [1], Mulinacci [74], Scott [89], Gukhal [37], and multiplicity of recent papers on the mathematical finance. There are other applications to risk models in insurance; Gerber [32], Dufresne and Gerber [22], Embrechts and Schmidli [23], soil moisture models; Mtundu and Koch [75], to hydrology;

Bodo and Thomason [16], and to population models; Hanson and Tuckwell [40], Guttorp and Kulperger [38], and so on. However, in any application, observable data are always discrete in practice, and we face the trouble of the inference for such continuous-time models from discrete samples.

In the early period of modeling by continuous-time stochastic processes, modeling by diffusion processes was prevalent, and estimation problems for discretely observed diffusion processes have been studied by many authors very well. Though there are some observation schemes, we are here interested in the one called "*rapidly increasing experimental design*": $nh_n \rightarrow \infty$ but $nh_n^2 \rightarrow 0$ as $n \rightarrow \infty$; it is also-called *frequent data* in financial literatures. The earlier work on this scheme is seen in Prakasa-Rao [77, 78]. He studied the least squares approach. Florens-Zmirou [29] considered an estimation of one-dimensional diffusion processes with constant diffusion coefficients under the less restrictive condition $nh_n^3 \rightarrow 0$ with convergence rate $\sqrt{nh_n}$ for a diffusion estimation. Yoshida [118] studied the case where the drift-diffusion parameter estimation cannot be split, and showed the joint convergence of an adaptive estimator with \sqrt{n} convergence rate for diffusion parameter. After that, Kessler [49, 50] improved it to a more general case with the design $nh_n^p \rightarrow 0$ for arbitrary fixed $p \geq 2$. For other schemes, for instance $nh_n = \text{constant}$ or $h_n = \text{constant}$, see Genon-Catalot and Jacod [31], Dacunha-Castelle and Florens-Zmirou [21], Bibby and Sørensen [12], and the references therein. Moreover, for small diffusion models, see Uchida [105, 106] and Sørensen and Uchida [100]. In this way, the inference for discretely observed diffusion processes have been well discussed.

On the other hand, the inference for discretely observed diffusion processes with jumps is still developing although jump-diffusion models are also well used in many fields. We propose some useful methodologies for such a critical problem in this thesis.

Our work gives several estimation methods, which are practical and mathematically valid for such extremely important models. Though the author's work is only a small part of the enormous history of the asymptotic statistics as described in Section 1.1, the author strongly believes that this work provides an absolutely significant foothold for the future's development of the inference theory for stochastic processes.

This thesis mainly consists of three already published papers that are devoted to the inference for stochastic differential equations with jumps from discrete observations; Shimizu and Yoshida [96], Shimizu [92, 93], and some unpublished new results; Shimizu [94].

In Chapter 2, we shall first present the definition of diffusion processes with jumps. We identify two types of models with a jump mechanism. One is a *finite activity model*

in which jumps occur only finitely many times in any finite time interval, and an *infinite activity model* in which jumps occur infinitely many times in any finite time interval. We introduce an aspect that the infinite activity model can be approximated by the finite activity model in some sense under some regularity conditions. This aspect is the key in Chapter 4. We also introduce Ito's formula and an ergodicity for diffusion processes with jumps, which are important in Chapter 3 and Chapter 4 to obtain asymptotic properties of estimators.

Chapter 3 is devoted to the parametric estimation of finite activity models under the scheme that $nh_n^2 \rightarrow 0$. We assume the ergodicity of the model to discuss the asymptotic behavior of estimators. We construct a single contrast function to estimate the drift and the diffusion parameters jointly. The exact likelihood function would be used if we would know the form of transition probabilities. However it is generally impossible to write it down explicitly, therefore we have to approximate it by a suitable function. We present an estimating function having two parts: the first part is the log likelihood of a local Gaussian process, and the second one is modeled after the likelihood of Poisson random measures. This estimating function divides increments of neighboring data according to their magnitudes and assigns them those parts, that is, a *small* increment that is less than or equals to a *threshold* r_n is regarded a Brownian shock, and is assigned to the first part that corresponds to the estimating function of the continuous part, and a *large* increment that is larger than the r_n is regarded a jump, and is assigned to the second part that corresponds to the one of the discontinuous part. This threshold r_n should satisfy some order-conditions. Evaluating the probability of misclassification, we prove the asymptotic normality of our estimator. Our estimator is of the maximum likelihood type, which is efficient in some sense.

In Chapter 4, we consider a more general ergodic jump-diffusion whose jump part is driven by a Lévy process z whose Lévy density f satisfies $\int_{\mathbb{R}} f(z) dz = \infty$; infinite activity models. This z can be split into two parts, the one is for a *small* jump part whose jumps are less than or equal to a positive value $\varepsilon_n (\downarrow 0)$, and the others are for a *large* jump part whose jumps are larger than ε_n . The former can be regarded as a *small* diffusion, and the latter becomes a compound Poisson Process for each n . This is the approximation of the infinite activity models by the finite activity models; see Chapter 2. This enables us to apply the idea of Chapter 3 to the infinite activity models. However we have to choose the sequence ε_n carefully for that purpose.

For estimation of parameters in the continuous part, we can take the same procedure as in Chapter 3. The parameters in the jump part are estimated via the method of moments fitting the higher-order moments of *large* increments. However, in order to

obtain the asymptotic normality, it needs some conditions for the intensity of *large* jumps: $\lambda^{(\varepsilon_n)} = \int_{|z| > \varepsilon_n} f(z) dz$. Therefore, our method can not always be applied to any infinite activity model.

In Chapter 5, the nonparametric estimation is discussed in reversal. In particular, we concentrate on the estimation of Lévy density, which is the most important in jump-processes, in finite activity models. The procedure is the kernel density method. We regard *large* increments as jumps approximately, and apply the idea of the usual density estimation. Evaluating the negligible gap between the approximated jump size and the true jump size, we prove the consistency of the density-type estimators in the sense of the mean squared error. The restriction of the Lévy density which was imposed in Chapter 3 is removed in this chapter although the experimental design becomes more rapid; $nh_n^{1+\delta} \rightarrow 0$ for a $\delta \in (0, 1/2)$.

Up to Chapter 4, we need an ergodicity of the model to obtain asymptotic results. However, this assumption is often strong in some applications. The nonparametric procedure proposed in Chapter 5 enables us to estimate non-ergodic models from sampled data.

In the simulation study; Section 5.4, we point out an important problem: How should we determine the threshold r_n for fixed n ? As described above, r_n is restricted only by some order-conditions, and can not be determined uniquely by asymptotic theories. However since the sample size n is finite in real data, we have to select this threshold r_n according to models and the sample size n . This is the very gap between the theory and the practice. In Section 5.4.3 and 5.4.4, we shall give some intuitive methods to select the threshold, and in a special case of the model, we shall propose more theoretical and practical methods than them in Chapter 6.

1.3 General notation

1. $\mathbb{R} = (-\infty, \infty)$, $\mathbb{R}_+ = [0, \infty)$, $\mathbb{N} = \{1, 2, \dots\}$, $\mathbb{N}_0 = \{0\} \cup \mathbb{N}$. Moreover $\mathcal{B}(X)$ means the family of Borel sets on a normed space X .
2. $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ is a filtered probability space with a filtration $(\mathcal{F}_t)_{t \geq 0}$ and a probability measure P . We denote by E the integral with respect to P .
3. For a sequence of increasing positive numbers $\{t_i^n\}_{i=1}^n$ for each $n \in \mathbb{N}$, we put $\mathcal{F}_{i-1}^n := \mathcal{F}_{t_{i-1}^n}$, $P_{i-1}^n\{\cdot\} := P\{\cdot | \mathcal{F}_{i-1}^n\}$ and $E_{i-1}^n[\cdot] = E[\cdot | \mathcal{F}_{i-1}^n]$.
4. For a function $g(x, y)$, we denote the value $g(x, X_{t_{i-1}^n})$ by $g_{i-1}(x)$.

5. For a stochastic flow X ., we set $\Delta X_t := X_t - X_{t-}$ and $\Delta_i X^n := X_{t_i^n} - X_{t_{i-1}^n}$.
6. The symbol $\mathcal{N}_d(\mu, \Sigma)$ means a d -dimensional Gaussian distribution with the mean vector μ and the covariance matrix Σ . As $d = 1$, we simply write $\mathcal{N}(\mu, \Sigma)$.
7. The symbol \xrightarrow{P} means the convergence in probability under the measure P . The symbol \xrightarrow{d} means the convergence in distribution.
8. $Q_n = O_p(r_n)$ means that, for every $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ and $n_\varepsilon > 0$ such that $P(|Q_n/r_n| > \delta_\varepsilon) < \varepsilon$ for any $n > n_\varepsilon$. $Q_n = o_p(r_n)$ means that $Q_n/r_n \rightarrow 0$ as $n \rightarrow \infty$.
9. For $\kappa = (\kappa_1, \dots, \kappa_d)$, $\partial_\kappa := (\partial_{\kappa_1}, \dots, \partial_{\kappa_d})^*$, $\partial_{\kappa_j} := \frac{\partial}{\partial \kappa_j}$, $\partial_{\kappa_j}^2 := \frac{\partial^2}{\partial \kappa_j^2}$, $\partial_{\kappa_i \kappa_j}^2 := \frac{\partial^2}{\partial \kappa_i \partial \kappa_j}$, and so on, where $*$ stands for the transpose. Moreover, for any integer l , denote $\partial_\kappa^l f(\kappa) = (\partial_{\kappa_{i_1} \dots \kappa_{i_l}}^l f(\kappa))_{1 \leq i_1, \dots, i_l \leq d}$; a tensor on $(\mathbb{R}^d)^{\otimes l}$.
10. For a tensor A , we express its components with upper index, for example, if A is a matrix, then its (k, l) -component is $A^{(k, l)}$. Moreover $|A|^2$ is the sum of squares of the components of A .
11. Let u_n be a real valued sequence and s, x be some vectors whose components are real valued. We denote by $R(s, u, x)$ a real valued function for which there exists a constant C such that

$$R(s, u_n, x) \leq u_n C(1 + |x|)^C$$
 uniformly in s . Furthermore we set $\tilde{R}(s, u_n, x) = 1 - R(s, u_n, x)$.
12. When ϑ is an unknown parameter, we express the true value of ϑ with the subscript zero: ϑ_0 is the true value of a parameter ϑ .
13. We sometimes omit the true values of parameters for simplicity of notations without specially mentioning. For example, we simply write $f(X_{t_{i-1}^n})$ or more simply f_{i-1} for $f(X_{t_{i-1}^n}, \theta_0)$, or $q(ds, dz)$ for $q_{\theta_0}(ds, dz)$, and so on, where θ is an unknown parameter.
14. We often use the notation C (resp. C_k) as an universal positive constant (resp. depending on the index k), therefore we sometimes use the same character for different constants from line to line without specially mentioning.

Chapter 2

Diffusion processes with jumps

When statisticians observe natural phenomena, they usually suppose some dynamical systems which have generated the obtained data. A dynamical system seems to always be disturbed by some stochastic perturbation; *noise*, and the *noise* makes it difficult to construct the model and to predict the future. Particularly, when we deal with time-continuous systems, modeling of the noise is complicated.

Fortunately, we have the powerful tool to model the noise. One type of which is a continuous noise modeled by a stochastic integral with respect to a Wiener process. The other is a jump type noise modeled by a stochastic integral with respect to a martingale measure generated by a point process. Stochastic processes modeled by differential equations including such kinds of noise terms are called stochastic differential equations (SDE) with jumps.

Our major interest of this thesis is the statistical inference for a certain class of SDE's with jumps; diffusion processes with jumps, or simply called jump-diffusions, from data obtained in the past. In this chapter, we give the definition and some properties about diffusion processes with jumps, and present some auxiliary results which are useful in the later chapters.

2.1 Stochastic differential equations with jumps

2.1.1 Solution-processes

On a probability space (Ω, \mathcal{F}, P) with a filtration $(\mathcal{F}_t)_{t \geq 0}$, we consider a d -dimensional stochastic differential equation with jumps: X_0 is a random variable, and

$$dX_t = A(\omega, t) dt + B(\omega, t) dw_t + \int_{\mathcal{E}} C(\omega, t, z) r(\omega, dt, dz), \quad (2.1)$$

where $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, $w_t^* = (w_t^1, \dots, w_t^r)$ ($t \geq 0$) is an r -dimensional Wiener process, $A(\omega, t)$ and $B(\omega, t)$ are jointly measurable and \mathcal{F}_t -adapted processes, $C(\omega, t, z)$ is also jointly measurable, but an \mathcal{F}_t -predictable process for each $z \in \mathcal{E}$, $r(\omega, dt, dz)$ is a compensated Poisson martingale measure of the form $r(\omega, dt, dz) = p(\omega, dt, dz) - q(dt, dz)$, p is an extended Poisson random measure independent of w , and q is its intensity measure, that is, $q(dt, dz) = E[p(\cdot, dt, dz)]$; see Jacod and Shiriyayev [43] for details of random measures.

Denote by (D_T, \mathcal{G}_T) the measurable space of càdlàg functions $x = (x_t)_{0 \leq t \leq T}$ for each $T > 0$ with a filtration $\mathcal{G}_t = \sigma\{x_s; s \leq t\}$.

Let us present a general definition of diffusion processes with jumps.

Definition 2.1 *A stochastic process $X = (X_t)_{0 \leq t \leq T}$ satisfying the equation (2.1) is called a diffusion process with jumps, or a jump-diffusion if there exist jointly measurable (s, x) functions $a(s, x)$ and $b(s, x)$, which are \mathcal{G}_{s+} -measurable for each s such that, for almost all $\omega \in \Omega$ and $s \in [0, T]$, $A(\omega, s) = a(s, X(\omega))$, $B(\omega, s) = b(s, X(\omega))$. Moreover there exist a jointly measurable (s, x, z) function $c(s, x, z)$, which are \mathcal{G}_{s-} -measurable for each s and $z \in \mathcal{E}$ such that, for almost all $\omega \in \Omega$, $s \in [0, T]$ and $z \in \mathcal{E}$, $C(\omega, s, z) = c(s, X(\omega), z)$.*

In this thesis, we particularly consider the following d -dimensional Markovian diffusion processes with jumps: $X_0 = x$, and

$$dX_t = a(X_t) dt + b(X_t) dw_t + \int_{\mathcal{E}} c(X_{t-}, z) r(\omega, dt, dz), \quad (2.2)$$

where x is a random variable.

Definition 2.2 *A solution-process to (2.2) is a càdlàg process X defined on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$ endowed with (w, p) as the driving terms, such that for each $t \geq 0$,*

$$X_t = x + \int_0^t a(X_s) ds + \int_0^t b(X_s) dw_s + \int_0^t \int_{\mathcal{E}} c(X_{s-}, z) r(\omega, ds, dz).$$

The following classical result of the existence and the uniqueness of the solution-process is found in Jacod and Shiriyayev [43]; see Theorem III, 2.32.

Theorem 2.1 *Assume the following two conditions*

- (i) **(Local Lipschitz continuity)** *For each $n \in \mathbb{N}$, there exist a constant L_n and a function $\zeta_n : \mathcal{E} \rightarrow \mathbb{R}_+$ with $\int_{\mathcal{E}} \zeta_n^2(z) f(z) dz < \infty$ such that, for any $|x| \leq n$, $|y| \leq n$,*

$$|a(x) - a(y)| + |b(x) - b(y)| \leq L_n |x - y|, \quad |c(x, z) - c(y, z)| \leq \zeta_n(z) |x - y|.$$

- (ii) **(Linear growthness)** For each $n \in \mathbb{N}$, there are L_n and ζ_n as above, such that for all $x \in \mathbb{R}^d$,

$$|c(x, z)| \leq \zeta_n(z)(1 + |x|).$$

Then the equation (2.2) has the unique solution X on the probability space (Ω, \mathcal{F}, P) . Here the uniqueness of the solution means that, $P\{X_t = Y_t, \text{ for all } t \in \mathbb{R}_+\} = 1$ if there exists an another solution-process Y .

2.1.2 A jump mechanism

A solution-process X to (2.2) is a càdlàg process, which transits continuously unless a jump occurs. Jumps are driven by the point process $p_t(z) = p((0, t] \times (-\infty, z])$, that is, X has a jump $\Delta X_t = c(X_{t-}, z)$ if the point process $p_t(z)$ has a jump $\Delta p_t = p_t(z + dz) - p_{t-}(z)$ at a point (t, z) . Therefore p_t may be considered as the random counting measure of jumps of a càdlàg process X , that is,

$$p(dt, dz) = \sum_{s \geq 0} \mathbf{1}_{\{\Delta X_s \neq 0\}} \mathbf{1}_{\{(s, z); \Delta X_s = c(X_{s-}, z)\}}(dt, dz) \quad (2.3)$$

On the other hand, we impose the following assumptions on the intensity measure q throughout this thesis: $q(\{t\} \times \mathcal{E}) = 0$ for each $t \in \mathbb{R}_+$, and it has the form

$$q(dt, dz) = dt \times f(z) dz, \quad (2.4)$$

where f is a density of a positive σ -finite measure. In this case p is called a *time-homogeneous Poisson random measure*, and these conditions are equivalent to that q is the compensator of p . Hence $r((0, t], \cdot) = (p - q)((0, t], \cdot)$ becomes a martingale measure.

In order to construct such a Poisson random measure, it is convenient to consider that the jump mechanism of X is controlled by an \mathbb{R}^d -valued Lévy process $z = (z_t)_{t \geq 0}$ defined on $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, P)$ whose Lévy density is $f(z)$. That is, we consider that X has a jump if z does. Then the expression of (2.3) is rewritten as follows:

$$p(dt, dz) = \sum_{s \geq 0} \mathbf{1}_{\{\Delta z_s \neq 0\}} \mathbf{1}_{\{(s, \Delta z_s)\}}(dt, dz). \quad (2.5)$$

We call such a p a *random measure associated to z* . Intuitively speaking, z_t has a jump at time t , then X has a jump $c(X_{t-}, \Delta z_t)$, and this is actually an integer-valued random measure; see Jacod and Shiriyayev [43], Proposition II.1.16. If $(z_t)_{t \geq 0}$ is a compound Poisson process, which is the most important case in applications, we have the following proposition:

Proposition 2.1 *Let N be a Poisson process with the intensity λ and $\{\varepsilon_i\}_{i \in \mathbb{N}}$ is an i.i.d. sequence with a probability density $F(z)$. Let p be a random measure (2.5) associated to a compound Poisson process $z_t = \sum_{i=1}^{N_t} \varepsilon_i$. Then $q(dt, dz) = dt \times \lambda F(z) dz$.*

Therefore we see that $f(z) = \lambda F(z)$. More generally, the following proposition implies that the intensity measure q of a random measure p of the form (2.5) also satisfies (2.4).

Proposition 2.2 *Let A be a Borel subset of \mathbb{R}^d with $0 \notin \bar{A}$, and g is a Borel function which is finite on A . Let f be the Lévy density of a Lévy process $(z_t)_{t \geq 0}$. Then*

$$E \left[\int_0^t \int_A g(z) p(ds, dz) \right] = t \int_A g(z) f(z) dz.$$

See e.g. Protter [82] for these results. We have (2.4) if $g \equiv 1$. Proposition 2.2 implies that the integral $\int_A f(z) dz$ can be interpreted as the average of the number of jumps whose sizes are in the set A per unit of time.

Now, let us suppose that $\int_{\mathcal{E}} c(x, z) f(z) dz < \infty$ for all $x \in \mathbb{R}^d$. Under this condition, the stochastic integral with respect to the compensated random measure r can be split into the two integrals with respect to p and q ; see Jacod and Shiriyayev [43], Proposition II.1.28, and the SDE of (2.2) can be rewritten as follows:

$$dX_t = \tilde{a}(X_t) dt + b(X_t) dw_t + \int_{\mathcal{E}} c(X_{t-}, z) p(dt, dz),$$

where $\tilde{a}(x) = a(x) - \int_{\mathcal{E}} c(x, z) f(z) dz$. This SDE implies that X follows the diffusion process $dX_t = \tilde{a}(X_t) dt + b(X_t) dw_t$ while z_t does not jump.

In this way, considering that the random measure p is a random measure associated to a Lévy process, we can easily understand the pathwise properties of diffusion processes with jumps. Therefore we often stand on such a point of view in the latter chapters to make the intuitive discussion be clear.

2.2 Finite and infinite activity for jump parts

In this section, let us investigate a structure of jump's mechanism.

Let $z = (z_t)_{t \geq 0}$ be a multidimensional Lévy process with the Lévy density f . If z is a compound Poisson process with $f = \lambda F$ then Proposition 2.1 implies that z can be represented as

$$z_t = \sum_{s \in [0, t]} \Delta z_s = \int_0^t \int_{\mathcal{E}} z p(ds, dz), \quad (2.6)$$

where p is a random measure associated to z . Since $\int_{\mathcal{E}} f(z) dz = \lambda < \infty$, the stochastic integral in (2.6) is finite for each $t > 0$. However if the above z is a general Lévy process then $\int_{\mathcal{E}} f(z) dz$ would possibly be infinite; f can have a singularity at the origin. Then there can be infinitely many small jumps in any finite time interval, and the sum (2.6) does not necessarily converge. On the other hand, it is well known that a Lévy measure satisfies, for any $\varepsilon > 0$, that

$$\int_{|z|>\varepsilon} f(z) dz < \infty, \quad \int_{0<|z|\leq\varepsilon} |z|^2 f(z) dz < \infty. \quad (2.7)$$

The first condition implies that the sum $z_t^{(\varepsilon)} = \sum_{s \in [0,t]} \Delta z_s \mathbf{1}_{\{|\Delta z_s|>\varepsilon\}}$ is finite almost surely, that is, it is a compound Poisson process. We have the following *Lévy-Ito decomposition* for the process $z - z^{(\varepsilon)}$; see Protter [82], Theorem I.42.

Proposition 2.3 *Let $z = (z_t)_{t \geq 0}$ be a d -dimensional Lévy process, and r be a compensated Poisson random measure associated to z_t . Then there exist constants a, σ and a d -dimensional standard Brownian motion B such that*

$$z_t - z_t^{(\varepsilon)} = at + \sigma B_t + \lim_{\delta \downarrow 0} \int_0^t \int_{\delta \leq z \leq \varepsilon} z r(ds, dz) \quad a.s. \quad (2.8)$$

for any $\varepsilon > 0$, where $z_t^{(\varepsilon)} = \int_0^t \int_{|z|>\varepsilon} z p(ds, dz)$. The terms in the right-hand side and $z^{(\varepsilon)}$ are independent each other. the constant a might depend on ε . In particular, if $\sigma \equiv 0$ then we call z a pure jump Lévy process.

Again let us consider the SDE (2.2). Let $p^{(\varepsilon)}$ be a random measure associated to a compound Poisson process $z^{(\varepsilon)}$ and $q^{(\varepsilon)}$ be its intensity measure. It follows from Proposition 2.2 that $q^{(\varepsilon)}(dt, dz) = dt \times f^{(\varepsilon)}(z) dz$, where $f^{(\varepsilon)}(z) = f(z) \mathbf{1}_{\{|z|>\varepsilon\}}$. Furthermore, let $r^{(\varepsilon)} = p^{(\varepsilon)} - q^{(\varepsilon)}$. Noticing the above decomposition, we can rewrite the SDE as follows:

$$dX_t = \tilde{a}^{(\varepsilon)}(X_t) dt + b(X_t) dw_t + dB_t^{(\varepsilon)} + \int_{\mathcal{E}} c(X_{t-}, z) p^{(\varepsilon)}(dt, dz), \quad (2.9)$$

where

$$\begin{aligned} \tilde{a}^{(\varepsilon)}(x) &= a(x) - \int_{\mathcal{E}} c(x, z) f^{(\varepsilon)}(z) dz, \\ B_t^{(\varepsilon)} &= \int_0^t \int_{0<|z|\leq\varepsilon} c(x, z) r(ds, dz). \end{aligned}$$

The last term in the decomposed SDE (2.9) corresponds to the jump part driven by a compound Poisson process $z^{(\varepsilon)}$. If the underlying z is a compound Poisson process,

then $B^{(\varepsilon)}$ vanishes by taking $\varepsilon = 0$, and we call such a model the *Poisson type* or the *finite activity model*. Otherwise $B^{(\varepsilon)}$ corresponds to the jump part driven by a Lévy process with infinitely many small jumps, and we call such a model the *Lévy type* or the *infinite activity model*.

On the latter condition of (2.7), we set $\sigma^2(\varepsilon) := \int_{0 < |z| \leq \varepsilon} |z|^2 f(z) dz$. This corresponds to the dispersion of the small jump part $B^{(\varepsilon)}$. The following proposition implies that the process $B^{(\varepsilon)}$ can be regarded as a Brownian noise with small covariance if $\varepsilon > 0$ is *small*; when the dispersion $\sigma(\varepsilon)$ converges to zero more slowly than the level of truncation.

Proposition 2.4 *Suppose that $\varepsilon^{-1}\sigma(\varepsilon) \rightarrow \infty$ as $\varepsilon \rightarrow 0$. Then*

$$\sigma(\varepsilon)^{-1}B^{(\varepsilon)} \xrightarrow{d} B$$

in $D[0, 1]$ space equipped with the uniform metric, where B is a d -dimensional standard Brownian motion.

See Asmussen and Rosinski [4] for the proof. On the other hand, if $\sigma(\varepsilon)$ converges to zero so fast then we can regard $B^{(\varepsilon)}$ as the negligible noise. Consequently, it indicates that an infinite activity model can be approximated by a finite activity model. We stand on such a viewpoint in Chapter 4.

2.3 Ito's formula and a differential operator

When we discuss various problems for stochastic differential equations (with jumps), one of the most important tools is Ito's formula, which is well known in the stochastic calculus. It provides an integral-differential calculus for sample paths of semimartingales. Ito's formula says that a "smooth" function of a semimartingale is a semimartingale again, and provides its decomposition.

In this section, we introduce a version of Ito's formula for diffusion processes with jumps of the form (2.2), and give some auxiliary results to be used repeatedly in Chapter 3 and 4.

We denote by L an integro-differential operator for a C^2 -class function g of the form

$$\begin{aligned} Lg(x) &= \partial_x^* g(x) a(x) + \frac{1}{2} \text{tr} [\partial_x^2 g(x) b(x) b^*(x)] \\ &\quad + \int_{\mathcal{E}} \{g(x + c(x, z)) - g(x) - \partial_x^* g(x) c(x, z)\} f(z) dz, \end{aligned}$$

This operator is known for the *infinitesimal generator* of X as a Markov process, and it plays an important role in the stochastic calculus although we do not describe the details about their properties here; see e.g. Ethier and Kurtz [24]. In this thesis, this is used as just an operator to make the notation be simple. Using the operator L , Ito's formula for (2.2) is provided as follows.

Theorem 2.2 (Ito's formula) *Let X be a solution-process to the stochastic differential equation (2.2) and let g be a C^2 -class function. Assume that $\int_{\mathcal{E}} c(x, z)f(z) dz < \infty$ for any $x \in \mathbb{R}^d$. Then $g(X)$ is a semimartingale with jumps that follows the following stochastic differential equation:*

$$g(X_t) - g(x) - \int_0^t Lg(X_s) ds = \int_0^t \partial_x^* g(X_s) b(X_s) dw_s + \int_0^t \int_{\mathcal{E}} \mathcal{D}g(X_{s-}, z) r(ds, dz),$$

where $\mathcal{D}g(x, z) = g(x + c(x, z)) - g(x)$.

The following results are useful throughout this thesis. The one is a version of Fubini's theorem for conditional distributions on a σ -field, and the other is the Ito-Taylor expansion for semimartingales.

Proposition 2.5 *Let Y be an \mathcal{F}_t -adapted càdlàg process, and assume that*

$$E \left[\sup_{0 \leq s \leq t} |Y_s| \right] < \infty \quad (2.10)$$

for each $t \geq 0$. Then $E[Y_t | \mathcal{F}_s]$ ($s \leq t$) is $\mathcal{F}_s \otimes \mathcal{B}(\mathbb{R}_+)$ -measurable and

$$E \left[\int_a^b Y_t dt \middle| \mathcal{F}_s \right] = \int_a^b E[Y_t | \mathcal{F}_s] dt \quad a.s.$$

for any $a, b \in \mathbb{R}_+$.

Proof . For $(\omega, t) \in \Omega \times \mathbb{R}_+$, $n \in \mathbb{N}$ and $s \in [0, t]$, we define

$$F_n(\omega, t) := \sum_{k=0}^{2^n-1} E \left[Y_{\frac{k+1}{2^n}t} \middle| \mathcal{F}_s \right] \mathbf{1}_{(\frac{kt}{2^n}, \frac{k+1}{2^n}t]}(t).$$

Notice that $\lim_{n \rightarrow \infty} \sum_{k=0}^{2^n-1} Y_{\frac{k+1}{2^n}t} \mathbf{1}_{(\frac{kt}{2^n}, \frac{k+1}{2^n}t]}(t) = Y_t$ for almost all $(\omega, t) \in \Omega \times \mathbb{R}_+$ since Y is càdlàg. Moreover notice that F_n is \mathcal{F}_s -measurable for each $t \geq 0$. Therefore it follows for any $A \in \mathcal{F}_s$ that

$$E \left[\lim_{n \rightarrow \infty} F_n(\omega, t) \mathbf{1}_A(\omega) \right] = \lim_{n \rightarrow \infty} E \left[E \left[\sum_{k=0}^{2^n-1} Y_{\frac{k+1}{2^n}t} \mathbf{1}_{(\frac{kt}{2^n}, \frac{k+1}{2^n}t]}(t) \mathbf{1}_A(\omega) \middle| \mathcal{F}_s \right] \right] \quad (2.11)$$

$$= E \left[\lim_{n \rightarrow \infty} \sum_{k=0}^{2^n-1} Y_{\frac{(k+1)t}{2^n}} \mathbf{1}_{(\frac{kt}{2^n}, \frac{k+1}{2^n}t]}(t) \mathbf{1}_A(\omega) \right] \quad (2.12)$$

$$= E[Y_t \mathbf{1}_A]. \quad (2.13)$$

We used Lebesgue's convergence theorem in the equalities (2.11) and (2.12) by noticing the assumption (2.10) and that $\left| \sum_{k=0}^{2^n-1} Y_{\frac{(k+1)t}{2^n}} \mathbf{1}_{(\frac{kt}{2^n}, \frac{k+1}{2^n}t]}(t) \mathbf{1}_A(\omega) \right| \leq \sup_{0 \leq s \leq t} |Y_s|$ for almost all $(\omega, t) \in \Omega \times \mathbb{R}_+$. The equality (2.13) implies that $\lim_{n \rightarrow \infty} F_n(\omega, t) = E[Y_t | \mathcal{F}_s]$. Hence $E[Y_t | \mathcal{F}_s]$ is $\mathcal{F}_s \otimes \mathcal{B}(\mathbb{R}_+)$ -measurable since $F_n(\omega, t)$ is $\mathcal{F}_s \otimes \mathcal{B}(\mathbb{R}_+)$ -measurable.

Moreover, since $E[|Y_t| | \mathcal{F}_s]$ is nonnegative and also $\mathcal{F}_s \otimes \mathcal{B}(\mathbb{R}_+)$ -measurable by the same argument as above, we see from (2.10) and Fubini's theorem that

$$0 \leq \iint_{[a,b] \times \Omega} E[|Y_t| | \mathcal{F}_s] dt \times dP(\omega) \leq \int_a^b E[|Y_t|] dt < \infty.$$

Therefore $E[Y_t | \mathcal{F}_s]$ is integrable on $[a, b] \times \Omega$, and by Fubini's theorem again, we see that $\int_a^b E[Y_t | \mathcal{F}_s] dt$ is \mathcal{F}_s -measurable and

$$E \left[\left(\int_a^b E[Y_t | \mathcal{F}_s] dt \right) \mathbf{1}_A \right] = \int_a^b E[Y_t \mathbf{1}_A] dt = E \left[\left(\int_a^b Y_t dt \right) \mathbf{1}_A \right]$$

for any $A \in \mathcal{F}_s$. This implies the consequence. \square

Proposition 2.6 *Let g be a $C^{2(l+1)}$ -class function whose derivatives up to $2(l+1)$ th are of polynomial growth. Assume that the coefficient $a(x), b(x)$ and $c(x, z)$ in (2.2) are C^{2l} -class functions whose derivatives with respect to x up to $2l$ th are of polynomial growth. Furthermore, assume that $\sup_t E|X_t|^p < \infty$ for $p > 0$ large enough. Then the following expansion holds for $t > s$ and $\Delta = t - s$:*

$$E[g(X_t) | \mathcal{F}_s] = \sum_{j=0}^l \Delta^j \frac{L^j}{j!} g(X_s) + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_l} E[L^{l+1} g(X_{s+u_{l+1}}) | \mathcal{F}_s] du_1 \cdots du_{l+1}.$$

Proof . Note that all the stochastic integrals with respect to the Wiener process w and the random martingale measure r which appear in the sequel become martingales from the assumptions: see Protter [82] for details.

It follows by Ito's formula, Proposition 2.5 and a martingale property that

$$\begin{aligned} E[g(X_t) | \mathcal{F}_s] &= g(X_s) + \int_s^t E[Lg(X_v) | \mathcal{F}_s] dv \\ &= g(X_s) + \int_0^\Delta E[Lg(X_{v+s}) | \mathcal{F}_s] dv. \end{aligned}$$

Applying Ito's formula to $Lg(X_{v+s})$, we have

$$E[g(X_t)|\mathcal{F}_s] = g(X_s) + \Delta Lg(X_s) + \int_0^\Delta \int_0^v E[L^2g(X_{u_1+s})] du_1 dv.$$

Applying Ito's formula again to $L^2g(X_{u_1+s})$, we have

$$\begin{aligned} E[g(X_t)|\mathcal{F}_s] &= g(X_s) + \Delta Lg(X_s) + \int_0^\Delta \int_0^v L^2g(X_s) du_1 dv \\ &\quad + \int_0^\Delta \int_0^v \int_0^{u_1} E[L^3g(X_{u_2+s})] du_2 du_1 dv \\ &= g(X_s) + \Delta Lg(X_s) + \frac{\Delta^2}{2} L^2g(X_s) \\ &\quad + \int_0^\Delta \int_0^v \int_0^{u_1} E[L^3g(X_{u_2+s})] du_2 du_1 dv. \end{aligned}$$

In this way, we can obtain the consequence by the induction. \square

2.4 Ergodic diffusion processes with jumps

In the statistical inference, the assumption of an *ergodicity* of the processes is sometimes important to investigate the asymptotic behavior of estimators. Actually we assume an ergodicity of X to (2.2) in Chapter 3 and 4.

General ergodic theorems for Markov processes are usually described as the combination of two kinds of theorems. One is on the existence of the limit in probability of the time-mean:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T f(X_t) dt, \quad (2.14)$$

where X is a stochastic process and f is a measurable function. The other is on the existence of the limit for a transition probability $p(t, x, A)$ of a Markov process or its time-mean:

$$\lim_{T \rightarrow \infty} p(T, x, A) \quad \text{or} \quad \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T p(t, x, A) dt, \quad (2.15)$$

and the cases where these limits do not depend on the initial state X_0 are of major interest.

In this thesis, we often assume the former ergodicity (2.14) of diffusion processes with jumps. This is so-called the weak law of large numbers for a stochastic process X .

Let X be a solution-process to (2.2) and $p(t, x, A)$ be a transition probability defined by

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

where A is a Borel subset on \mathbb{R}^d and $t, s \geq 0$. Since X is a Markov process, the above $p(t, x, A)$ is independent of s because of the Markov property of X , and the distribution of X is uniquely determined by p . In particular, if π be a probability distribution of the initial value X_0 then $\pi T_t := \int p(t, x, A) \pi(dx)$ is a probability distribution of X_t for each $t > 0$.

The distribution π is said to be *invariant* if and only if the equality $\pi T_t = \pi$ holds for all $t \in \mathbb{R}_+$. The existence of the invariant measure for a Markov process is essential in the ergodic theory since the limit (2.14) is written by a kind of integral by the invariant measure. Moreover, note that if the initial distribution π is invariant then all of the distributions of X_t for each $t > 0$ is also π , that is, X is *stationary*. Hereafter we use the word “stationarity” in this sense. Such π is also called the *stationary distribution*.

Now let us give the definition of the ergodicity in this thesis of a solution-process to (2.2).

Definition 2.3 *Let X be a solution-process to (2.2). The process X is ergodic if and only if there exists an invariant measure π such that*

$$\frac{1}{T} \int_0^T f(X_t) dt \xrightarrow{p} \int_{\mathbb{R}^d} f(x) \pi(dx) \quad (T \rightarrow \infty)$$

for any π -integrable function f defined on \mathbb{R}^d .

This might not be a general definition of an ergodicity but the one we require as an assumption in the later chapters. However it is generally difficult to check this ergodicity for general diffusion processes with jumps. Nevertheless we can find some sufficient conditions of the ergodicity for some special diffusion processes with jumps.

Meyn and Tweedie [71, 72] gave a general ergodic theory for general Markov processes, and one could find some ergodic jump-diffusions in above sense by investigating their papers carefully. Applying their theory to jump-diffusions (2.2) in which the coefficient of the jump part has the form $c(x, z) = \zeta(x)z$ for a function ζ , Masuda [67] provided more explicit conditions for the ergodicity. For example, the irreducibility, Foster-Lyapunov criteria, the stationarity and some moment conditions with respect to the invariant measure π yield the exponential ergodicity which is stronger than the one in our sense. Moreover, Masuda [66] says that Lévy driven Ornstein-Uhlenbeck

processes can be ergodic under some mild conditions. Otherwise, some stability conditions, that is, the irreducibility and Feller property, yield the positive Harris recurrency, and they deduce $1/T \int_0^T f(X_t) dt \rightarrow \int f(x) d\pi$ P -almost surely. One can find this result by the combination of the results in Kwon and Lee [56] and Meyn and Tweedie [71]. Considering the above facts, we shall give some examples of ergodic diffusions with jumps.

Example 2.1 Consider multidimensional Lévy driven Ornstein-Uhlenbeck processes:

$$dX_t = -\theta X_t dt + dz_t^\alpha, \quad (2.16)$$

where θ and α are also multidimensional parameters, z_t^α is a Lévy process with a Lévy density f_θ satisfying $\int_{\mathcal{E}} |z|^q f_\theta(z) dz < \infty$ for some $q > 0$. This is one of the most important models in applications. In this model, it is known that there exists the unique invariant measure π such that $\int |x|^q d\pi(x) < \infty$, and X is exponentially ergodic if π is the initial distribution. See Masuda [66] for details.

Example 2.2 Consider the following 1-dimensional SDE's with a multidimensional parameter θ , and $\sigma_1, \sigma_2 > 0$:

$$dX_t = b(X_t, \theta) dt + \left[\sigma_1 (1 + X_t^2)^{-1/2} + \sigma_2 \right] dw_t + dz_t^\theta, \quad (2.17)$$

If z^θ is a compound Poisson process with $\int_{\mathcal{E}} z^2 f_\theta(z) dz < \infty$ then X is irreducible, and if the Foster-Lyapunov criterion: $2xb(x) \leq -\kappa x^2$ for some $\kappa > 0$ is satisfied for sufficiently large $|x|$, then X is exponentially ergodic.

Example 2.3 Consider the following SDE's with one-dimensional parameters:

$$dX_t = \theta_1 X_t dt + \sigma dw_t + \theta_2 X_t dz_t^{\theta_3}, \quad (2.18)$$

where $\sigma > 0$ and z^{θ_3} is a compound Poisson process with $\int_{\mathcal{E}} z^2 f_{\theta_3}(z) dz < \infty$. X is ergodic if the Foster-Lyapunov criterion: $2\theta_1 + \theta_2^2 \int_{\mathcal{E}} z^2 f_{\theta_3}(z) dz < 0$ is satisfied.

Chapter 3

Parametric estimation in finite activity models

This chapter is devoted to the parametric inference for finite activity jump-diffusion models. We construct a single contrast function to estimate parameters in the drift, the diffusion and the jump part jointly. The contrast function has two parts: one is the log-likelihood of a local Gaussian process, which is the direct discretization of the log-likelihood of an usual diffusion process and corresponds to the contrast function for parameters in continuous part of jump-diffusions. The other is the contrast function modeled after the log-likelihood of Poisson random measures. The key idea is the jump-judging procedure, which is described in Section 3.2.2, and this idea would be also the key in whole of this thesis.

3.1 Setting of the model

Let us consider a d -dimensional solution process X to the following stochastic differential equation with jumps on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$:

$$\begin{cases} dX_t = a(X_t, \theta) dt + b(X_t, \sigma) dw_t + \int_{\mathcal{E}} c(X_{t-}, z, \theta) (p - q_\theta)(dt, dz), \\ X_0 = x_0, \end{cases} \quad (3.1)$$

where $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, $\theta \in \Theta \subset \mathbb{R}^{m_1}$, $\sigma \in \Pi \subset \mathbb{R}^{m_2}$ are parameters, and $\alpha = (\theta, \sigma)$ belongs to a parameter space $\Xi = \Theta \times \Pi$ which is a compact convex subset of $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$. Let $m = m_1 + m_2$. $(w_t)_{t \geq 0}$ is an r -dimensional Wiener process, $p(dt, dz)$ is a time-homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^d$, and $q_\theta(dt, dz)$ is its intensity measure, that is, $E[p(dt, dz)] = q_\theta(dt, dz)$. We set $q_\theta(dt, dz) = f_\theta(z) dz dt$ and $f_\theta(z) =$

$\lambda(\theta)F_\theta(z)$, where $\lambda(\theta)$ is a strictly positive and *bounded* function of θ and F_θ is a probability density. The coefficients a and c are known \mathbb{R}^d -valued Borel functions defined on $\mathbb{R}^d \times \Theta$, $\mathbb{R}^d \times \mathcal{E} \times \Theta$ respectively, and b is a known $\mathbb{R}^d \otimes \mathbb{R}^r$ -valued Borel function defined on $\mathbb{R}^d \times \Pi$.

One of the most simple but important examples of one-dimensional case is written as follows:

$$dX_t = \tilde{a}(X_t, \mu) dt + \tilde{b}(X_t, \sigma) dw_t + \tilde{c}(X_{t-}, \vartheta) dz_t^\vartheta, \quad (3.2)$$

where z^ϑ is a Lévy process with parameter ϑ . This belongs to the class of (3.1) with $c(x, z, \theta) = \tilde{c}(x, \vartheta)z$, $a(x, \theta) = \tilde{a}(x, \mu) + \int_{\mathcal{E}} c(x, \vartheta)z f_\vartheta(z) dz$ and $\theta = (\mu, \vartheta)$.

We study estimation of the parameter $\alpha = (\theta, \sigma)$ from discrete observations. For that purpose, we observe $n + 1$ data $\{X_{t_i^n}\}_{i=0}^n$, $t_i^n = ih_n$, and show the consistency and the asymptotic normality of an estimator under the rapidly experimental design such that $h_n \rightarrow 0$, $nh_n \rightarrow \infty$, $nh_n^2 \rightarrow 0$.

3.2 Discussion and conclusions

3.2.1 Assumptions and examples

We make the following assumptions.

A 1 *There exists a constant $L > 0$ and a function $\zeta(z)$ which satisfies $|\zeta(z)| \leq C(1 + |z|)^C$ for a constant $C > 0$ such that*

$$\begin{aligned} |a(x, \theta_0) - a(y, \theta_0)| + |b(x, \sigma_0) - b(y, \sigma_0)| &\leq L|x - y|, \\ |c(x, z, \theta_0) - c(y, z, \theta_0)| &\leq \zeta(z)|x - y|, \quad |c(x, z, \theta_0)| \leq \zeta(z)(1 + |x|). \end{aligned}$$

A 2 *The process X is ergodic and stationary for $\alpha = \alpha_0$ with an invariant measure π in the sense of Section 2.4.*

A 3 *For every $p \geq 1$,*

$$\sup_{t \geq 0} E[|X_t|^p] < \infty.$$

A 4 *For fixed θ and σ , the derivatives $\partial_x^l a(x, \theta)$ and $\partial_x^l b(x, \sigma)$ ($l = 1, 2$) exist on \mathbb{R}^d and they are continuous in x . Moreover, for fixed x , the derivatives $\partial_\theta^l a(x, \theta)$ and $\partial_\sigma^l b(x, \sigma)$ ($l = 1, 2$) exist on Θ and Π respectively, and a , b , and their all derivatives are of polynomial growth in x uniformly in α : for $l = 0, 1, 2$,*

$$|\partial_x^l a(x, \theta)|, |\partial_x^l b(x, \sigma)|, |\partial_\theta^l a(x, \theta)|, |\partial_\sigma^l b(x, \sigma)| \leq C(1 + |x|)^C.$$

A 5 $\inf_x |c(x, z, \theta_0)| \geq c_0 |z|$ for some $c_0 > 0$.

A 6 There exist constants $r, K > 0$ and $\gamma > 3$ such that $f_{\theta_0}(z) \mathbf{1}_{\{|z| \leq r\}} \leq K |z|^\gamma$, and that

$$\sup_{\theta \in \Theta} \int_{\mathcal{E}} |z|^p f_{\theta}(z) dz < \infty$$

for all $p \geq 1$.

A 7 For each θ and x , the mapping $z \mapsto y = c(x, z, \theta)$ has an inverse $z = c^{-1}(x, y, \theta)$ which is differentiable with respect to y , and we set

$$\Psi_{\theta}(y, x) := f_{\theta}(c^{-1}(x, y, \theta)) J(x, y, \theta),$$

where $J(x, y, \theta)$ is the absolute value of the Jacobian of $c^{-1}(x, y, \theta)$.

A 8 The matrix $\beta(x, \sigma) := b(x, \sigma) b^*(x, \sigma)$ is a positive definite and $\inf_{x, \sigma} \det \beta(x, \sigma) > 0$.

A 9 The function $\Psi_{\theta}(y, x)$ is differentiable with respect to x and y , and three times continuously differentiable with respect to θ . Moreover we assume that

$$|\partial_{\theta}^k \Psi_{\theta}(y, x)| \leq L_1(y) (1 + |x|)^C \quad (k = 0, 1, 2, 3), \quad (3.3)$$

$$|\partial_x \partial_{\theta}^l \Psi_{\theta}(y, x)| \leq L_2(y) \quad (l = 0, 1, 2), \quad (3.4)$$

where L_1 and L_2 are bounded and dy -integrable functions. Furthermore,

$$|\partial_y \partial_{\theta}^l \Psi_{\theta}(y, x)| \leq C (1 + |y|)^C (1 + |x|)^C \quad (l = 0, 1, 2), \quad (3.5)$$

$$\int \sup_{\theta} |\partial_{\theta}^k \log \Psi_{\theta}(y, x) \cdot \Psi_{\theta_0}(y, x)| dy \leq C (1 + |x|)^C \quad (k = 0, 1, 2, 3). \quad (3.6)$$

A 10 The following identifiability condition holds: $\sigma = \sigma_0$ if and only if $\det \beta(x, \sigma) = \det \beta(x, \sigma_0)$ for almost all x . Moreover, $\theta = \theta_0$ if and only if $a(x, \theta) = a(x, \theta_0)$ and $\Psi_{\theta}(y, x) = \Psi_{\theta_0}(y, x)$ for almost all x and y .

A 11 A sequence of real valued functions $\{\varphi_n(x, y)\}_{n \in \mathbb{N}}$ satisfies the following properties: $0 \leq \varphi_n \leq 1$ and $\varphi_n \rightarrow 1$ $dy \times d\pi$ -a.s. as $n \rightarrow \infty$. There exist some $M > 0$ such that

$$\varphi_n(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \left\{ \inf_{\theta \in \Theta} \Psi_{\theta}(y, x) > M \right\} \\ 0 & \text{if } (x, y) \in \left\{ \inf_{\theta \in \Theta} \Psi_{\theta}(y, x) \leq \varepsilon_n \right\} \end{cases},$$

where $\varepsilon_n = b_n^{-1/10}$ for a sequence b_n satisfying $b_n \rightarrow \infty$, $nh_n^2 b_n \rightarrow 0$ and $\frac{b_n}{nh_n} \rightarrow 0$. Moreover,

$$\sup_{x,y} |\partial_x \varphi_n| + \sup_{x,y} |\partial_y \varphi_n| = O(\varepsilon_n^{-1}), \quad (3.7)$$

and

$$\partial_x \varphi_n = \partial_y \varphi_n = 0 \quad \text{on} \quad \left\{ \inf_{\theta \in \Theta} \Psi_\theta(y, x) \leq \varepsilon_n \right\}. \quad (3.8)$$

Let us give some examples of $\Psi_\theta(y, x)$ and check conditions in A9.

Example 3.1 We consider the following one-dimensional stochastic differential equation.

$$dX_t = a(X_t) dt + b(X_t) dw_t + c(X_{t-}, \theta) dz_t^\theta, \quad (3.9)$$

where z^θ is a Lévy process with Lévy density $f_\theta(z)$. We assume, for some $M, K > 0$,

$$c(x, \theta) = \frac{1}{c}, \quad f_\theta(z) = M z^\alpha (K - z)^\beta \mathbf{1}_{\{0 \leq z \leq K\}},$$

and $\theta = (c, \alpha, \beta)$, where $\alpha, \beta > 3$.

$$\begin{aligned} \Psi_\theta(y, x) &= M c (cy)^\alpha (K - cy)^\beta, \\ \log \Psi_\theta(y, x) &= \log M + (\alpha + 1) \log c + \alpha \log y + \beta \log(K - cy). \end{aligned}$$

Noticing the parameter space Θ is compact, it is easy to verify (3.3) - (3.5). Moreover all $\int \sup_\theta |\partial_\theta^k \log \Psi_\theta(y, x) \cdot \Psi_{\theta_0}(y, x)| dy$ become finite, so (3.6) is satisfied.

Example 3.2 For SDE (3.9), we suppose that $\text{supp}(f_\theta) \subset \mathbb{R}_+$ and

$$c(x, \theta) = \frac{1}{c}, \quad f_\theta(z) \mathbf{1}_{\{0 < z \leq M\}} = e^{-\frac{\gamma}{z}} \mathbf{1}_{\{0 < z \leq M\}}$$

and put $\theta = (c, \gamma)$, where $c > 0$, $\gamma > 0$ and $M > 0$. It is easy to check A9 in the neighborhood of the origin since

$$\Psi_\theta(y, x) = e^{-\frac{\gamma}{cy}}, \quad \log \Psi_\theta(y, x) = -\frac{\gamma}{cy}$$

on the set $\{0 < z \leq M\}$.

Example 3.3 For SDE (3.9), we suppose that

$$c(x, \theta) = \frac{1}{c}, \quad f_\theta(z) = \frac{\alpha^\beta}{\Gamma(\beta)} z^{\beta-1} e^{-\alpha z},$$

and put $\theta = (c, \alpha, \beta)$, where $\alpha > 0$, $\beta > 4$. Then

$$\begin{aligned} \Psi_\theta(y, x) &= \frac{\alpha^\beta}{\Gamma(\beta)} (cy)^{\beta-1} e^{-\alpha cy}, \\ \log \Psi_\theta(y, x) &= \beta \log \alpha - \log \Gamma(\beta) + (\beta - 1) \log cy - \alpha cy. \end{aligned}$$

Then

$$|\partial_\theta^k \Psi_\theta(y, x)|, \quad |\partial_x \partial_\theta^l \Psi_\theta(y, x)|, \quad |\partial_y \partial_\theta^l \Psi_\theta(y, x)|$$

are all dominated by $C(1 + |\log y|) y^{\beta'} e^{-\alpha' y}$ for some $C > 0$, $\beta' > 0$, $\alpha' > 0$, so Conditions (3.3) - (3.5) are satisfied. Moreover we obtain

$$\sup_\theta |\partial_\theta^k \log \Psi_\theta(y, x)| \leq C(1 + |y| + |\log y|).$$

This implies (3.6).

Under Conditions A1 and A6, the stochastic differential equation (3.1) can be rewritten as follows:

$$dX_t = \bar{a}(X_t, \theta) dt + b(X_t, \sigma) dw_t + \int_{\mathcal{E}} c(X_{t-}, z, \theta) p(dt, dz), \quad (3.10)$$

where $\bar{a}(x, \theta) = a(x, \theta) - \int_{\mathcal{E}} c(x, z, \theta) q_\theta(dt, dz)$. This expression implies that X follows diffusion process $dX_t = \bar{a}(X_t, \theta) dt + b(X_t, \sigma) dw_t$, in the interval in which no jump occurred. We start with the stochastic differential equation (3.10) to construct the contrast function.

3.2.2 Contrast functions and efficient estimators

Now we present a contrast function for estimating parameters. In Section 2.3, we show how to obtain it.

Definition 3.1 For $\frac{2}{\gamma+1} \leq \rho < \frac{1}{2}$, we define the contrast function $l_n(\alpha)$ as follows .

$$l_n(\alpha) = \bar{l}_n(\theta, \sigma) + \tilde{l}_n(\theta),$$

where

$$\begin{aligned}
\bar{l}_n(\alpha) &= -\frac{1}{2h_n} \sum_{i=1}^n (\bar{X}_{i,n})^*(\theta) \beta_{i-1}^{-1}(\sigma) \bar{X}_{i,n}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \\
&\quad - \sum_{i=1}^n \frac{1}{2} \log \det \beta_{i-1}(\sigma) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\
\tilde{l}_n(\theta) &= \sum_{i=1}^n \left\{ \log \Phi_n(\theta, X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\
&\quad - h_n \sum_{i=1}^n \int \Phi_n(\theta, X_{t_{i-1}^n}, y) dy,
\end{aligned}$$

$$\bar{X}_{i,n}(\theta) = X_{t_i^n} - X_{t_{i-1}^n} - h_n \bar{a}_{i-1}(\theta), \quad \Phi_n(\theta, x, y) = \Psi_\theta(y, x) \varphi_n(x, y).$$

Intuitively speaking, this contrast function is very natural since $\bar{l}_n(\alpha)$ corresponds to the contrast for an usual diffusion process, and $\tilde{l}_n(\theta)$ does to the discretization of the likelihood function of an compound Poisson process with Lévy density f_θ .

Our main theorem is the following. The proof will be presented in Section 3.5

Theorem 3.1 *Under Conditions A1 to A11 and $nh_n \rightarrow \infty$ as $n \rightarrow \infty$, the estimator $\hat{\alpha}_n$ which satisfies*

$$l_n(\hat{\alpha}_n) = \sup_{\alpha \in \Xi} l_n(\alpha)$$

is consistent:

$$\hat{\alpha}_n \xrightarrow{P} \alpha_0 \quad (n \rightarrow \infty).$$

If, in addition, $nh_n^2 \rightarrow 0$ and the true value α_0 is in the interior of Ξ , then

$$\left(\sqrt{nh_n}(\hat{\theta}_n - \theta_0), \sqrt{n}(\hat{\sigma}_n - \sigma_0) \right) \xrightarrow{d} \mathcal{N}_m(0, K^{-1}),$$

where

$$K := \begin{pmatrix} K_1 & \mathbf{0} \\ \mathbf{0} & K_2 \end{pmatrix},$$

$$\begin{aligned}
K_1^{(p,q)} &= \int (\partial_{\theta_p} \bar{a})^* \beta^{-1}(\partial_{\theta_{p'}} \bar{a})(x, \alpha_0) d\pi + \iint \frac{\partial_{\theta_p} \Psi_{\theta_0} \partial_{\theta_{p'}} \Psi_{\theta_0}}{\Psi_{\theta_0}}(y, x) dy d\pi, \\
K_2^{(p,q)} &= \frac{1}{2} \int \text{tr} \left[(\partial_{\sigma_q} \beta) \beta^{-1} (\partial_{\sigma_{q'}} \beta) \beta^{-1} \right] (x, \sigma_0) d\pi.
\end{aligned}$$

Remark 3.1 We impose Assumption A6 in order to show the asymptotic results under the asymptotics $nh_n^2 \rightarrow 0$. However we can relax A6 to, for example, “ f_{θ_0} is bounded” if we impose a more rapid experimental design as $nh_n^{1+\delta} \rightarrow 0$ for a $\delta \in (0, 1)$. The setting in Chapter 5 is such a case.

Remark 3.2 We use a truncation function φ_n to ensure the P -integrability of the logarithm term and its derivatives with respect to parameters. If we knew

$$\sup_{i,n} E \left[\partial_y \partial_\theta^l \log \Psi_\theta(\Delta_i X^n, X_{t_{i-1}^n}) \right] < \infty \quad (l = 0, 1, 2)$$

by some reasons, φ_n would not be needed, however φ_n is needed generally.

Remark 3.3 This result can be applied to pure jump type processes with $b(x, \sigma) \equiv 0$, that is, X is a solution process to the following stochastic differential equation:

$$dX_t = a(X_t, \theta) dt + \int_E c(X_{t-}, z, \theta) (p - q_\theta)(dz, dt).$$

The contrast function of jump part is similar to the non-degenerate case since we estimate jump parameters from only the number of jumps and their amplitudes. For diffusion part, however, we can not make use of $\bar{l}_n(\alpha)$ any more because we can not approximate the path of X by the local Gaussian approximation in the no jump intervals. We can overcome this difficulty by estimating drift parameters as least square estimators, that is,

$$\bar{l}_n(\theta) = -\frac{1}{2h_n} \sum_{i=1}^n (\bar{X}_{i,n})^*(\theta) \bar{X}_{i,n}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}},$$

In this case, $\hat{\theta}_n$ has also consistency and asymptotic normality with asymptotic variance K , $K^{(p,q)} = \iint \frac{\partial_{\theta_p} \Psi_{\theta_0} \partial_{\theta_{p'}} \Psi_{\theta_0}}{\Psi_{\theta_0}}(y, x) dy d\pi$. The proof is the same as for the non-degenerate diffusion case.

Remark 3.4 The asymptotic efficiency for $\hat{\theta}_n$ is obtained since K_1 is the asymptotic variance of the estimator for the continuously observed ergodic diffusion processes with jumps; see Sørensen [99], which discusses the inference for diffusion processes with jumps from continuous observations under the setting which includes the non-ergodic case. Particularly, when you compute the asymptotic variance in the ergodic case, it could be clearer to refer Section 3 in Barndorff-Nielsen and Sørensen [11], which is a review of the general likelihood theory.

3.2.3 Construction of contrast functions

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 $|\Delta_i X^n|$, 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. The way we will take is the following: for $\rho \in [0, 1/2)$, if the increment exceeds h_n^ρ in an interval then we regard it as the interval in which a jump has occurred and if not, as the interval in which no jump occurred. This is because the increment of a diffusion without jumps exceeds h_n^ρ with small probability, and the increment of a diffusion with a single jump also exceeds h_n^ρ with a large probability. Although they are intuitive argument, these are justified by Lemma 3.2 described below.

The value ρ has to be chosen carefully. For instance if ρ is too large, and therefore h_n^ρ is too small, the probability of getting the increment h_n^ρ by the continuous diffusion can not be ignored, on the other hand, if ρ is too small, and therefore h_n^ρ is too large, we cannot ignore the probability of getting an increment less than h_n^ρ when a jump occurs in an interval. Later we will choose ρ as $2/(\gamma + 1) \leq \rho < 1/2$.

First, we state the well-known Gronwall's inequality. This lemma is often used in this theses.

Lemma 3.1 (Gronwall's inequality) *Let $f(t)$ and $g(t)$ be continuous functions such that $g(t) \geq 0$. For a function $u(t)$, suppose that*

$$u(t) \leq f(t) + \int_c^t g(s)u(s) ds,$$

for $t > c$. Then the following inequality is valid: for $t > c$,

$$u(t) \leq f(t) + \int_c^t g(s)f(s) \exp\left(\int_s^t g(v) dv\right) ds.$$

The following result and its idea is the key throughout this thesis.

Lemma 3.2 *Define random times τ_i^n and η_i^n as*

$$\begin{aligned} \tau_i^n &:= \inf \{t \in [t_{i-1}^n, t_i^n); |\Delta X_t| > 0\}, \\ \eta_i^n &:= \sup \{t \in [t_{i-1}^n, t_i^n); |\Delta X_t| > 0\}. \end{aligned}$$

If the infimum or supremum on the right hand side does not exist, we define the random times as t_i^n . Assume Conditions A1, A3 and A6. Then, for any $\rho \in [0, 1/2)$ and any

$p \geq 1$,

$$P_{i-1}^n \left\{ \sup_{t \in [t_{i-1}^n, \tau_i^n)} |X_t - X_{t_{i-1}^n}| > h_n^\rho \right\} = R(\alpha, h_n^p, X_{t_{i-1}^n}), \quad (3.11)$$

$$P_{i-1}^n \left\{ \sup_{t \in [\eta_i^n, t_i^n)} |X_{t_i^n} - X_t| > h_n^\rho \right\} = R(\alpha, h_n^p, X_{t_{i-1}^n}), \quad (3.12)$$

where R is given in Section 1.3, and each function R does not depend on i .

Proof . First , we show (3.11) . On the interval $[t_{i-1}^n, \tau_i^n)$, X follows the stochastic differential equation

$$dX_t = \bar{a}(X_t) dt + b(X_t) dw_t,$$

hence for $t \in [t_{i-1}^n, \tau_i^n)$,

$$\begin{aligned} |X_t - X_{t_{i-1}^n}| &= \left| (t - t_{i-1}^n) \bar{a}(X_{t_{i-1}^n}) + \int_{t_{i-1}^n}^t (\bar{a}(X_s) - \bar{a}(X_{t_{i-1}^n})) ds + \int_{t_{i-1}^n}^t b(X_s) dw_s \right| \\ &\leq h_n |\bar{a}(X_{t_{i-1}^n})| + \sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| + L \int_{t_{i-1}^n}^t |X_s - X_{t_{i-1}^n}| ds. \end{aligned}$$

Gronwall's inequality yields that

$$\begin{aligned} |X_t - X_{t_{i-1}^n}| &\leq h_n |\bar{a}(X_{t_{i-1}^n})| + \sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| \\ &\quad + Le^{Lh_n} h_n \left(h_n |\bar{a}(X_{t_{i-1}^n})| + \sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| \right). \end{aligned}$$

If n is sufficiently large , we obtain that

$$|X_t - X_{t_{i-1}^n}| \leq C \left(h_n |\bar{a}(X_{t_{i-1}^n})| + \sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| \right), \quad (3.13)$$

therefore Markov's inequality and Burkholder-Davis-Gundy's inequality yield

$$\begin{aligned} &P_{i-1}^n \left\{ \sup_{t \in [t_{i-1}^n, \tau_i^n)} |X_t - X_{t_{i-1}^n}| > h_n^\rho \right\} \\ &\leq P_{i-1}^n \left\{ Ch_n |\bar{a}(X_{t_{i-1}^n})| > \frac{h_n^\rho}{2} \right\} + P_{i-1}^n \left\{ C \sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| > \frac{h_n^\rho}{2} \right\} \\ &\leq C_p \left\{ h_n^{p(1-\rho)} E_{i-1}^n \left[|\bar{a}(X_{t_{i-1}^n})|^p \right] + h_n^{-2p\rho} E_{i-1}^n \left[\sup_{u \in [t_{i-1}^n, t_i^n)} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right|^{2p} \right] \right\} \end{aligned}$$

$$\begin{aligned}
&\leq R(\alpha, h_n^{p(1-\rho)}, X_{t_{i-1}^n}) + C_p h_n^{-2p\rho} E_{i-1}^n \left[\left| \int_{t_{i-1}^n}^{t_i^n} b^2(X_s) ds \right|^p \right] \\
&= R(\alpha, h_n^{p(1-2\rho)}, X_{t_{i-1}^n}).
\end{aligned}$$

We used Lemma 3.5 (3.17) in the last equality. We should notice that one can take p arbitrary larger here.

The almost same argument holds for (3.12). Actually, for $t \in [\eta_i^n, t_i^n)$,

$$|X_{t_i^n} - X_t| = \left| (t_i^n - t) \bar{a}(X_t) + \int_t^{t_i^n} (\bar{a}(X_s) - \bar{a}(X_t)) ds + \int_t^{t_i^n} b(X_s) dw_s \right|.$$

By the same argument as (3.11)

$$|X_{t_i^n} - X_t| \leq C \left(h_n |\bar{a}(X_{t_i^n})| + \sup_{u \in [t_{i-1}^n, t_i^n]} \left| \int_{t_{i-1}^n}^u b(X_s) dw_s \right| \right), \quad (3.14)$$

and then,

$$\begin{aligned}
&P_{i-1}^n \left\{ \sup_{t \in [\eta_i^n, t_i^n)} |X_{t_i^n} - X_t| > h_n^\rho \right\} \\
&\leq C_p \left\{ h_n^{p(1-\rho)} E_{i-1}^n [|\bar{a}(X_{t_i^n})|^p] + h_n^{-2p\rho} E_{i-1}^n \left[\left| \int_{t_{i-1}^n}^{t_i^n} b^2(X_s) ds \right|^p \right] \right\}.
\end{aligned}$$

Lemma 3.5 (3.17) completes the proof. \square

By these facts, the probability whether the value of $|\Delta_i X^n|$ exceeds h_n^ρ or not is evaluated in the next lemma. In the following discussion, let J_i^n be the number of jumps in the interval $[t_{i-1}^n, t_i^n)$ and we set

$$\{|\Delta_i X^n| \leq Lh_n^\rho\} = \bigcup_{j=0}^2 C_{i,j}^n, \quad \{|\Delta_i X^n| > Lh_n^\rho\} = \bigcup_{j=0}^2 D_{i,j}^n,$$

where

$$\begin{aligned}
C_{i,0}^n &= \{J_i^n = 0, |\Delta_i X^n| \leq Lh_n^\rho\}, \\
C_{i,1}^n &= \{J_i^n = 1, |\Delta_i X^n| \leq Lh_n^\rho\}, \\
C_{i,2}^n &= \{J_i^n \geq 2, |\Delta_i X^n| \leq Lh_n^\rho\}, \\
D_{i,0}^n &= \{J_i^n = 0, |\Delta_i X^n| > Lh_n^\rho\}, \\
D_{i,1}^n &= \{J_i^n = 1, |\Delta_i X^n| > Lh_n^\rho\}, \\
D_{i,2}^n &= \{J_i^n \geq 2, |\Delta_i X^n| > Lh_n^\rho\}.
\end{aligned}$$

Lemma 3.3 Assume Conditions A1, A3, A5 and A6. Let $\frac{2}{\gamma+1} \leq \rho < \frac{1}{2}$, where γ is the constant given in condition A6. For any $p \geq 1$, as $n \rightarrow \infty$

$$\begin{aligned} P_{i-1}^n\{C_{i,0}^n\} &= e^{-\lambda_0 h_n} \tilde{R}(\alpha, h_n^p, X_{t_{i-1}^n}), & P_{i-1}^n\{D_{i,0}^n\} &= e^{-\lambda_0 h_n} R(\alpha, h_n^p, X_{t_{i-1}^n}), \\ P_{i-1}^n\{C_{i,1}^n\} &= R(\alpha, h_n^3, X_{t_{i-1}^n}), & P_{i-1}^n\{D_{i,1}^n\} &= \lambda_0 h_n e^{-\lambda_0 h_n} \tilde{R}(\alpha, h_n^2, X_{t_{i-1}^n}), \\ P_{i-1}^n\{C_{i,2}^n\} &\leq \lambda_0^2 h_n^2, & P_{i-1}^n\{D_{i,2}^n\} &\leq \lambda_0^2 h_n^2, \end{aligned}$$

where R and \tilde{R} are given in Section 1.3.

Proof . It is obvious that $P_{i-1}^n\{C_{i,2}^n\} \leq h_n^2 \lambda_0^2$, and so is $P_{i-1}^n\{D_{i,2}^n\}$. On $C_{i,1}^n$,

$$\begin{aligned} &P_{i-1}^n\{C_{i,1}^n\} \\ &\leq \left[P \left\{ \left| (X_{t_i^n} - X_{\tau_i^n}) + (X_{\tau_i^n} - X_{t_{i-1}^n}) + \Delta X_{\tau_i^n} \right| \leq h_n^\rho, |\Delta z_{\tau_i^n}| > \frac{2h_n^\rho}{c_0} \middle| \mathcal{F}_{i-1}^n, J_i^n = 1 \right\} \right. \\ &\quad \left. + P \left\{ |\Delta z_{\tau_i^n}| \leq \frac{2h_n^\rho}{c_0} \middle| \mathcal{F}_{i-1}^n, J_i^n = 1 \right\} \right] P\{J_i^n = 1\}, \end{aligned}$$

where $\Delta z_{\tau_i^n}$ has density F_{θ_0} . If $\left| (X_{t_i^n} - X_{\tau_i^n}) + (X_{\tau_i^n} - X_{t_{i-1}^n}) + \Delta X_{\tau_i^n} \right| \leq h_n^\rho$ then

$$|X_{t_i^n} - X_{\tau_i^n}| + |X_{\tau_i^n} - X_{t_{i-1}^n}| \geq c_0 |\Delta z_{\tau_i^n}| - h_n^\rho,$$

hence, applying Lemma 3.2, we have

$$\begin{aligned} P_{i-1}^n\{C_{i,1}^n\} &\leq \lambda_0 h_n e^{-\lambda_0 h_n} P_{i-1}^n \left\{ \sup_{t \in [t_{i-1}^n, \tau_i^n]} |X_t - X_{t_{i-1}^n}| + \sup_{t \in [\tau_i^n, t_i^n]} |X_{t_i^n} - X_t| > h_n^\rho \right\} \\ &\quad + \lambda_0 h_n e^{-\lambda_0 h_n} \int_{-2h_n^\rho/c_0}^{2h_n^\rho/c_0} M|z|^\gamma dz \\ &= R(\alpha, h_n^p, X_{t_{i-1}^n}) + C h_n^{\rho(\gamma+1)+1} \\ &= R(\alpha, h_n^3, X_{t_{i-1}^n}). \end{aligned}$$

On $C_{i,0}^n$, applying Lemma 3.2 again, we have

$$\begin{aligned} P_{i-1}^n\{C_{i,0}^n\} &= P_{i-1}^n\{J_i^n = 0\} - P_{i-1}^n\{|\Delta_i X^n| > h_n^\rho, J_i^n = 0\} \\ &= e^{-\lambda_0 h_n} - P_{i-1}^n\left\{ |X_{\tau_i^n} - X_{t_{i-1}^n}| > h_n^\rho, \tau_i^n = t_i^n \right\} \\ &= e^{-\lambda_0 h_n} \tilde{R}(\alpha, h_n^p, X_{t_{i-1}^n}). \end{aligned}$$

Finally,

$$P_{i-1}^n\{D_{i,0}^n\} = P_{i-1}^n\{J_i^n = 0\} - P_{i-1}^n\{C_{i,0}^n\}$$

$$= e^{-\lambda_0 h_n} R(\alpha, h_n^p, X_{t_{i-1}^n}),$$

and

$$\begin{aligned} P_{i-1}^n \{D_{i,1}^n\} &= P_{i-1}^n \{J_i^n = 1\} - P_{i-1}^n \{|\Delta_i X^n| \leq h_n^\rho, J_i^n = 1\} \\ &= \lambda_0 h_n e^{-\lambda_0 h_n} - R(\alpha, h_n^3, X_{t_{i-1}^n}) \\ &= \lambda_0 h_n e^{-\lambda_0 h_n} \tilde{R}(\alpha, h_n^2, X_{t_{i-1}^n}). \quad \square \end{aligned}$$

This lemma implies that we can judge the interval $[t_{i-1}^n, t_i^n]$ has no jump if $|\Delta_i X^n| \leq h_n^\rho$ and the interval has a single jump if $|\Delta_i X^n| > h_n^\rho$, and that we can ignore the events which include more than two jumps in the interval.

Remark 3.5 One can easily find in the above two lemmas that the "jump-judgement" threshold; h_n^ρ can be replaced by Lh_n^ρ for any constant $L > 0$, and the main result in this chapter; Theorem 3.1 is of course valid for such thresholds, that is, the asymptotic behavior of estimators are invariant for any constant $L > 0$. However here we consider only the trivial threshold as $L \equiv 1$ since we are interested in the asymptotic inference in this chapter. It would be a big problem in practice that how we choose the suitable L in addition to ρ according to the sample size n . Such a practical problem will be discussed in Chapter 6.

Let $(z_t)_{t \geq 0}$ be a compound Poisson process which is independent of w and has the form $z_t = \sum_{i=1}^{N_t} \varepsilon_i$, where $(N_t)_{t \geq 0}$ is a Poisson process with the intensity $\lambda(\theta)$, $(\varepsilon_i)_{i \in \mathbb{N}}$ is a sequence of d -dimensional random vectors which are independent of each other and identically distributed with density $F_\theta(x)$. N and $(\varepsilon_i)_{i \in \mathbb{N}}$ are also independent of each other. In our setting, the random measure p can be regarded as the one associated with the process z ; see Chapter 2.

$$p(dt, dz) = \sum_{s \geq 0} \mathbf{1}_{\{\Delta z_s \neq 0\}} \mathbf{1}_{(s, \Delta z_s)}(dt, dz).$$

Hence if the flow z has a jump of size z at time t , then X will have a jump of the size $c(X_{t-}, z, \theta)$ at the same time.

Now let us discuss the approximation of the transition probability. First, we consider the transition probability from $X_{t_{i-1}^n}$ to $X_{t_i^n}$ in the case of single jump in the interval $[t_{i-1}^n, t_i^n]$. We set $\tau_i^n := \inf\{t; |\Delta X_t| > 0, t_{i-1}^n \leq t < t_i^n\}$. Since no jump occurs in $[t_{i-1}^n, \tau_i^n]$, we approximate the transition by the one of

$$X_{\tau_i^n-} = X_{t_{i-1}^n} + \bar{a}_{i-1}(\tau_i^n - t_{i-1}^n) + b_{i-1}Z,$$

where $Z \sim \mathcal{N}_d(0, (\tau_i^n - t_{i-1}^n)I)$. The above X is not the same as the solution process to (3.1), but the above $X_{t_{i-1}^n}$ has the same value as the one of (3.1). Next, since we suppose that no jump occurs after a jump at the time τ_i^n , we can take the same approximation as above in $[\tau_i^n, t_i^n)$, that is,

$$X_{t_i^n} = X_{\tau_i^n-} + \bar{a}(X_{\tau_i^n})(t_i^n - \tau_i^n) + b(X_{\tau_i^n})Z' + c(X_{\tau_i^n-}, \Delta z_{\tau_i^n}),$$

where $Z' \sim \mathcal{N}_d(0, (t_i^n - \tau_i^n)I)$. Let $\phi(x; A, B)$ be a Gaussian density with the mean vector A and variance matrix B . Since the distribution of the jump time τ_i^n conditional on $\{J_i^n = 1\}$ becomes the uniform distribution on $[t_{i-1}^n, t_i^n)$,

$$\begin{aligned} & P_{i-1}^n \{X_{t_i^n} \in A, J_i^n = 1\} / P\{J_i^n = 1\} \\ &= \int_A \int_z \int_{x'} \int_{t_{i-1}^n}^{t_i^n} \frac{1}{h_n} \phi\left(x'; X_{t_{i-1}^n} + \bar{a}_{i-1}(s - t_{i-1}^n), \beta_{i-1}(s - t_{i-1}^n)\right) \times \\ & \quad \times \phi(x; x' + c(x', z) + \bar{a}(x' + c(x', z))(t_i^n - s), \beta(x' + c(x', z))(t_i^n - s)) \times \\ & \quad \times F(z) ds dx' dz dx. \end{aligned}$$

We denote by $p_{i,n}^d(x)$ the above probability density function. Secondly, using the local Gaussian approximation for $J_i^n = 0$,

$$P_{i-1}^n \{X_{t_i^n} \in A, J_i^n = 0\} / P\{J_i^n = 0\} = \int_A \phi(x; X_{t_i^n} + \bar{a}_{i-1}h_n, \beta h_n) dx.$$

We denote by $p_{i,n}^c(x)$ the integrand in the right-hand side. Finally,

$$P_{i-1}^n \{X_{t_i^n} \in A, J_i^n \geq 2\} = O_p(h_n^2).$$

Since

$$\begin{aligned} P_{i-1}^n \{X_{t_i^n} \in A\} &= P_{i-1}^n \{X_{t_i^n} \in A, J_i^n = 0\} + P_{i-1}^n \{X_{t_i^n} \in A, J_i^n = 1\} \\ & \quad + P_{i-1}^n \{X_{t_i^n} \in A, J_i^n \geq 2\} \\ &= \sum_{j=0}^2 [P_{i-1}^n \{X_{t_i^n} \in A, C_{i,j}^n\} + P_{i-1}^n \{X_{t_i^n} \in A, D_{i,j}^n\}], \end{aligned}$$

and by Lemma 3.3, we have the following relations:

$$\begin{aligned} P_{i-1}^n \{X_{t_i^n} \in A, C_{i,0}^n\} &= e^{-\lambda_0 h_n} \int_A \mathbf{1}_{\{|x - X_{t_{i-1}^n}| \leq h_n^2\}} p_{i,n}^c(x) dx, \\ P_{i-1}^n \{X_{t_i^n} \in A, C_{i,1}^n\} &\leq P_{i-1}^n \{C_{i,1}^n\} = O_p(h_n^3), \\ P_{i-1}^n \{X_{t_i^n} \in A, C_{i,2}^n\} &= O_p(h_n^2), \end{aligned}$$

$$\begin{aligned}
P_{i-1}^n\{X_{t_i^n} \in A, D_{i,0}^n\} &\leq P\{D_{i,0}^n\} = O_p(h_n^3), \\
P_{i-1}^n\{X_{t_i^n} \in A, D_{i,1}^n\} &= \lambda_0 h_n e^{-\lambda_0 h_n} \int_A \mathbf{1}_{\{|x - X_{t_{i-1}^n}| > h_n^\rho\}} p_{i,n}^d(x) dx, \\
P_{i-1}^n\{X_{t_i^n} \in A, D_{i,2}^n\} &= O_p(h_n^2).
\end{aligned}$$

Therefore we can approximate the transition density $p_{i,n}(x)$ by

$$\begin{aligned}
&\log p_{i,n}(x) \\
&\approx \mathbf{1}_{\{|x - X_{t_{i-1}^n}| \leq h_n^\rho\}} \log(p_{i,n}^c(x) e^{-\lambda_0 h_n}) + \mathbf{1}_{\{|x - X_{t_{i-1}^n}| > h_n^\rho\}} \log(p_{i,n}^d(x) \lambda_0 h_n e^{-\lambda_0 h_n}).
\end{aligned}$$

By the way, in the expression

$$\begin{aligned}
p_{i,n}^d(x) &= \int_z \int_{x'} \int_{t_{i-1}^n}^{t_i^n} \frac{1}{h_n} \phi\left(x'; X_{t_{i-1}^n} + \bar{a}_{i-1}(s - t_{i-1}^n), \beta_{i-1}(s - t_{i-1}^n)\right) \times \\
&\quad \times \phi(x; x' + c(x', z) + \bar{a}(x' + c(x', z))(t_i^n - s), \beta(x' + c(x', z))(t_i^n - s)) \times \\
&\quad \times F_{\theta_0}(z) ds dx' dz,
\end{aligned}$$

we can approximate ϕ to δ -function if h_n decreases rapidly, and then

$$\begin{aligned}
p_{i,n}^d(x) &\approx \int_z \int_{x'} \delta_{X_{t_{i-1}^n}}(x') \delta_{x-x'}(c(x', z)) F_{\theta_0}(z) dx' dz \\
&= \int_y \int_{x'} \delta_{X_{t_{i-1}^n}}(x') \delta_{x-x'}(y) F_{\theta_0}(c^{-1}(x', y)) J(x', y, \theta_0) dx' dy \\
&= F_{\theta_0}\left(c^{-1}(X_{t_{i-1}^n}, x - X_{t_{i-1}^n})\right) J(X_{t_{i-1}^n}, x - X_{t_{i-1}^n}, \theta_0) \\
&= \lambda_0^{-1} \Psi_{\theta_0}(x - X_{t_{i-1}^n}, X_{t_{i-1}^n}).
\end{aligned}$$

Moreover, since $\lambda(\theta) = \iint \Psi_\theta(y, x) dy d\pi$, $\lambda(\theta)$ can be approximated by the data as

$$\frac{1}{n} \sum_{i=1}^n \int \Psi_\theta(y, X_{t_{i-1}^n}) dy,$$

thanks to the ergodicity of X . These considerations lead the contrast function in Definition 3.1.

3.3 Moment estimates in the finite activity case

In this section, we introduce some useful moment inequalities, and they will be used repeatedly in the proofs below.

First, we prepare some notations. We consider the following inequality which is immediately obtained from the expression of the stochastic differential equation (3.1): for a constant $p \in \mathbb{N}$,

$$\begin{aligned} |X_t - X_{t_{i-1}^n}|^p &\leq 3^{p-1} \left\{ \left| \int_{t_{i-1}^n}^t a(X_s) ds \right|^p + \left| \int_{t_{i-1}^n}^t b(X_s) dw_s \right|^p \right. \\ &\quad \left. + \left| \int_{t_{i-1}^n}^t \int c(X_{s-}, z) (p-q)(ds, dz) \right|^p \right\}. \end{aligned} \quad (3.15)$$

Let

$$\begin{aligned} H_t &:= \left| \int_{t_{i-1}^n}^t a(X_s) ds \right|^p + \left| \int_{t_{i-1}^n}^t b(X_s) dw_s \right|^p, \\ M_t &:= \int_{t_{i-1}^n}^t \int c(X_{s-}, z) (p-q)(ds, dz), \end{aligned}$$

and also

$$N_t := \int_{t_{i-1}^n}^t \int |c(X_{s-}, z)|^2 (p-q)(ds, dz).$$

We obtain the following lemma.

Lemma 3.4 *Assume Conditions A1, A3 and A6. For $p = 2^q$, $q \in \mathbb{N}$, $t_{i-1}^n \leq t \leq t_i^n$,*

$$E_{i-1}^n [|M_t|^p] \leq C_p E_{i-1}^n \left[\int \int_{t_{i-1}^n}^t |c(X_s, z)|^p q(ds, dz) \right].$$

Proof . This proof follows from Bichteler and Jacod [13].

By using Doob's inequality ,

$$\begin{aligned} E_{i-1}^n \left[\sup_{t_{i-1}^n \leq s \leq t} |M_s|^2 \right] &\leq 4E_{i-1}^n [|M_t|^2] \leq 4E_{i-1}^n [|\langle M, M \rangle_t|] \\ &\leq 4E_{i-1}^n \left[\int \int_{t_{i-1}^n}^t |c(X_s, z)|^2 q(ds, dz) \right], \end{aligned}$$

hence Lemma 3.4 holds for $q = 1$. Next we suppose that Lemma holds for $p = 2^q$. We notice that $[M, M]_t = N_t + \langle M, M \rangle_t$, for all $p \in \mathbb{N}$, then

$$|[M, M]_t|^p \leq 2^{p-1} (|N_t|^p + |\langle M, M \rangle_t|^p).$$

Using the Burkholder-Davis-Gundy's inequality and the above inequality ,

$$\begin{aligned} E_{i-1}^n [|M_t|^{2p}] &\leq C_p E_{i-1}^n [|N_t|^p + \langle M, M \rangle_t^p] \\ &\leq C_p E_{i-1}^n \left[\int_{t_{i-1}^n}^t |c(X_s, z)|^{2p} q(ds, dz) + \left(\int_{t_{i-1}^n}^t |c(X_s, z)|^2 q(ds, dz) \right)^p \right]. \end{aligned}$$

Applying Jensen's inequality to the last second term, we have

$$E_{i-1}^n [|M_t|^{2p}] \leq C_p E \left[\int_{t_{i-1}^n}^t |c(X_s, z)|^{2p} q(ds, dz) \right].$$

Therefore Lemma holds for $p = 2^{q+1}$. \square

Lemma 3.5 Assume Conditions A1, A3 and A6. For $2 \leq k$, $k \in \mathbb{N}$, $t \in [t_{i-1}^n, t_i^n]$,

$$E_{i-1}^n [|X_t - X_{t_{i-1}^n}|^k] \leq C_k |t - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^k. \quad (3.16)$$

If g is a function defined on $\mathbb{R}^d \times \Xi$ and is of polynomial growth in x uniformly in α , then it follows that

$$E_{i-1}^n [|g(X_t, \alpha)|] \leq C(1 + |X_{t_{i-1}^n}|)^C. \quad (3.17)$$

Proof . First, we consider the case $p = 2^q$, $q \in \mathbb{N}$. Let us notice the inequality (3.15). Applying the linear growthness of a, b and Burkholder-Davis-Gundy's inequality, we easily obtain that

$$E_{i-1}^n [H_t] \leq C_p |t - t_{i-1}^n|^{p/2} (1 + |X_{t_{i-1}^n}|)^p + C_p \int_{t_{i-1}^n}^t E_{i-1}^n [|X_s - X_{t_{i-1}^n}|^p] ds. \quad (3.18)$$

Applying Lemma 3.4 and the linear growth of $c(x, z)$, we obtain

$$E_{i-1}^n [|M_t|^p] \leq C_p |t - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^p + C_p \int_{t_{i-1}^n}^t E_{i-1}^n [|X_s - X_{t_{i-1}^n}|^p] ds. \quad (3.19)$$

the inequalities (3.18) , (3.19) and the Gronwall's inequality yield , for all $q \in \mathbb{N}$,

$$E_{i-1}^n [|X_t - X_{t_{i-1}^n}|^p] \leq C_p |t - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^p.$$

For arbitrary $k \geq 2$, if we write $k = \sum_{q=0}^l \delta_q 2^q$ by the binary expansion, where $\delta_q = 0$ or 1 then

$$E_{i-1}^n [|X_t - X_{t_{i-1}^n}|^k] = E_{i-1}^n \left[\prod_{q=0}^l |X_t - X_{t_{i-1}^n}|^{\delta_q 2^q} \right].$$

Therefore we obtain the inequality (3.16) by using Cauchy-Schwarz's inequality repeatedly.

On (3.17), we can write

$$\begin{aligned} E_{i-1}^n [|g(X_t, \alpha)|] &\leq C E_{i-1}^n [(1 + |X_t|)^C] \\ &\leq C \left(1 + |X_{t_{i-1}^n}|^C + E_{i-1}^n [|X_t - X_{t_{i-1}^n}|^C] \right), \end{aligned}$$

and (3.16) ends the proof. \square

Remark 3.6 If we take the same argument as in the proof of Lemma 3.5 then we obtain the following moment inequality: let $k \geq 1$, $t_{i-1}^n \leq t \leq t_i^n$. For any $p \geq 1$,

$$E_{i-1}^n [|X_t - X_{t_{i-1}^n}|^k \mathbf{1}_{C_{i,0}^n}] \leq C_k |t - t_{i-1}^n|^{k/2} e^{-\lambda_0 h_n} (1 + |X_{t_{i-1}^n}|)^k + R(\alpha, h_n^p, X_{t_{i-1}^n}). \quad (3.20)$$

Lemma 3.6 Assume Conditions A1, A3 and A4 - A6. Let $\bar{X}_{i,n} = X_{t_i^n} - X_{t_{i-1}^n} - h_n \bar{a}_{i-1}(\theta_0)$. For all $k_j = 1, 2, \dots, d$ ($j = 1, 2, 3, 4$),

$$E_{i-1}^n [\bar{X}_{i,n}^{(k_1)} \mathbf{1}_{C_{i,0}^n}] = R(\alpha, h_n^2, X_{t_{i-1}^n}), \quad (3.21)$$

$$E_{i-1}^n [\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \mathbf{1}_{C_{i,0}^n}] = h_n e^{-\lambda_0 h_n} \beta_{i-1}^{(k_1, k_2)} + R(\alpha, h_n^2, X_{t_{i-1}^n}), \quad (3.22)$$

$$E_{i-1}^n [\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \mathbf{1}_{C_{i,0}^n}] = R(\alpha, h_n^2, X_{t_{i-1}^n}), \quad (3.23)$$

$$\begin{aligned} E_{i-1}^n [\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{C_{i,0}^n}] \\ = h_n^2 e^{-\lambda_0 h_n} \left(\beta_{i-1}^{(k_1, k_2)} \beta_{i-1}^{(k_3, k_4)} + \beta_{i-1}^{(k_1, k_3)} \beta_{i-1}^{(k_2, k_4)} + \beta_{i-1}^{(k_1, k_4)} \beta_{i-1}^{(k_2, k_3)} \right) \\ + R(\alpha, h_n^3, X_{t_{i-1}^n}). \end{aligned} \quad (3.24)$$

Proof . We prove (3.24); the others are done similarly .

Let Y be a solution to the following stochastic differential equation

$$dY_t = \bar{a}(Y_t) dt + b(Y_t) dw_t,$$

which is independent of J_i^n . A simple calculation deduces the multidimensional case of Lemma 7 in Kessler [50], that is,

$$\begin{aligned} E_{i-1}^n [\bar{Y}_{i,n}^{(k_1)} \bar{Y}_{i,n}^{(k_2)} \bar{Y}_{i,n}^{(k_3)} \bar{Y}_{i,n}^{(k_4)}] \\ = h_n^2 \left(\beta_{i-1}^{(k_1, k_2)} \beta_{i-1}^{(k_3, k_4)} + \beta_{i-1}^{(k_1, k_3)} \beta_{i-1}^{(k_2, k_4)} + \beta_{i-1}^{(k_1, k_4)} \beta_{i-1}^{(k_2, k_3)} \right) \\ + R(\alpha, h_n^3, X_{t_{i-1}^n}). \end{aligned} \quad (3.25)$$

Since J_i^n is independent of \mathcal{F}_{i-1}^n , we have

$$\begin{aligned}
& E_{i-1}^n \left[\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{\{J_i^n=0\}} \right] \\
&= E_{i-1}^n \left[\bar{Y}_{i,n}^{(k_1)} \bar{Y}_{i,n}^{(k_2)} \bar{Y}_{i,n}^{(k_3)} \bar{Y}_{i,n}^{(k_4)} \mathbf{1}_{\{J_i^n=0\}} \right] \\
&= E_{i-1}^n \left[\bar{Y}_{i,n}^{(k_1)} \bar{Y}_{i,n}^{(k_2)} \bar{Y}_{i,n}^{(k_3)} \bar{Y}_{i,n}^{(k_4)} \right] P_0 \{J_i^n = 0\} \\
&= h_n^2 e^{-\lambda_0 h_n} \left(\beta_{i-1}^{(k_1,k_2)} \beta_{i-1}^{(k_3,k_4)} + \beta_{i-1}^{(k_1,k_3)} \beta_{i-1}^{(k_2,k_4)} + \beta_{i-1}^{(k_1,k_4)} \beta_{i-1}^{(k_2,k_3)} \right) \\
&\quad + R(\alpha, h_n^3, X_{t_{i-1}^n}).
\end{aligned}$$

On the other hand,

$$\begin{aligned}
& E_{i-1}^n \left[\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{\{J_i^n=0\}} \right] \\
&= E_{i-1}^n \left[\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{C_{i,0}^n} \right] + E_{i-1}^n \left[\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{D_{i,0}^n} \right].
\end{aligned}$$

According to (3.16) and Lemma 3.3 ,

$$\begin{aligned}
& \left| E_{i-1}^n \left[\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)} \mathbf{1}_{D_{i,0}^n} \right] \right| \\
&\leq \sqrt{E_{i-1}^n \left[(\bar{X}_{i,n}^{(k_1)} \bar{X}_{i,n}^{(k_2)} \bar{X}_{i,n}^{(k_3)} \bar{X}_{i,n}^{(k_4)})^2 \right] P_{i-1}^n \{D_{i,0}^n\}} \\
&\leq \sqrt{E_{i-1}^n [C (|\Delta_i X^n|^8 + |\bar{a}_{i-1} h_n|^8)] P_{i-1}^n \{D_{i,0}^n\}} \\
&\leq R(\alpha, h_n^p, X_{t_{i-1}^n}).
\end{aligned}$$

This completes the proof. \square

3.4 Limit theorems and some remarks

The asymptotic properties of estimators are usually deduced from the asymptotic behavior of the estimating function. We therefore prepare some limit theorems for triangular arrays of the data. Since we assumed the ergodicity for the continuous data such as A2; the weak law of large numbers, this property is inherited to the discontinuous array.

Our contrast function is divided into two parts, the discretization of the likelihood of the continuous part and the one of the jump part. The limit theorems corresponding to the continuous part can be proved as if the data are from a diffusion process, and the proofs are similar to those of classical limit theorems for arrays by diffusion data, see e.g. Kessler [50]. On the other hand, the limit theorems corresponding to the discontinuous part; Proposition 3.4, are proved by a kind of complicated, but useful

techniques. That argument will be used again in Section 3.5. In addition, we describe some remarks, in which it will be described that those theorems can be applied to our contrast functions.

In the sequel, we denote by b_n a real valued sequence satisfying

$$b_n \rightarrow \infty, \quad nh_n^2 b_n \rightarrow 0, \quad \frac{b_n}{nh_n} \rightarrow 0.$$

There certainly exists such a sequence; e.g. if $h_n = n^{-2/3}$ then $b_n = n^{1/4}$ and so on.

Proposition 3.1 *Assume Conditions A1 - A3, A5 - A7, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Moreover suppose that $g^{(n)}$ is a function: $\mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ which satisfies the following conditions: $|g^{(n)}(x, \alpha)|^4 \leq L(x, \alpha)$, $|\partial_x g^{(n)}(x, \alpha)| \leq O(\sqrt{b_n})(1 + |x|)^C$ and $|\partial_\alpha g^{(n)}(x, \alpha)| \leq C(1 + |x|)^C$, where L is a π -integrable function for all α , and that there exist a function g for each α such that, as $n \rightarrow \infty$,*

$$g^{(n)}(x, \alpha) \longrightarrow g(x, \alpha) \quad \pi\text{-a.s.}$$

Then g is a π -integrable function and the following (i) , (ii) and (iii) hold as $n \rightarrow \infty$:

$$\begin{aligned} \text{(i)} \quad & \sup_{\alpha \in \Xi} \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \int g(x, \alpha) \pi(dx) \right| \xrightarrow{P} 0, \\ \text{(ii)} \quad & \sup_{\alpha \in \Xi} \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} - \int g(x, \alpha) \pi(dx) \right| \xrightarrow{P} 0, \\ \text{(iii)} \quad & \sup_{\alpha \in \Xi} \left| \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} - \lambda_0 \int g(x, \alpha) \pi(dx) \right| \xrightarrow{P} 0. \end{aligned}$$

Proof . The π -integrability of $g(x)$ is led from the uniform integrability of $g^{(n)}(x, \alpha)$. Let us prove that each convergence holds for fixed α . We start with the proof of (i) .

$$\begin{aligned} & P \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \int g(x, \alpha) \pi(dx) \right| > \varepsilon \right\} \\ & \leq P \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds \right| > \frac{\varepsilon}{3} \right\} \\ & \quad + P \left\{ \left| \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds - \frac{1}{nh_n} \int_0^{nh_n} g(X_s, \alpha) ds \right| > \frac{\varepsilon}{3} \right\} \\ & \quad + P \left\{ \left| \frac{1}{nh_n} \int_0^{nh_n} g(X_s, \alpha) ds - \int g(x, \alpha) \pi(dx) \right| > \frac{\varepsilon}{3} \right\}. \end{aligned}$$

The third term on the right-hand side converges to zero by the assumption of ergodicity. Let us call the first and second terms P_n^1 and P_n^2 , respectively, then

$$\begin{aligned} P_n^1 &\leq \frac{3}{\varepsilon} E \left[\left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds \right| \right] \\ &\leq \frac{3}{\varepsilon} E \left[\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |g^{(n)}(X_s, \alpha) - g_{i-1}^{(n)}(\alpha)| ds \right]. \end{aligned}$$

Applying Taylor's formula and Schwarz' inequality, we see that

$$\begin{aligned} P_n^1 &\leq \frac{3}{nh_n \varepsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left(E \left[|X_s - X_{t_{i-1}^n}|^2 \right] \right)^{\frac{1}{2}} \\ &\quad \times \left(E \left[\left(\int_0^1 \partial_x g^{(n)}(X_{t_{i-1}^n} + u(X_s - X_{t_{i-1}^n})) du \right)^2 \right] \right)^{\frac{1}{2}} ds. \end{aligned}$$

The inequalities (3.16) and (3.17) of Lemma 3.5 yield

$$\begin{aligned} P_n^1 &\leq \frac{3}{nh_n \varepsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left(E \left[C |s - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^C \right] \right)^{\frac{1}{2}} \times \\ &\quad \times \left(E \left[O(b_n) (1 + |X_{t_{i-1}^n}|)^C \right] \right)^{\frac{1}{2}} ds \\ &\leq \frac{O(\sqrt{b_n})}{nh_n \varepsilon} \sum_{i=1}^n \left(\int_{t_{i-1}^n}^{t_i^n} |s - t_{i-1}^n|^{1/2} ds \right) \\ &\leq O(\sqrt{h_n b_n}). \end{aligned}$$

Moreover

$$\begin{aligned} P_n^2 &\leq \frac{3}{\varepsilon nh_n} \int_0^{nh_n} E |g^{(n)}(X_t, \alpha) - g(X_t, \alpha)| dt \\ &= \frac{3}{\varepsilon} \int |g^{(n)}(x, \alpha) - g(x, \alpha)| d\pi. \end{aligned}$$

This converges to zero by Lebesgue's convergence theorem.

Next we show the convergence (iii). What we should show are the following (a) and (b) thanks to Lemma 9 in Genon-Catalot and Jacod [31]:

$$(a) \quad \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{|\Delta_i X^n| > h_n^p\}} \right] \xrightarrow{P} \lambda_0 \int g(x, \alpha) d\pi(x),$$

$$(b) \quad \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n^2 h_n^2} \left(g_{i-1}^{(n)}(\alpha) \right)^2 \mathbf{1}_{\{|\Delta_i X^n| > h_n^p\}} \right] \xrightarrow{P} 0.$$

(a) : By the same argument as for (i), it is sufficient to show that

$$I_n := P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} - \lambda_0 \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds \right| > \varepsilon \right\} \longrightarrow 0.$$

This is easily seen as follows:

$$\begin{aligned} I_n &\leq \frac{1}{\varepsilon} E \left[\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left| g_{i-1}^{(n)}(\alpha) \frac{1}{h_n} P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} - \lambda_0 g^{(n)}(X_s, \alpha) \right| ds \right] \\ &\leq \frac{1}{nh_n \varepsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left[E \left| g_{i-1}^{(n)}(\alpha) \left(\frac{1}{h_n} P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} - \lambda_0 \right) \right| \right. \\ &\quad \left. + \lambda_0 E |g_{i-1}^{(n)}(\alpha) - g^{(n)}(X_s, \alpha)| \right] ds \\ &\leq \frac{1}{nh_n \varepsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left[\sqrt{E |g_{i-1}^{(n)}(\alpha)|^2} \sqrt{E \left| \frac{1}{h_n} P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} - \lambda_0 \right|^2} \right. \\ &\quad \left. + \lambda_0 E |g_{i-1}^{(n)}(\alpha) - g^{(n)}(X_s, \alpha)| \right] ds = O(\sqrt{h_n b_n}). \end{aligned}$$

Here we applied Lemma 3.3 to the term $P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \}$ and the same argument as the proof of (i) to the term $E |g_{i-1}^{(n)}(\alpha) - g^{(n)}(X_s, \alpha)|$.

(b) :

$$\begin{aligned} &P \left\{ \left| \frac{1}{n^2 h_n^2} \sum_{i=1}^n \left(g_{i-1}^{(n)}(\alpha) \right)^2 P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} \right| > \varepsilon \right\} \\ &\leq \frac{1}{n^2 h_n^2 \varepsilon} \sum_{i=1}^n E \left| \left(g_{i-1}^{(n)}(\alpha) \right)^2 P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} \right| \\ &\leq \frac{1}{n^2 h_n \varepsilon} \sum_{i=1}^n \sqrt{E |g_{i-1}^{(n)}(\alpha)|^4 E \left| \frac{1}{h_n} P_{i-1}^n \{ |\Delta_i X^n| > h_n^\rho \} \right|^2} \\ &= O \left(\frac{1}{nh_n} \right). \end{aligned}$$

We can easily deduce (ii) for each fixed α from (i) and (iii) since

$$\frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{ |\Delta_i X^n| \leq h_n^\rho \}} = \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - h_n \left(\frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{ |\Delta_i X^n| > h_n^\rho \}} \right).$$

Finally we have to show the uniformity of the convergence in α . We only show (i); the uniformity in (ii) can be proved similarly and that in (iii) is shown by the

same argument as the proof of more general Proposition 3.4, so we omit the proof here. Let $s_n(\alpha) = \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha)$, and we regard this as a random element taking values in $(C(\Xi), \|\cdot\|_\infty)$. It suffices to see the tightness of this sequence; see Remark B.1 in Appendix B. Since we already showed the convergence of the marginal distributions of $s_n(\alpha)$, the tightness is implied by

$$\sup_n E \left[\sup_\alpha |\partial_\alpha s_n(\alpha)| \right] < \infty;$$

see Corollary B.1. And this is clear if we use the stationarity and Assumption A3. \square

Remark 3.7 Below, we sometimes use (i) as

$$g^{(n)}(x, \alpha) = \int \partial_\theta^k \Phi_n(\theta, x, y) dy \quad (k = 0, 1, 2).$$

We are able to check the above conditions for these $g^{(n)}$ by A9 and A11. Actually, for these $g^{(n)}$,

$$|g^{(n)}(x, \alpha)| \leq \left(\int L_1(y) dy \right) (1 + |x|)^C$$

from condition (3.3), and it is obvious that $|\partial_\alpha g^{(n)}(x, \alpha)| \leq C(1 + |x|)^C$ similarly. Moreover,

$$\begin{aligned} |\partial_x g^{(n)}(x, \alpha)| &\leq \int |\partial_x \partial_\theta^k \Psi_\theta(y, x) \cdot \varphi_n| dy + \int |\partial_\theta^k \Psi_\theta(y, x)| |\partial_x \varphi_n| dy \\ &\leq O(\varepsilon_n^{-1})(1 + |x|)^C \\ &\leq O(\sqrt{b_n})(1 + |x|)^C \end{aligned}$$

by (3.3), (3.4) and (3.7).

Proposition 3.2 Assume Conditions A1 - A7, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Suppose that a function $g : \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ and its derivatives $\partial_\alpha g$ and $\partial_x g$ are of polynomial growth uniformly in α . Then as $n \rightarrow \infty$,

$$\sup_{\alpha \in \Xi} \left| \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^p\}} - \int g(x, \alpha) \beta^{(k,l)}(x) \pi(dx) \right| \xrightarrow{P} 0$$

for $k, l = 1, 2, \dots, d$.

Proof . We set

$$\zeta_i^n(\alpha) := \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}},$$

and show that

$$\begin{aligned} A_n &:= \sum_{i=1}^n E_{i-1}^n [\zeta_i^n(\alpha)] \xrightarrow{P} \int g(x, \alpha) \beta^{(k,l)}(x) d\pi(x), \\ B_n &:= \sum_{i=1}^n E_{i-1}^n [(\zeta_i^n(\alpha))^2] \xrightarrow{P} 0. \end{aligned}$$

Using (3.22), we have

$$\begin{aligned} A_n &= \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \sum_{j=0}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \\ &= \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \left\{ h_n e^{-\lambda_0 h_n} \beta_{i-1}^{(k,l)} + R(\alpha, h_n^2, X_{t_{i-1}^n}) \right. \\ &\quad \left. + \sum_{j=1}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \right\}. \end{aligned}$$

Here, for sufficiently large n , we have

$$\left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right| \leq 2 \left\{ (h_n^\rho)^2 + |\bar{a}_{i-1} h_n|^2 \right\} = R(\alpha, h_n^{2\rho}, X_{t_{i-1}^n}).$$

Hence

$$\begin{aligned} \left| \sum_{j=1}^2 \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \right| &\leq \sum_{j=1}^2 \frac{1}{nh_n} \sum_{i=1}^n R(\alpha, h_n^{2\rho}, X_{t_{i-1}^n}) P_{i-1}^n \{C_{i,j}^n\} \\ &= O_p(h_n^{1+2\rho}). \end{aligned}$$

Therefore, by Proposition 3.1 (i),

$$A_n = \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) h_n e^{-\lambda_0 h_n} \beta_{i-1}^{(k,l)} + O_p(h_n) \xrightarrow{P} \int g(x, \alpha) \beta^{(k,l)}(x) d\pi(x).$$

The convergence of B_n can be proved similarly as for A_n .

The proof of the uniformity of convergence is the same as for Proposition 3.1 (i), that is, we set

$$s_n(\alpha) = \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}},$$

and using (3.24) ,

$$\begin{aligned}
\sup_n E \left[\sup_\alpha |\partial_\alpha s_n(\alpha)| \right] &\leq \frac{1}{nh_n} \sum_{i=1}^n E \left[C(1 + |X_{t_{i-1}^n}|)^C \left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right| \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\leq \frac{C}{nh_n} \sum_{i=1}^n \sqrt{\sum_{j=0}^2 E \left[\left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right|^2 \mathbf{1}_{C_{i,j}^n} \right]} \\
&\leq \frac{C}{nh_n} \sum_{i=1}^n E[R(\alpha, h_n, X_{t_{i-1}^n})] < \infty.
\end{aligned}$$

This completes the proof. \square

Proposition 3.3 *Under the same assumptions as in Proposition 3.2,*

$$\sup_{\alpha \in \Xi} \left| \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right| \xrightarrow{P} 0$$

for $k = 1, 2, \dots, d$ as $n \rightarrow \infty$.

Proof . We prove the convergence for each α in a similar way as for Proposition 3.2. Let

$$\zeta_i^n(\alpha) = \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}.$$

It suffices to show that

$$A_n := \sum_{i=1}^n E_{i-1}^n [\zeta_i^n(\alpha)] \xrightarrow{P} 0, \quad B_n := \sum_{i=1}^n E_{i-1}^n [(\zeta_i^n(\alpha))^2] \xrightarrow{P} 0.$$

Using (3.21), we have

$$\begin{aligned}
A_n &= \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \sum_{j=0}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \mathbf{1}_{C_{i,j}^n} \right] \\
&= \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \left\{ R(\alpha, h_n^2, X_{t_{i-1}^n}) + \sum_{j=1}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \mathbf{1}_{C_{i,j}^n} \right] \right\}.
\end{aligned}$$

Here, for sufficiently large n , we have

$$\left| \bar{X}_{i,n}^{(k)} \right| \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \leq h_n^\rho + |\bar{a}_{i-1} h_n| = R(\alpha, h_n^\rho, X_{t_{i-1}^n}).$$

Hence

$$\left| \sum_{j=1}^2 \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \mathbf{1}_{C_{i,j}^n} \right] \right| \leq \sum_{j=1}^2 \frac{1}{nh_n} \sum_{i=1}^n R(\alpha, h_n^\rho, X_{t_{i-1}^n}) P_{i-1}^n \{C_{i,j}^m\}$$

$$= O_p(h_n^{1+\rho}).$$

Therefore $A_n = O_p(h_n)$. The convergence of B_n can be proved similarly as for A_n .

Next, we show the tightness of

$$\begin{aligned} s_n(\alpha) &:= \sum_{i=1}^n \zeta_i^n(\alpha) \\ &= \sum_{i=1}^n \bar{\zeta}_i^n(\alpha) + \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}, \end{aligned}$$

where $\bar{\zeta}_i^n(\alpha) = \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(k)}$. The tightness of the second term in the last right-hand side is shown by the same argument as in the proof of Proposition 3.4 below. Therefore we only show the tightness of the first term $\sum_{i=1}^n \bar{\zeta}_i^n(\alpha)$.

According to Theorem B.8 in Appendix B, we should verify the following criterion: for any $N \in \mathbb{N}$ and some positive constant H independent of n ,

$$E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] \leq H, \quad (3.26)$$

$$E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha_1) - \sum_{i=1}^n \bar{\zeta}_i^n(\alpha_2) \right)^{2N} \right] \leq H |\alpha_1 - \alpha_2|^{2N}. \quad (3.27)$$

Let $G(s, \alpha) = \sum_{i=1}^n g_{i-1}(\alpha) \mathbf{1}_{[t_{i-1}^n, t_i^n)}(s)$, we have

$$\begin{aligned} \sum_{i=1}^n \bar{\zeta}_i^n(\alpha) &= \frac{1}{nh_n} \left\{ \int_0^{nh_n} G(s, \alpha) \bar{a}^{(k)}(X_s) ds + \sum_{j=1}^r \int_0^{nh_n} G(s, \alpha) b^{(k,j)}(X_s) dW_s^{(j)} \right. \\ &\quad \left. + \int_0^{nh_n} \int G(s, \alpha) c^{(k)}(X_{s-}, z) (p - q)(ds, dz) - \sum_{i=1}^n g_{i-1}(\alpha) \bar{a}_{i-1}^{(k)} h_n \right\}. \end{aligned}$$

Therefore, the left-hand side of (3.26) is evaluated as follows:

$$\begin{aligned} &E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] \\ &\leq C_N \left\{ E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} G(s, \alpha) \bar{a}^{(k)}(X_s) ds \right)^{2N} \right] \right. \end{aligned}$$

$$\begin{aligned}
& + \sum_{j=1}^r E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} G(s, \alpha) b^{(k,j)}(X_s) dW_s^{(j)} \right)^{2N} \right] \\
& + E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} \int G(s, \alpha) c^{(k)}(X_{s-}, z) (p - q)(ds, dz) \right)^{2N} \right] \\
& + E \left[\left(\frac{1}{n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{a}_{i-1}^{(k)} \right)^{2N} \right] \Bigg\}.
\end{aligned}$$

Applying Jensen's and Burkholder-Davis-Gundy's inequality, we see that

$$\begin{aligned}
E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] & \leq C_N \left\{ \frac{1}{nh_n} \int_0^{nh_n} E \left[G^{2N}(s, \alpha) (\bar{a}^{(k)}(X_s))^{2N} \right] ds \right. \\
& + \frac{1}{(nh_n)^{N+1}} \sum_{j=1}^r \int_0^{nh_n} E \left[G^{2N}(s, \alpha) (b^{(k,j)}(X_s))^{2N} \right] ds \\
& + \frac{1}{(nh_n)^{N+1}} \int_0^{nh_n} \int E \left[G^{2N}(s, \alpha) (c^{(k)}(X_{s-}, z))^{2N} \right] q(ds, dz) \\
& \left. + E \left[\left(\frac{1}{n} \sum_{i=1}^n g_{i-1}(\alpha) \bar{a}_{i-1}^{(k)} \right)^{2N} \right] \right\}.
\end{aligned}$$

One can see that these all are bounded because of Assumption A3 . On (3.27), by using Jensen's inequality and Burkholder-Davis-Gundy's inequality, we have that

$$\begin{aligned}
& E \left[\left(\frac{1}{|\alpha_1 - \alpha_2|} \sum_{i=1}^n \{ \bar{\zeta}_i^n(\alpha_1) - \bar{\zeta}_i^n(\alpha_2) \} \right)^{2N} \right] \\
& \leq C_N \left\{ \frac{1}{nh_n} \int_0^{nh_n} E \left[\left\{ \frac{G(s, \alpha_1) - G(s, \alpha_2)}{|\alpha_1 - \alpha_2|} \right\}^{2N} (\bar{a}^{(k)}(X_s))^{2N} \right] ds \right. \\
& + \frac{1}{(nh_n)^{N+1}} \sum_{j=1}^r \int_0^{nh_n} E \left[\left\{ \frac{G(s, \alpha_1) - G(s, \alpha_2)}{|\alpha_1 - \alpha_2|} \right\}^{2N} (b^{(k,j)}(X_s))^{2N} \right] ds \\
& + \frac{1}{(nh_n)^{N+1}} \int_0^{nh_n} \int E \left[\left\{ \frac{G(s, \alpha_1) - G(s, \alpha_2)}{|\alpha_1 - \alpha_2|} \right\}^{2N} (c^{(k)}(X_{s-}, z))^{2N} \right] q(ds, dz) \\
& \left. + \frac{1}{n} \sum_{i=1}^n E \left[\left\{ \frac{g_{i-1}(\alpha_1) - g_{i-1}(\alpha_2)}{|\alpha_1 - \alpha_2|} \right\}^{2N} \{ \bar{a}_{i-1}^{(k)} \}^{2N} \right] \right\}. \tag{3.28}
\end{aligned}$$

Since $\partial_\alpha g$ is of polynomial growth uniformly in α , we have

$$\frac{|G(s, \alpha_1) - G(s, \alpha_2)|}{|\alpha_1 - \alpha_2|} \leq \sum_{i=1}^n \sup_{\alpha} |\partial_\alpha g_{i-1}(\alpha)| \mathbf{1}_{[t_{i-1}^n, t_i^n)}(s)$$

$$\leq \sum_{i=1}^n C(1 + |X_{t_{i-1}^n}|)^C \mathbf{1}_{[t_{i-1}^n, t_i^n]}(s),$$

and

$$\frac{|g_{i-1}(\alpha_1) - g_{i-1}(\alpha_2)|}{|\alpha_1 - \alpha_2|} \leq \sup_{\alpha} |\partial_{\alpha} g_{i-1}(\alpha)| \leq C(1 + |X_{t_{i-1}^n}|)^C.$$

Hence we see that (3.28) is bounded, and this completes the proof. \square

Proposition 3.4 *Assume Conditions A1 - A3, A5 - A7, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Suppose $g_n(\alpha, y, x) : \Xi \times \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies that, for $y_1, y_2 \in \mathbb{R}^d$ and $\eta \in [0, 1]$,*

$$|\partial_{\alpha} g_n(\alpha, y_1, x) - \partial_{\alpha} g_n(\alpha, y_2, x)| \leq \tilde{g}_n(\alpha, \eta y_1 + (1 - \eta)y_2, x) |y_1 - y_2|, \quad (3.29)$$

where $\tilde{g}_n(\alpha, y, x) \leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C$, and

$$|\partial_{\alpha}^m g_n(\alpha, y, x)| \leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C \quad (m = 0, 1). \quad (3.30)$$

Assume that the integral $G_n(\alpha, x) = \int g_n(\alpha, y, x) \Psi(y, x) dy$ exists for all x and α , and that, for a π -integrable function $L(x, \alpha)$,

$$|G_n(\alpha, x)|^4 \leq L(x, \alpha), \quad (3.31)$$

$$|\partial_x G_n(\alpha, x)| \leq O(\sqrt{b_n})(1 + |x|)^C. \quad (3.32)$$

Moreover there exists a function g such that

$$G_n(\alpha, x) \longrightarrow \int g(\alpha, y, x) \Psi(y, x) dy \quad \pi\text{-a.s.} \quad (3.33)$$

and the last integral is a π -integrable function for all α . Furthermore, assume that

$$\int \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y, x)| \Psi(y, x) dy \leq C(1 + |x|)^C \quad (3.34)$$

for all x . Then it follows as $n \rightarrow \infty$ that

$$\sup_{\alpha \in \Xi} \left| \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{\{|\Delta_i X^n| > h_n^{\rho}\}} - \iint g(\alpha, y, x) \Psi(y, x) dy d\pi(x) \right| \xrightarrow{P} 0.$$

Proof . First, we show the convergence for each α .

Applying Hölder's inequality, for $p > 1, \delta \in (0, 1/3]$ which satisfies $\frac{1}{p} + \frac{1}{1+\delta} = 1$ and $\varepsilon > 0$,

$$\sum_{j=0,2} P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,j}^n} \right| > \varepsilon \right\}$$

$$\begin{aligned}
&\leq \sum_{j=0,2} \frac{1}{\varepsilon n h_n} \sum_{i=1}^n E \left[\left| g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,j}^n} \right| \right] \\
&\leq \sum_{j=0,2} \frac{1}{\varepsilon n h_n} \sum_{i=1}^n \left(E \left| g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \right|^p \right)^{1/p} (P\{D_{i,j}^n\})^{\frac{1}{1+\delta}} \\
&= O \left(h_n^{\frac{1-\delta}{1+\delta}} \sqrt{b_n} \right) \\
&= O \left(\sqrt{h_n b_n} \cdot h_n^{\frac{1-3\delta}{2+2\delta}} \right) = o(1).
\end{aligned}$$

From this, we have the following inequality:

$$\begin{aligned}
&P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{\{|\Delta_i X^n| > h_n^2\}} - \iint g_n(\alpha, y, x) \Psi(y, x) dy d\pi \right| > 3\varepsilon \right\} \\
&= P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n \sum_{j=0}^2 g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,j}^n} - \iint g_n(\alpha, y, x) \Psi(y, x) dy d\pi \right| > 3\varepsilon \right\} \\
&\leq P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,1}^n} - \iint g_n(\alpha, y, x) \Psi(y, x) dy d\pi \right| > \varepsilon \right\} \\
&\quad + \sum_{j=0,2} P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,j}^n} \right| > \varepsilon \right\} \\
&\leq \sum_{k=1}^5 I_k + o(1),
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,1}^n} - \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,1}^n} \right| > \frac{\varepsilon}{5} \right\}, \\
I_2 &= P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{D_{i,1}^n} - \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} \right| > \frac{\varepsilon}{5} \right\}, \\
I_3 &= P \left\{ \left| \frac{1}{n h_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} - \frac{1}{n h_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) p(ds, dz) \right| > \frac{\varepsilon}{5} \right\},
\end{aligned}$$

$$\begin{aligned}
I_4 &= P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}^n) p(ds, dz) \right. \right. \\
&\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}^n) q(ds, dz) \right| > \frac{\varepsilon}{5} \right\}, \\
I_5 &= P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}^n) q(ds, dz) \right. \right. \\
&\quad \left. \left. - \iint g(\alpha, y, x) \Psi(y, x) dy d\pi \right| > \frac{\varepsilon}{5} \right\}.
\end{aligned}$$

Let us evaluate these terms. Applying Hölder's inequality, for $p > 1$ and $q > 1$ with $1/p + 1/q = 1$, we see

$$\begin{aligned}
I_1 &\leq \frac{5}{\varepsilon nh_n} \sum_{i=1}^n E \left[|g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}^n) - g_n(\alpha, \Delta X_{\tau_i^n}^n, X_{t_{i-1}^n}^n)| \mathbf{1}_{D_{i,1}^n} \right] \\
&\leq \frac{5}{\varepsilon nh_n} \sum_{i=1}^n E \left[\left| \tilde{g}_n(\alpha, \xi_i^n, X_{t_{i-1}^n}^n) \right| \left(|X_{t_i^n}^n - X_{\tau_i^n}^n| + |X_{\tau_i^n}^n - X_{t_{i-1}^n}^n| \right) \mathbf{1}_{\{J_i^n=1\}} \right] \\
&\leq \frac{5}{\varepsilon nh_n} \sum_{i=1}^n E \left[\left(E \left[\left| \tilde{g}_n(\alpha, \xi_i^n, X_{t_{i-1}^n}^n) \right|^2 \middle| J_i^n = 1 \right] \right)^{1/2} \right. \\
&\quad \left. \left(E \left[\left(|X_{t_i^n}^n - X_{\tau_i^n}^n| + |X_{\tau_i^n}^n - X_{t_{i-1}^n}^n| \right)^2 \middle| J_i^n = 1 \right] \right)^{1/2} \mathbf{1}_{\{J_i^n=1\}} \right],
\end{aligned}$$

where $\xi_i^n = \eta \Delta_i X^n + (1 - \eta) \Delta X_{\tau_i^n}^n$ for some $[0, 1]$ -valued random variable η . Here, we notice that X follows the following stochastic differential equation on the set $\{J_i^n = 1\}$;

$$\tilde{X}_t - \tilde{X}_{t_{i-1}^n} = H_t + \int_{t_{i-1}^n}^t \bar{a}(\tilde{X}_s) ds + \int_{t_{i-1}^n}^t b(\tilde{X}_s) dw_s,$$

where $\tilde{X}_{t_{i-1}^n} = X_{t_{i-1}^n}^n$, $H_t = c(X_{u-}, z) \mathbf{1}_{[u, t_i^n]}(t)$, u is a $[t_{i-1}^n, t_i^n]$ -valued uniform random variable which is independent of $(w_t)_{t \geq 0}$ and J_i^n , and z is a random variable with density F_{θ_0} which is independent of $(w_t)_{t \geq 0}$. Therefore, for example,

$$\begin{aligned}
E \left[|X_{\tau_i^n}^n - X_{t_{i-1}^n}^n|^2 \middle| J_i^n = 1 \right] &= E \left[|X_{\tau_i^n}^n - X_{t_{i-1}^n}^n|^2 \mathbf{1}_{\{J_i^n=1\}} \right] / P\{J_i^n = 1\} \\
&= E \left[|\tilde{X}_{u-} - \tilde{X}_{t_{i-1}^n}^n|^2 \right].
\end{aligned}$$

Applying the Burkholder-Davis-Gundy's inequality to (3.13), we see that

$$E \left[\sup_{t \in [t_{i-1}^n, \tau_i^n -]} |\tilde{X}_t - \tilde{X}_{t_{i-1}^n}^n|^2 \right] \leq C \left\{ h_n^2 + E \left[\int_{t_{i-1}^n}^{t_i^n} b^2(\tilde{X}_s) ds \right] \right\} = O(h_n).$$

Hence $E \left[\left| X_{\tau_i^n} - X_{t_{i-1}^n} \right|^2 \middle| J_i^n = 1 \right] = O(h_n)$. Similarly, by (3.14),

$$E \left[\sup_{t \in [\tau_i^n, t_i^n]} |\tilde{X}_{t_i^n} - \tilde{X}_t|^2 \right] = O(h_n),$$

and $E \left[\left| X_{t_i^n} - X_{\tau_i^n} \right|^2 \middle| J_i^n = 1 \right] = O(h_n)$. Hence $I_1 = O(\sqrt{h_n b_n})$.

$$\begin{aligned} I_2 &= P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{C_{i,1}^n} \right| > \frac{\varepsilon}{5} \right\} \\ &\leq \frac{5}{\varepsilon n h_n} \sum_{i=1}^n E \left[\left| g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{C_{i,1}^n} \right| \right] \\ &\leq \frac{C}{n h_n} \sum_{i=1}^n O(\sqrt{b_n}) \sqrt{P\{C_{i,1}^n\}} = O(\sqrt{h_n b_n}). \end{aligned}$$

$$\begin{aligned} I_3 &\leq P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, c_{i-1}(\Delta z_{\tau_i^n}), X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} \right| > \frac{\varepsilon}{10} \right\} \\ &\quad + P \left\{ \left| \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, c_{i-1}(\Delta z_{\tau_i^n}), X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) p(ds, dz) \right| > \frac{\varepsilon}{10} \right\}. \end{aligned}$$

The first term of the right-hand side becomes $O(\sqrt{h_n b_n})$ by the same argument as I_1 .

Denote the second term by I'_3 then

$$\begin{aligned} I'_3 &\leq \frac{10}{\varepsilon n h_n} \sum_{i=1}^n E \left[\left| g_n(\alpha, c_{i-1}(\Delta z_{\tau_i^n}), X_{t_{i-1}^n}) \mathbf{1}_{\{J_i^n=1\}} \right. \right. \\ &\quad \left. \left. - \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) p(ds, dz) \right| \right] \\ &= \frac{10}{n h_n} \sum_{i=1}^n E \left[\left| \int_{t_{i-1}^n}^{t_i^n} \int \mathbf{1}_{\{J_i^n \geq 2\}} g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) p(ds, dz) \right| \right] \\ &\leq \frac{10}{n h_n} \sum_{i=1}^n (P\{J_i^n \geq 2\})^{1/2} \left\| \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) p(ds, dz) \right\|_{L^2(P)} \end{aligned}$$

$$\leq \frac{C}{n} \sum_{i=1}^n \left\{ E \left[\int_{t_{i-1}^n}^{t_i^n} \int g_n^2(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) q(ds, dz) \right] \right\}^{1/2} = O(\sqrt{h_n b_n}).$$

Hence $I_3 = O(\sqrt{h_n b_n})$. Furthermore

$$\begin{aligned} I_4 &\leq \frac{25}{\varepsilon^2} E \left[\left(\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) (p - q)(ds, dz) \right)^2 \right] \\ &\leq \frac{25}{\varepsilon^2 n^2 h_n^2} \sum_{i=1}^n E \left[\int_{t_{i-1}^n}^{t_i^n} \int g_n^2(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) q(ds, dz) \right] \\ &\quad + \frac{50}{\varepsilon^2 n^2 h_n^2} \sum_{i < j} E \left[\int_{t_{i-1}^n}^{t_i^n} \int g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n}) (p - q)(ds, dz) \right. \\ &\quad \times \left. E \left[\int_{t_{j-1}^n}^{t_j^n} \int g_n(\alpha, c_{j-1}(z), X_{t_{j-1}^n}) (p - q)(ds, dz) \middle| \mathcal{F}_{j-1}^n \right] \right] = O\left(\frac{b_n}{nh_n}\right). \end{aligned}$$

The term I_5 clearly converges to zero by Proposition 3.1 (i); see Remark 3.9.

Let us show the uniformity of convergence. Set

$$s_n(\alpha) = \frac{1}{nh_n} \sum_{i=1}^n g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n}) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}.$$

We prove the tightness of $\{s_n(\alpha)\}$. The expectation

$$\begin{aligned} E \left[\sup_{\alpha} |\partial_{\alpha} s_n(\alpha)| \right] &\leq \frac{1}{nh_n} \sum_{i=1}^n \sum_{j=0}^2 E \left[\sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,j}^n} \right] \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n E \left[\sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right] + o\left(\sqrt{h_n b_n}\right) \end{aligned}$$

by condition (3.30) and Hölder's inequality, and we can show that

$$\begin{aligned} &\frac{1}{nh_n} \sum_{i=1}^n E \left[\sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right] \\ &= \iint \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y, x)| \Psi(y, x) dy d\pi + O\left(\sqrt{h_n b_n}\right). \end{aligned}$$

Indeed, by the same argument as above,

$$\left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right] \right|$$

$$\left| - \iint \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y, x)| \Psi(y, x) dy d\pi \right| \leq \sum_{k=1}^5 H_k,$$

where

$$\begin{aligned} H_1 &= \left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta_i X^n, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right] \right|, \\ H_2 &= \left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n})| \mathbf{1}_{D_{i,1}^n} \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n})| \mathbf{1}_{\{J_i^n=1\}} \right] \right|, \\ H_3 &= \left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, \Delta X_{\tau_i^n}, X_{t_{i-1}^n})| \mathbf{1}_{\{J_i^n=1\}} \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n})| p(ds, dz) \right] \right|, \\ H_4 &= \left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n})| p(ds, dz) \right. \right. \\ &\quad \left. \left. - \frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n})| q(ds, dz) \right] \right|, \\ H_5 &= \left| E \left[\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, c_{i-1}(z), X_{t_{i-1}^n})| q(ds, dz) \right. \right. \\ &\quad \left. \left. - \iint \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y, x)| \Psi(y, x) dy d\pi \right] \right|. \end{aligned}$$

We obtain that $H_1 = O(\sqrt{h_n b_n})$ by the same argument as for I_1 since

$$\begin{aligned} \left| \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y_1, x)| - \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y_2, x)| \right| &\leq \sup_{\alpha} |\partial_{\alpha} g_n(\alpha, y_1, x) - \partial_{\alpha} g_n(\alpha, y_2, x)| \\ &\leq \tilde{g}_n(\alpha, \eta y_1 + (1 - \eta) y_2, x) |y_1 - y_2|. \end{aligned}$$

Similarly, we can obtain that $H_2 + H_3 = O(\sqrt{h_n b_n})$. Moreover it is also easy to see that $H_4 = H_5 = 0$. Hence, $\sup_n E \left[\sup_{\alpha} |\partial_{\alpha} s_n(\alpha)| \right] < \infty$. This ends the proof. \square

Remark 3.8 Condition (3.29) is satisfied if $\partial_\alpha g_n(\alpha, y, x)$ is differentiable with respect to y and

$$|\partial_y \partial_\alpha g_n(\alpha, y, x)| \leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C. \quad (3.35)$$

Remark 3.9 Thanks to Conditions (3.29) - (3.34), we are able to apply Proposition 3.1 (i) to $G_n(\alpha, x)$.

Remark 3.10 Below, we use this proposition as

$$g_n(\alpha, y, x) = \partial_\theta^k \log \Phi_n(\theta, x, y) \cdot \varphi_n(x, y) \quad (k = 0, 1, 2).$$

We are able to check the above conditions for this g_n by A9. Indeed,

$$\begin{aligned} |\partial_y g_n(\alpha, y, x)| &\leq |\partial_y \partial_\theta^k \log \Phi_n| |\varphi_n| + |\partial_y \varphi_n| |\partial_\theta^k \log \Phi_n| \\ &\leq O(\varepsilon_n^{-2})(1 + |y|)^C(1 + |x|)^C + O(\varepsilon_n^{-3})(1 + |y|)^C(1 + |x|)^C \\ &\leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C \end{aligned}$$

from (3.3), (3.5), (3.7) and (3.8). Similarly,

$$\begin{aligned} |\partial_y \partial_\alpha g_n(\alpha, y, x)| &= |\partial_y [\partial_\theta^{k+1} \log \Psi_\theta(y, x) \cdot \varphi_n(x, y)]| \\ &\leq O(\varepsilon_n^{-4})(1 + |y|)^C(1 + |x|)^C \\ &\leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C. \end{aligned}$$

Therefore (3.35) is satisfied, and this implies (3.29).

On the inequality (3.30),

$$\begin{aligned} |\partial_\alpha^m g_n(\alpha, y, x)| &= |\partial_\theta^{k+m} \log \Psi_\theta(y, x) \cdot \varphi_n(x, y)| \\ &\leq O(\varepsilon_n^{-2-m})(1 + |y|)^C(1 + |x|)^C \\ &\leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C. \end{aligned}$$

On the inequality (3.31),

$$|g_n(\alpha, y, x) \Psi(y, x)| \leq \{|\partial_\theta^k \log \Psi_\theta(y, x)| + 1\} \Psi(y, x), \quad (3.36)$$

therefore (3.3) and (3.6) yield (3.31).

Next,

$$\begin{aligned} |\partial_x G_n(\alpha, x)| &\leq \int |\partial_x (\partial_\theta^k \log \Phi_n \cdot \varphi_n) \Psi| dy + \int |\partial_\theta^k \log \Phi_n \cdot \varphi_n| |\partial_x \Psi| dy \\ &\leq \int |\partial_x \varphi_n \partial_\theta^k \log \Phi_n| |\Psi| dy + \int |\varphi_n \partial_x \partial_\theta^k \log \Phi_n| |\Psi| dy \end{aligned}$$

$$+ \int |\partial_\theta^k \log \Phi_n \cdot \varphi_n| |\partial_x \Psi| dy.$$

The inequalities (3.3), (3.4), (3.7) and (3.8) imply that

$$\begin{aligned} |\partial_x \varphi_n \partial_\theta^k \log \Phi_n| &\leq O(\varepsilon_n^{-3}) L(y) (1 + |x|)^C, \\ |\varphi_n \partial_x \partial_\theta^k \log \Phi_n| &\leq O(\varepsilon_n^{-3}) L(y) (1 + |x|)^C, \\ |\partial_\theta^k \log \Phi_n \cdot \varphi_n| &\leq O(\varepsilon_n^{-2}) L(y) (1 + |x|)^C, \end{aligned}$$

where $L(y)$ is a bounded dy -integrable function. Then

$$\begin{aligned} |\partial_x G_n(\alpha, x)| &\leq \int O(\varepsilon_n^{-3}) L(y) (1 + |x|)^C \Psi dy + \int O(\varepsilon_n^{-2}) L_2(y) L(y) (1 + |x|)^C dy \\ &\leq O(\sqrt{b_n}) (1 + |x|)^C. \end{aligned}$$

Therefore (3.32) is satisfied. The condition (3.33) is obtained from Lebesgue's theorem thanks to (3.36).

The inequality (3.34) is obtained from (3.6) directly with $g(\alpha, x, y) = \partial_\theta^k \log \Phi_\theta(y, x)$.

3.5 Proof of the main theorem

We proceed the proof of Theorem 3.1.

3.5.1 Proof of consistency

Let us prove the consistency of $\hat{\alpha}_n$.

Applying Propositions 3.1 (i), (ii), 3.2 and 3.4, we can easily obtain that

$$\frac{1}{n} \bar{l}_n(\alpha) \xrightarrow{P} U_1(\sigma, \sigma_0) = -\frac{1}{2} \int \{ \text{tr} (\beta^{-1}(x, \sigma) \beta(x, \sigma_0)) + \log \det \beta(x, \sigma) \} d\pi, \quad (3.37)$$

$$\frac{1}{nh_n} \tilde{l}_n(\theta) \xrightarrow{P} U_2(\theta, \theta_0) = \iint \{ (\log \Psi_\theta(y, x)) \Psi_{\theta_0}(y, x) - \Psi_\theta(y, x) \} dy d\pi \quad (3.38)$$

uniformly in σ and θ . See Remark 3.7 and Remark 3.10 on the conditions for the convergence (3.38).

In order to prove the consistency of $\hat{\alpha}_n$, we may assume that the convergences of (3.37) and (3.38) take place almost surely and uniformly in the parameters, and prove that it implies $\hat{\alpha}_n \rightarrow \alpha_0$ almost surely since the convergence in probability implies that, for any subsequence, the existence of a subsequence converging almost surely.

For fixed $\omega \in \Omega$, thanks to the compactness of Ξ , there exists a subsequence n_k such that $\hat{\alpha}_{n_k} \rightarrow \alpha_\infty = (\theta_\infty, \sigma_\infty)$. Since the mapping $\sigma \rightarrow U_1(\sigma, \sigma_0)$ is continuous,

$$\frac{1}{n_k} l_{n_k}(\hat{\alpha}_{n_k}) \longrightarrow U_1(\sigma_\infty, \sigma_0),$$

and, by the definition of $\hat{\sigma}_n$, we have $U_1(\sigma_\infty, \sigma_0) \geq U_1(\sigma_0, \sigma_0)$. On the other hand, notice the following inequality:

$$\log \frac{\det \beta(x, \sigma_0)}{\det \beta(x, \sigma_\infty)} \leq \text{tr} [\beta^{-1}(x, \sigma_\infty) \beta(x, \sigma_0)] - d,$$

then we have $U_1(\sigma_\infty, \sigma_0) \leq U_1(\sigma_0, \sigma_0)$. Hence the equality $U_1(\sigma_\infty, \sigma_0) = U_1(\sigma_0, \sigma_0)$ and Assumption A10 lead that $\sigma_\infty = \sigma_0$. This implies that any convergent subsequence of $\hat{\sigma}_n$ tends to σ_0 . This means the consistency of $\hat{\sigma}_n$.

Next, let us show the consistency of $\hat{\theta}_n$. Since the mapping $\theta \rightarrow U_2(\theta, \theta_0)$ is also continuous,

$$\frac{1}{n_k h_{n_k}} \tilde{l}_{n_k}(\hat{\theta}_{n_k}) \longrightarrow U_2(\theta_\infty, \theta_0)$$

for fixed $\omega \in \Omega$. Here we prepare a lemma.

Lemma 3.7 *Assume Conditions A1 - A11, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Then*

$$\begin{aligned} & \frac{1}{nh_n} \bar{l}_n(\theta, \sigma) - \frac{1}{nh_n} \bar{l}_n(\theta_0, \sigma) \\ & \xrightarrow{P} -\frac{1}{2} \int (\bar{a}(x, \theta_0) - \bar{a}(x, \theta))^* \beta^{-1}(x, \sigma) (\bar{a}(x, \theta_0) - \bar{a}(x, \theta)) d\pi \end{aligned} \quad (3.39)$$

uniformly in α as $n \rightarrow \infty$.

Proof . By simple computation,

$$\begin{aligned} & \frac{1}{nh_n} \bar{l}_n(\theta, \sigma) - \frac{1}{nh_n} \bar{l}_n(\theta_0, \sigma) \\ & = -\frac{1}{2n} \sum_{i=1}^n (\bar{a}_{i-1}(\theta_0) - \bar{a}_{i-1}(\theta))^* \beta_{i-1}^{-1}(\sigma) (\bar{a}_{i-1}(\theta_0) - \bar{a}_{i-1}(\theta)) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \\ & \quad - \frac{1}{nh_n} \sum_{i=1}^n (\bar{a}_{i-1}(\theta_0) - \bar{a}_{i-1}(\theta))^* \beta_{i-1}^{-1}(\sigma) (\Delta_i X^n - \bar{a}_{i-1}(\theta_0) h_n) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}. \end{aligned}$$

Propositions 3.1 (ii) and 3.3 end the proof . \square

Thanks to this lemma and the continuity of the limit function (3.39),

$$\begin{aligned} \frac{1}{nh_n} l_n(\hat{\theta}_{n_k}, \hat{\sigma}_n) - \frac{1}{nh_n} l_n(\theta_0, \hat{\sigma}_n) \\ \xrightarrow{P} -\frac{1}{2} \int (\bar{a}(x, \theta_0) - \bar{a}(x, \theta_\infty))^* \beta^{-1}(x, \sigma_0) (\bar{a}(x, \theta_0) - \bar{a}(x, \theta_\infty)) d\pi \\ - \{U_2(\theta_0, \theta_0) - U_2(\theta_\infty, \theta_0)\}. \end{aligned}$$

The above limit is positive because of the definition of $\hat{\theta}_n$. Therefore θ_∞ satisfies $\Psi_{\theta_\infty}(x, z) = \Psi_{\theta_0}(x, z)$ and $\bar{a}(x, \theta_0) = \bar{a}(x, \theta_\infty)$ since β^{-1} is a positive definite and $U_2(\theta, \theta_0)$ will be maximized if and only if $\Psi_\theta(x, z) = \Psi_{\theta_0}(x, z)$. Thus the assumption A10 implies $\theta_\infty = \theta_0$. This ends the proof of consistency. \square

3.5.2 Proof of asymptotic normality

First, let us compute the first and the second derivatives of the contrast function. For $p, p' = 1, 2, \dots, m_1$ and $q, q' = 1, 2, \dots, m_2$,

$$\begin{aligned} \partial_{\theta_p} l_n(\alpha) &= \sum_{i=1}^n \{ \delta_{i,1}^p(\alpha) + \delta_{i,2}^p(\alpha) \}, \\ \delta_{i,1}^p(\alpha) &= \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\theta) (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(l)}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ \delta_{i,2}^p(\alpha) &= \partial_{\theta_p} \left\{ \log \Phi_n(\theta, X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\ &\quad - h_n \int \partial_{\theta_p} \Phi_n(\theta, X_{t_{i-1}^n}, y) dy, \\ \partial_{\sigma_q} l_n(\alpha) &= \sum_{i=1}^n \zeta_i^q(\alpha), \\ \zeta_i^q(\alpha) &= -\frac{1}{2} \left\{ \sum_{k,l=1}^d \frac{\partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)}(\sigma)}{h_n} \bar{X}_{i,n}^{(k)}(\theta) \bar{X}_{i,n}^{(l)}(\theta) \right. \\ &\quad \left. + \frac{\partial_{\sigma_q} \det \beta_{i-1}(\sigma)}{\det \beta_{i-1}(\sigma)} \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ \partial_{\theta_p \theta_{p'}}^2 l_n(\alpha) &= \sum_{i=1}^n \sum_{k,l=1}^d \left\{ \partial_{\theta_p \theta_{p'}}^2 \bar{a}_{i-1}^{(k)}(\theta) (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(l)}(\theta) \right. \\ &\quad \left. - (\partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\theta)) (\partial_{\theta_{p'}} \bar{a}_{i-1}^{(l)}(\theta)) (\beta_{i-1}^{-1})^{(k,l)}(\sigma) h_n \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \\ &\quad + \sum_{i=1}^n \left\{ \partial_{\theta_p \theta_{p'}}^2 \left\{ \log \Phi_n(\theta, X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right. \end{aligned}$$

$$\begin{aligned}
& - h_n \int \partial_{\theta_p \theta_{p'}}^2 \Phi_n(\theta, X_{t_{i-1}^n}, y) dy \Big\}, \\
\partial_{\sigma_q \sigma_{q'}}^2 l_n(\alpha) &= - \sum_{i=1}^n \left\{ \sum_{k,l=1}^d \frac{\partial_{\sigma_q \sigma_{q'}}^2 (\beta_{i-1}^{-1})^{(k,l)}(\sigma)}{2h_n} \bar{X}_{i,n}^{(k)}(\theta) \bar{X}_{i,n}^{(l)}(\theta) \right. \\
& \quad \left. + \frac{1}{2} \partial_{\sigma_q \sigma_{q'}}^2 \log \det \beta_{i-1}(\sigma) \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^2\}}, \\
\partial_{\theta_p \sigma_q}^2 l_n(\alpha) &= \partial_{\sigma_q \theta_p}^2 l_n(\alpha) \\
&= - \sum_{i=1}^n \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\theta) \partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(l)}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^2\}}.
\end{aligned}$$

We define the following notations:

$$M_n := \begin{pmatrix} \frac{1}{\sqrt{nh_n}} I_{m_1} & \mathbf{0} \\ \mathbf{0} & \frac{1}{\sqrt{n}} I_{m_2} \end{pmatrix},$$

where I_n is an n -dimensional identity matrix. Let

$$S_n := \begin{pmatrix} \sqrt{nh_n}(\hat{\theta}_n - \theta_0) \\ \sqrt{n}(\hat{\sigma}_n - \sigma_0) \end{pmatrix}, \quad L_n(\alpha) := \begin{pmatrix} -\frac{1}{\sqrt{nh_n}} \partial_{\theta} l_n(\alpha) \\ -\frac{1}{\sqrt{n}} \partial_{\sigma} l_n(\alpha) \end{pmatrix},$$

and

$$C_n(\alpha) := \begin{pmatrix} \frac{1}{nh_n} \partial_{\theta}^2 l_n(\alpha) & \frac{1}{n\sqrt{h_n}} \partial_{\theta\sigma}^2 l_n(\alpha) \\ \frac{1}{n\sqrt{h_n}} \partial_{\sigma\theta}^2 l_n(\alpha) & \frac{1}{n} \partial_{\sigma}^2 l_n(\alpha) \end{pmatrix}.$$

Then

$$M_n \partial_{\alpha}^2 l_n = \begin{pmatrix} \frac{1}{\sqrt{nh_n}} \partial_{\theta}^2 l_n(\alpha) & \frac{1}{\sqrt{nh_n}} \partial_{\theta\sigma}^2 l_n(\alpha) \\ \frac{1}{\sqrt{n}} \partial_{\sigma\theta}^2 l_n(\alpha) & \frac{1}{\sqrt{n}} \partial_{\sigma}^2 l_n(\alpha) \end{pmatrix} = C_n(\alpha) M_n^{-1}.$$

Now, by Taylor's formula,

$$\int_0^1 \partial_{\alpha}^2 l_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) du \begin{pmatrix} \hat{\theta}_n - \theta_0 \\ \hat{\sigma}_n - \sigma_0 \end{pmatrix} = -\partial_{\alpha} l_n(\alpha_0) \quad (3.40)$$

since $\partial l_n(\hat{\alpha}_n) = 0$. Then, multiplying both sides by M_n from the left, we have

$$\int_0^1 C_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) du S_n = L_n(\alpha_0). \quad (3.41)$$

Thus the asymptotic normality of S_n is proved by Lemmas 3.8 and 3.9 below.

Lemma 3.8 Assume Conditions A1 - A11, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Then the following statements hold:

- (i) $C_n(\alpha_0) \xrightarrow{P} B$, where $B = -K$ and K is given in Theorem 3.1.
- (ii) For any positive sequence ε_n tending to zero,

$$\sup_{|\alpha| \leq \varepsilon_n} |C_n(\alpha + \alpha_0) - C_n(\alpha_0)| \xrightarrow{P} 0 \quad (n \rightarrow \infty).$$

Proof . (i) We show the convergence of $n^{-1}\partial_\sigma^2 l_n(\alpha_0)$ only. The others are easily shown by using Propositions 3.1 (i), (ii), 3.2, 3.3 and 3.4.

Applying Propositions 3.1 (ii) and 3.2, we have

$$\frac{1}{n} \partial_{\sigma_q \sigma_{q'}}^2 l_n(\alpha_0) \xrightarrow{P} -\frac{1}{2} \int \text{tr} \left[\partial_{\sigma_q \sigma_{q'}}^2 \beta^{-1}(x) \beta(x) \right] d\pi - \frac{1}{2} \int \partial_{\sigma_q} \partial_{\sigma_{q'}} \log \det \beta(x) d\pi.$$

Noticing that $\partial_{\sigma_q} \log \det \beta(x, \sigma) = -\text{tr} [\partial_{\sigma_q} \beta^{-1}(x) \beta(x)]$, so also

$$\partial_{\sigma_q \sigma_{q'}}^2 \log \det \beta(x) = -\text{tr} \left[\partial_{\sigma_q \sigma_{q'}}^2 \beta^{-1}(x) \beta(x) \right] - \text{tr} \left[\partial_{\sigma_q} \beta^{-1}(x) \partial_{\sigma_{q'}} \beta(x) \right],$$

we can obtain that

$$\begin{aligned} \frac{1}{n} \partial_{\sigma_q \sigma_{q'}}^2 l_n(\alpha_0) &\xrightarrow{P} \frac{1}{2} \int \text{tr} \left[\partial_{\sigma_q} \beta^{-1}(x) \partial_{\sigma_{q'}} \beta(x) \right] d\pi \\ &= -\frac{1}{2} \int \text{tr} \left[(\partial_{\sigma_q} \beta) \beta^{-1} (\partial_{\sigma_{q'}} \beta) \beta^{-1} \right] (x) d\pi. \end{aligned}$$

- (ii) Let $B(\alpha)$ be the uniform limit of $C_n(\alpha)$, that is,

$$\sup_{\alpha \in H} |C_n(\alpha) - B(\alpha)| \xrightarrow{P} 0,$$

and $B(\alpha)$ is easily specified. Then, noticing $B(\alpha_0) = B$, we have

$$\begin{aligned} \sup_{|\alpha| \leq \varepsilon_n} |C_n(\alpha + \alpha_0) - C_n(\alpha_0)| \\ \leq 2 \sup_{|\alpha| \leq \varepsilon_n} |C_n(\alpha + \alpha_0) - B(\alpha + \alpha_0)| + \sup_{|\alpha| \leq \varepsilon_n} |B(\alpha + \alpha_0) - B|. \end{aligned}$$

The first term on the right-hand side converges to zero in probability by the uniformity of convergence, The second term also converges to zero in probability by the continuity of $B(\alpha)$. \square

Lemma 3.9 Assume Conditions A1 - A11, $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$, in addition, assume $nh_n^2 \rightarrow 0$. Then, as $n \rightarrow \infty$,

$$L_n(\alpha_0) \xrightarrow{d} L \sim \mathcal{N}_m(0, K).$$

Actually, by (3.41),

$$\left(\int_0^1 \{C_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) - C_n(\alpha_0)\} du + C_n(\alpha_0) \right) S_n = L_n(\alpha_0).$$

We find that the matrix

$$\int_0^1 \{C_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) - C_n(\alpha_0)\} du + C_n(\alpha_0) \quad (3.42)$$

converges in probability to the nonsingular matrix B . Hence, taking the limit on both sides after multiplying by the inverse of (3.42), we see by the continuous mapping theorem that

$$S_n \xrightarrow{d} B^{-1}L \sim \mathcal{N}_m(0, K^{-1}).$$

This ends the proof of the asymptotic normality of S_n . \square

Finally it remains to show Lemma 3.9.

Proof . Notice that the sequence $L_n(\alpha_0)$ is a triangular array. Therefore we can apply Theorem A.3 and its Remark A.1 in order to show the asymptotic normality of $L_n(\alpha_0)$. According to Remark 3.11 below, it suffices to show the following: For $p, p' = 1, \dots, m_1$, $q, q' = 1, \dots, m_2$ and some $\nu_1, \nu_2 > 0$,

$$\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \delta_{i,v}^p(\alpha_0) \right] \right| \xrightarrow{P} 0 \quad (v = 1, 2), \quad (3.43)$$

$$\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{n}} \zeta_i(\alpha_0)^q \right] \right| \xrightarrow{P} 0, \quad (3.44)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,1}^p(\alpha_0) \delta_{i,1}^{p'}(\alpha_0) \right] \xrightarrow{P} \int (\partial_{\theta_p} \bar{a})^* \beta^{-1} (\partial_{\theta_{p'}} \bar{a}(x, \alpha_0)) d\pi, \quad (3.45)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,2}^p(\alpha_0) \delta_{i,2}^{p'}(\alpha_0) \right] \xrightarrow{P} \iint \frac{\partial_{\theta_p} \Psi_{\theta_0} \partial_{\theta_{p'}} \Psi_{\theta_0}}{\Psi_{\theta_0}}(y, x) dy d\pi, \quad (3.46)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,1}^p(\alpha_0) \delta_{i,2}^{p'}(\alpha_0) \right] \xrightarrow{P} 0, \quad (3.47)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n} \zeta_i^q(\alpha_0) \zeta_i^{q'}(\alpha_0) \right] \xrightarrow{P} \frac{1}{2} \int \text{tr}[(\partial_{\sigma_q} \beta) \beta^{-1} (\partial_{\sigma_{q'}} \beta) \beta^{-1}] (x, \sigma_0) d\pi, \quad (3.48)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n \sqrt{h_n}} \delta_{i,v}^p(\alpha_0) \zeta_i^q(\alpha_0) \right] \xrightarrow{P} 0 \quad (v = 1, 2), \quad (3.49)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n h_n}} \delta_{i,v}^p(\alpha_0) \right|^{2+\nu_1} \right] \xrightarrow{P} 0 \quad (v = 1, 2), \quad (3.50)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n}} \zeta_i^q(\alpha_0) \right|^{2+\nu_2} \right] \xrightarrow{P} 0. \quad (3.51)$$

Remark 3.11 We use the central limit theorem for triangular arrays to show this lemma, so we have to check the Lindeberg condition for $L_n(\alpha_0) = \sum_{i=1}^n X_i^n$, that is,

$$\sum_{i=1}^n E_{i-1}^n \left[|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \right] \xrightarrow{P} 0$$

for any $\varepsilon > 0$. If X_i^n has an expression $X_i^n = Y_i^n + Z_i^n$ then

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \right] \\ & \leq 4 \sum_{i=1}^n E_{i-1}^n \left[|Y_i^n|^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon/2\}} \right] + 4 \sum_{i=1}^n E_{i-1}^n \left[|Z_i^n|^2 \mathbf{1}_{\{|Z_i^n| > \varepsilon/2\}} \right]. \end{aligned}$$

Hence, to check the above Lindeberg condition, it suffices to check the following Lyapunov conditions:

$$\sum_{i=1}^n E_{i-1}^n \left[|Y_i^n|^{2+\nu_1} \right], \quad \sum_{i=1}^n E_{i-1}^n \left[|Z_i^n|^{2+\nu_2} \right] \xrightarrow{P} 0$$

for some $\nu_1, \nu_2 > 0$. Here, it is not necessary that ν_1 and ν_2 are the same, so the above ν_i 's of (3.50) and (3.51) can be taken differently.

Proof of (3.43)

For $v = 1$, it follows that

$$\begin{aligned} & \sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \delta_{i,1}^p(\alpha_0) \right] \right| \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \sum_{j=0}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \right|. \end{aligned}$$

Since $|\bar{X}_{i,n}^{(l)}| \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} = R(\alpha, h_n^\rho, X_{t_{i-1}^n})$, we see that

$$\begin{aligned} & \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \sum_{j=1}^2 E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \right| \\ & \leq \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n R(\alpha, h_n^\rho, X_{t_{i-1}^n}) \sum_{j=1}^2 P_{i-1}^n \{C_{i,j}^n\} \\ & \leq \frac{1}{n} \sum_{i=1}^n R\left(\alpha, \sqrt{nh_n^{3+2\rho}}, X_{t_{i-1}^n}\right). \end{aligned}$$

Applying (3.21) to the term for $j = 0$, we have

$$\begin{aligned} & \sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \delta_{i,1}^p(\alpha_0) \right] \right| \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \right| R(\alpha, h_n^2, X_{t_{i-1}^n}) + o_p(\sqrt{nh_n^3}) \\ &= \frac{1}{n} \sum_{i=1}^n R(\alpha, \sqrt{nh_n^3}, X_{t_{i-1}^n}) + o_p(\sqrt{nh_n^3}) \xrightarrow{P} 0. \end{aligned}$$

For $v = 2$, it follows that

$$\begin{aligned} & \sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \delta_{i,2}^p(\alpha_0) \right] \right| \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| E_{i-1}^n \left[\sum_{j=0}^2 \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{D_{i,j}^n} \right. \right. \\ & \quad \left. \left. - h_n \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right] \right| \\ &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| E_{i-1}^n \left[\partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{D_{i,1}^n} \right. \right. \\ & \quad \left. \left. - h_n \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right] \right| + o_p\left(\sqrt{nh_n^2 b_n}\right) \end{aligned}$$

$$\leq \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \sum_{k=1}^5 I_k^i + o_p \left(\sqrt{nh_n^2 b_n} \right),$$

where

$$\begin{aligned} I_1^i &= \left| E_{i-1}^n \left[\partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{D_{i,1}^n} \right. \right. \\ &\quad \left. \left. - \partial_{\theta_p} \left\{ \log \Phi_n(\theta_0, X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n} \right] \right|, \\ I_2^i &= \left| E_{i-1}^n \left[\partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n} \right. \right. \\ &\quad \left. \left. - \partial_{\theta_p} \left\{ \log \Phi_n(\theta_0, X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}} \right] \right|, \\ I_3^i &= \left| E_{i-1}^n \left[\partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta X_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}} \right. \right. \\ &\quad \left. \left. - \int_{t_{i-1}^n}^{t_i^n} \int \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, c_{i-1}(z)) \right\} \varphi_n(X_{t_{i-1}^n}, c_{i-1}(z)) p(ds, dz) \right] \right|, \\ I_4^i &= \left| E_{i-1}^n \left[\int_{t_{i-1}^n}^{t_i^n} \int \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, c_{i-1}(z)) \right\} \varphi_n(X_{t_{i-1}^n}, c_{i-1}(z)) p(ds, dz) \right. \right. \\ &\quad \left. \left. - \int_{t_{i-1}^n}^{t_i^n} \int \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, c_{i-1}(z)) \right\} \varphi_n(X_{t_{i-1}^n}, c_{i-1}(z)) q(ds, dz) \right] \right|, \\ I_5^i &= \left| E_{i-1}^n \left[\int_{t_{i-1}^n}^{t_i^n} \int \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, c_{i-1}(z)) \right\} \varphi_n(X_{t_{i-1}^n}, c_{i-1}(z)) q(ds, dz) \right. \right. \\ &\quad \left. \left. - h_n \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right] \right|. \end{aligned}$$

Since, by Remark 3.10,

$$|\partial_y \partial_{\theta_p} \log \Phi_n(\theta_0, x, y) \varphi_n(x, y)| \leq O(\sqrt{b_n})(1 + |y|)^C(1 + |x|)^C,$$

we can apply the same argument to the terms $I_1^i - I_3^i$ as for $H_1 - H_3$ in the proof of Proposition 3.4, we easily obtain that

$$I_1^i = R\left(\alpha, \sqrt{h_n^3 b_n}, X_{t_{i-1}^n}\right), \quad I_2^i = R(\alpha, \sqrt{h_n^3 b_n}, X_{t_{i-1}^n}), \quad I_3^i = R\left(\alpha, \sqrt{h_n^3 b_n}, X_{t_{i-1}^n}\right).$$

It is also easy to see that $I_4^i = I_5^i = 0$. Thus

$$\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \delta_{i,2}^p(\alpha_0) \right] \right| = O_p \left(\sqrt{nh_n^2 b_n} \right).$$

Proof of (3.44)

Using Lemma 3.6 (3.22), we have

$$\begin{aligned}
\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{n}} \zeta_i^q(\alpha_0) \right] \right| &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| \sum_{j=0}^2 E_{i-1}^n \left[\sum_{k,l=1}^d \frac{\partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)}}{2h_n} \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,j}^n} \right] \right. \\
&\quad \left. + \frac{1}{2} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} P_{i-1}^n \{ |\Delta_i X^n| \leq h_n^\rho \} \right| \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| \frac{1}{2} \operatorname{tr} [\partial_{\sigma_q} \beta_{i-1}^{-1} \beta_{i-1}] e^{-\lambda_0 h_n} + R(\alpha, h_n, X_{t_{i-1}^n}) \right. \\
&\quad \left. + \frac{1}{2} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} \left(e^{-\lambda_0 h_n} + R(\alpha, h_n^3, X_{t_{i-1}^n}) + \lambda_0^2 h_n^2 \right) \right| \\
&= O_p(\sqrt{nh_n^2}).
\end{aligned}$$

We used the relation $\frac{\partial_{\sigma_q} \det \beta_{i-1}(\sigma)}{\det \beta_{i-1}(\sigma)} = -\operatorname{tr} [\partial_{\sigma_q} \beta_{i-1}^{-1}(\sigma) \beta_{i-1}(\sigma)]$.

Proof of (3.45)

Noticing that $|\bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} = R(\alpha, h_n^\rho, X_{t_{i-1}^n})$, we see from Lemma 3.6 (3.22) and Proposition 3.1 (i) that

$$\begin{aligned}
&\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,1}^p(\alpha_0) \delta_{i,1}^{p'}(\alpha_0) \right] \\
&= \frac{1}{nh_n} \sum_{i=1}^n E_{i-1}^n \left[\left(\sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \bar{X}_{i,n}^{(l)} \right) \right. \\
&\quad \left. \times \left(\sum_{k',l'=1}^d \partial_{\theta_{p'}} \bar{a}_{i-1}^{(k')} (\beta_{i-1}^{-1})^{(k',l')} \bar{X}_{i,n}^{(l')} \right) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\xrightarrow{P} \sum_{k,l,k',l'=1}^d \int \partial_{\theta_p} \bar{a}^{(k)} \partial_{\theta_{p'}} \bar{a}^{(k')} (\beta^{-1})^{(k,l)} (\beta^{-1})^{(k',l')} (\beta)^{(l,l')} (x, \alpha_0) d\pi \\
&= \int (\partial_{\theta_p} \bar{a})^* (\beta^{-1}) (\partial_{\theta_{p'}} \bar{a}) (x, \alpha) d\pi.
\end{aligned}$$

Proof of (3.46)

It follows by the direct computation that

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,2}^p(\alpha_0) \delta_{i,2}^{p'}(\alpha_0) \right]$$

$$\begin{aligned}
&= \frac{1}{nh_n} \sum_{i=1}^n E_{i-1}^n \left[\left\{ \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right. \right. \\
&\quad \left. \left. - h_n \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right\} \right. \\
&\quad \times \left\{ \partial_{\theta_{p'}} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right. \\
&\quad \left. \left. - h_n \int \partial_{\theta_{p'}} \Phi_n(X_{t_{i-1}^n}, y) dy \right\} \right] \\
&= \frac{1}{nh_n} \sum_{i=1}^n E_{i-1}^n \left[\partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right. \\
&\quad \times \partial_{\theta_{p'}} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \left. \right] + O_p(\sqrt{h_n b_n}) \\
&\xrightarrow{P} \iint \frac{\partial_{\theta_p} \Psi \partial_{\theta_{p'}} \Psi}{\Psi}(y, x) dy d\pi.
\end{aligned}$$

The last convergence is proved by the same argument as in the proof of (3.43) with $v = 2$ since, for $m = 0, 1$,

$$\left| \partial_y^m \left[\partial_{\theta_p} \log \Phi_n(\theta, x, y) \partial_{\theta_{p'}} \log \Phi_n(\theta, x, y) \varphi_n^2(x, y) \right] \right| \leq O(\sqrt{b_n})(1 + |x|)^C(1 + |y|)^C.$$

Proof of (3.47)

Noticing that $|\bar{X}_{i,n}^{(l)}| \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} = R(\alpha, h_n^\rho, X_{t_{i-1}^n})$, we see from Lemma 3.6 (3.21) that

$$\begin{aligned}
&\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \delta_{i,1}^p(\alpha_0) \delta_{i,2}^{p'}(\alpha_0) \right] \\
&= -\frac{1}{n} \sum_{i=1}^n E_{i-1}^n \left[\sum_{k,l=1}^d \left\{ \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right\} \partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\beta_{i-1}^{-1})^{(k,l)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&= -\frac{1}{n} \sum_{i=1}^n E_{i-1}^n \left[\sum_{k,l=1}^d \left\{ \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right\} \partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\beta_{i-1}^{-1})^{(k,l)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{C_{i,0}^n} \right] + o_p(h_n^2) \\
&= O(h_n^2).
\end{aligned}$$

Proof of (3.48)

Using the equalities (3.22), (3.24), and the relation

$$\frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} = -\text{tr} [\partial_{\sigma_q} \beta_{i-1}^{-1} \beta_{i-1}] = \text{tr} [\partial_{\sigma_q} \beta_{i-1} \beta_{i-1}^{-1}].$$

we see that

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n} \zeta_i^q(\alpha_0) \zeta_i^{q'}(\alpha_0) \right] \\
&= \frac{1}{4n} \sum_{i=1}^n E_{i-1}^n \left[\left\{ \sum_{k,l=1}^d \frac{\partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)}}{h_n} \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} + \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} \right\} \right. \\
&\quad \times \left. \left\{ \sum_{k',l'=1}^d \frac{\partial_{\sigma_{q'}} (\beta_{i-1}^{-1})^{(k',l')}}{h_n} \bar{X}_{i,n}^{(k')} \bar{X}_{i,n}^{(l')} + \frac{\partial_{\sigma_{q'}} \det \beta_{i-1}}{\det \beta_{i-1}} \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&= \frac{1}{4n} \sum_{i=1}^n \left[\sum_{k,l,k',l'=1}^d e^{-\lambda_0 h_n} (\partial_{\sigma_q} \beta_{i-1}^{-1})^{(k,l)} (\partial_{\sigma_{q'}} \beta_{i-1}^{-1})^{(k',l')} \left(\beta_{i-1}^{(k,l)} \beta_{i-1}^{(k',l')} + \beta_{i-1}^{(k,k')} \beta_{i-1}^{(l,l')} \right. \right. \\
&\quad \left. \left. + \beta_{i-1}^{(k,l')} \beta_{i-1}^{(k',l)} \right) + \sum_{k',l'=1}^d e^{-\lambda_0 h_n} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} \partial_{\sigma_{q'}} (\beta_{i-1}^{-1})^{(k',l')} \beta_{i-1}^{(k',l')} \right. \\
&\quad \left. + \sum_{k,l=1}^d e^{-\lambda_0 h_n} \frac{\partial_{\sigma_{q'}} \det \beta_{i-1}}{\det \beta_{i-1}} \partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)} \beta_{i-1}^{(k,l)} \right. \\
&\quad \left. + e^{-\lambda_0 h_n} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} \frac{\partial_{\sigma_{q'}} \det \beta_{i-1}}{\det \beta_{i-1}} \right] + O_p(h_n^{4\rho}) \\
&\xrightarrow{P} \frac{1}{2} \int \text{tr}[(\partial_{\sigma_q} \beta) \beta^{-1} (\partial_{\sigma_{q'}} \beta) \beta^{-1}] d\pi.
\end{aligned}$$

The last convergence is deduced by Proposition 3.1 (i).

Proof of (3.49)

For $v = 1$, it follows that

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \delta_{i,1}^p(\alpha_0) \zeta_i^q(\alpha_0) \right] \\
&= -\frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \sum_{k,l,k',l'=1}^d \frac{1}{2h_n} \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k',l')} \\
&\quad \times E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \bar{X}_{i,n}^{(k')} \bar{X}_{i,n}^{(l')} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\quad - \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \frac{1}{2} \sum_{k,l=1}^d \partial_{\theta_p} \bar{a}_{i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right].
\end{aligned}$$

Noticing that $|\bar{X}_{i,n}^{(l)}| \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} = R(\alpha, h_n^\rho, X_{t_{i-1}^n})$, we see from Lemma 3.6 (3.21) and (3.23) that

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \delta_{i,1}^p(\alpha_0) \zeta_i^q(\alpha_0) \right] = \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n R(\alpha, h_n, X_{t_{i-1}^n}) + O_p(\sqrt{h_n})$$

$$\xrightarrow{P} 0.$$

For $v = 2$, by using Lemma 3.6 (3.22), we see that

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \delta_{i,2}^p(\alpha_0) \zeta_i^q(\alpha_0) \right] \\ &= -\frac{1}{n\sqrt{h_n}} \sum_{i=1}^n E_{i-1}^n \left[\left\{ \sum_{k,l=1}^d \frac{\partial_{\sigma_q}(\beta_{i-1}^{-1})^{(k,l)}}{2h_n} \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} + \frac{1}{2} \frac{\partial_{\sigma_q} \det \beta_{i-1}}{\det \beta_{i-1}} \right\} \right. \\ & \quad \left. \times \left(h_n \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\ &= o_p(\sqrt{h_n}). \end{aligned}$$

Proof of (3.50)

For $v = 1$, it follows that

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \delta_{i,1}^p(\alpha_0) \right|^{2+\nu} \right] \\ & \leq \frac{C_\nu}{n^{1+\nu/2} h_n^{1+\nu/2}} \sum_{i=1}^n \sum_{k,l=1}^d \sum_{j=0}^2 \left| \partial_{\theta_p} \bar{a}_{i-1}^{(k)}(\beta_{i-1}^{-1})^{(k,l)} \right|^{2+\nu} E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{C_{i,j}^n} \right]. \end{aligned}$$

Noticing that $E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{C_{i,0}^n} \right] = R(\alpha, h_n^{1+\nu/2}, X_{t_{i-1}^n})$ from (3.20), we have

$$E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \delta_{i,1}^p(\alpha_0) \right|^{2+\nu} \right] = O_p\left(\frac{1}{n^{\nu/2}}\right) + o_p\left(\frac{h_n}{n^{\nu/2} h_n^{\nu/2}}\right).$$

For $v = 2$, it follows that

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \delta_{i,2}^p(\alpha_0) \right|^{2+\nu} \right] \\ & \leq \frac{C_\nu}{n^{1+\nu/2} h_n^{1+\nu/2}} \sum_{i=1}^n E_{i-1}^n \left[\left| \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right|^{2+\nu} \right. \\ & \quad \left. \times \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} + h_n^{2+\nu} \left| \int \partial_{\theta_p} \Phi_n(X_{t_{i-1}^n}, y) dy \right|^{2+\nu} \right] \\ & \leq \frac{C_\nu}{n^{1+\nu/2} h_n^{1+\nu/2}} \end{aligned}$$

$$\begin{aligned} & \times \sum_{i=1}^n E_{i-1}^n \left[\left| \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right|^{2+\nu} \left(\mathbf{1}_{D_{i,1}^n} + \mathbf{1}_{D_{i,2}^n} \right) \right] \\ & + O_p \left(\frac{h_n^{1+\nu/2}}{n^{\nu/2}} \right). \end{aligned}$$

Here, it follows from Assumption A9 (3.3) that

$$\begin{aligned} \left| \partial_{\theta_p} \left\{ \log \Phi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right\} \varphi_n(X_{t_{i-1}^n}, \Delta_i X^n) \right| & \leq \varepsilon_n^{-1} L_1(\Delta_i X^n) (1 + |X_{t_{i-1}^n}|)^C \\ & = R(\alpha, \varepsilon_n^{-1}, X_{t_{i-1}^n}) \end{aligned}$$

since L_1 is a bounded function given in (3.3). Then we have

$$\begin{aligned} \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n} h_n} \delta_{i,2}^p(\alpha_0) \right|^{2+\nu} \right] & \leq \frac{\varepsilon_n^{-(2+\nu)}}{n^{1+\nu/2} h_n^{1+\nu/2}} \sum_{i=1}^n R(\alpha, h_n, X_{t_{i-1}^n}) + O_p \left(\frac{h_n^{1+\nu/2}}{n^{\nu/2}} \right) \\ & = O_p \left(\left(\frac{b_n}{n h_n} \right)^{\nu/2} b_n^{(1-2\nu)/5} \right) + O_p \left(\frac{h_n^{1+\nu/2}}{n^{\nu/2}} \right). \end{aligned}$$

The last term converges to zero if $\nu \geq 1/2$.

Proof of (3.51)

It follows by the direct computation that

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n}} \zeta_i^q(\alpha_0) \right|^{2+\nu} \right] \\ & \leq \frac{C_\nu}{n^{1+\nu/2} h_n^{2+\nu}} \sum_{i=1}^n \sum_{k,l=1}^d E_{i-1}^n \left[\left| \partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)} \right|^{2+\nu} \left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\ & \quad + \frac{C_\nu}{n^{1+\nu/2}} \sum_{i=1}^n \left| \frac{\partial_q \det \beta_{i-1}}{\det \beta_{i-1}} \right|^{2+\nu} P_{i-1}^n \{ |\Delta_i X^n| \leq h_n^\rho \} \\ & \leq \frac{C}{n^{1+\nu/2} h_n^{2+\nu}} \sum_{i=1}^n \sum_{k,l=1}^d \sum_{j=0}^2 \left| \partial_{\sigma_q} (\beta_{i-1}^{-1})^{(k,l)} \right|^{2+\nu} E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{C_{i,j}^n} \right] \\ & \quad + O_p \left(\frac{1}{n^{\nu/2}} \right). \end{aligned}$$

We notice that $E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{C_{i,0}^n} \right] = R(\alpha, h_n^{2+\nu}, X_{t_{i-1}^n})$ from (3.20) and that

$$E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{C_{i,j}^n} \right] \leq R(\alpha, h_n^{2\rho(2+\nu)}, X_{t_{i-1}^n}) P_{i-1}^n \{ C_{i,j}^n \}$$

$$= R(\alpha, h_n^{2\rho(2+\nu)+2}, X_{t_{i-1}^n})$$

for $j = 1, 2$. Then we have

$$\begin{aligned} \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n}} \zeta_i^q(\alpha_0) \right|^{2+\nu} \right] &= \frac{1}{n^{\nu/2} h_n^{2+\nu}} O_p(h_n^{2\rho(2+\nu)+2}) + O_p\left(\frac{1}{n^{\nu/2}}\right) \\ &= O_p\left(\frac{h_n^\mu}{n^{\nu/2} h_n^{\nu/2}}\right) + O_p\left(\frac{1}{n^{\nu/2}}\right), \end{aligned}$$

where $\mu = 2\rho(2+\nu) - \nu/2 = 2(2+\nu) \left(\rho - \frac{\nu}{4(\nu+2)} \right)$. If we take $\nu > 0$ sufficiently small, then $\mu > 0$ since $2/(\gamma+1) \leq \rho < 1/2$. This completes the proof. \square

Chapter 4

Parametric estimation in infinite activity models

In Chapter 3, we treated finite activity models. The method proposed there worked if the Lévy density f satisfies $\sup_z |f(z)| < B$ for a constant $B > 0$; see Remark 3.1. However, we sometimes need a more general measure f admitting $\int_{\mathcal{E}} f(z) dz = \infty$ in some applications. In this chapter, we consider the inference for such infinite activity models from sampled data. The essential idea is the same as in Chapter 3; classifying the increments of the data. We construct a single estimating function having two parts: the continuous part is constructed by the *small* increments in the same way as in the previous chapter, and the jump part is a moment-type estimating function constructed by the *large* increments. The meaning of *small* and *large* are important, and we will make them be clear in Section 4.2.

4.1 Models and assumptions

On a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$, we consider a d -dimensional stochastic process $X = (X_t)_{t \geq 0}$ which is a solution to the following stochastic differential equation:

$$\begin{cases} dX_t = a(X_t, \theta) dt + b(X_t, \sigma) dw_t + \int_{\mathbb{R}^k} c(X_{t-}, z, \theta) r_\theta(dt, dz), \\ X_0 = x_0, \end{cases} \quad (4.1)$$

where x_0 is a random variable on \mathbb{R}^d , θ and σ are parameters, and their parameter spaces Θ and Π are compact convex subsets of \mathbb{R}^{m_1} and \mathbb{R}^{m_2} respectively, the coefficient a, b and c are known functions such that $a : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$, $b : \mathbb{R}^d \times \Pi \rightarrow \mathbb{R}^d \otimes \mathbb{R}^r$ and

$c : \mathbb{R}^d \times \mathbb{R}^k \times \Theta \rightarrow \mathbb{R}^d$. In this chapter, we admit the case where $k \neq d$. Moreover $(w_t)_{t \geq 0}$ is an r -dimensional Wiener process, $r_\theta(dt, dz) := p(dt, dz) - q_\theta(dt, dz)$ is a compensated Poisson random measure, that is, p is a time-homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}^k$, and q_θ is its intensity measure of the form $q(dt, dz) = f_\theta(z) dz dt$, where $\int_{\mathbb{R}^k} |z|^2 \wedge 1 f_\theta(z) dz < \infty$ for all θ . We set $\alpha := (\theta, \sigma) \in \Xi := \Theta \times \Pi$ and $m := m_1 + m_2$.

Our interest is to estimate the parameters θ and σ jointly from sampled data $\{X_{t_i^n}\}_{i=1}^n$, where $t_i^n = ih_n$ under the condition that $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$.

We make the following assumptions A1-A5.

A 1 *There exists a constant $L > 0$ and a function $\zeta(z)$ which satisfies $|\zeta(z)|\mathbf{1}_{\{|z| \leq 1\}} \leq C|z|$ and $|\zeta(z)| \leq C(1 + |z|)^C$ for a constant $C > 0$ such that*

$$\begin{aligned} |a(x, \theta_0) - a(y, \theta_0)| + |b(x, \sigma_0) - b(y, \sigma_0)| &\leq L|x - y|, \\ |c(x, z, \theta_0) - c(y, z, \theta_0)| &\leq \zeta(z)|x - y|, \quad |c(x, z, \theta_0)| \leq \zeta(z)(1 + |x|). \end{aligned}$$

A 2 *The process X is ergodic and stationary for $\alpha = \alpha_0$ with an invariant measure π in the sense of Section 2.4, and π satisfies that*

$$\int |x|^p \pi(dx) < \infty \quad (4.2)$$

for any $p \geq 0$.

A 3 *β is positive definite and $\inf_{x, \sigma} \det \beta(x, \sigma) > 0$.*

A 4 *For $l = 0, 1, 2$ and $v = x, \alpha$, there exist the partial derivatives $\partial_v^l a(x, \theta)$ and $\partial_v^l b(x, \sigma)$ such that $|\partial_v^l a(x, \theta)| + |\partial_v^l b(x, \sigma)| \leq C(1 + |x|)^C$, and they are continuous for fixed α . Moreover, there exist the partial derivatives $\partial_v^l \int_{\mathbb{R}^k} \prod_{j=1}^p c^{(l_j)}(x, z, \theta) f_\theta(z) dz$ and $\partial_v^l c(x, z, \theta)$ for any $p \in \mathbb{N}$ such that*

$$|\partial_v^l c(x, z, \theta)| + \left| \partial_v^l \int_{\mathbb{R}^k} \prod_{j=1}^p \{c^{(l_j)}(x, z, \theta)\} f_\theta(z) dz \right| \leq \zeta(z)(1 + |x|)^C,$$

where $l_j = 1, \dots, d$ and we allow the case where $l_j = l_i$ for $i \neq j$.

As well as in Chapter 3, we shall identify p with a random counting measure associated with a Lévy process z with the Lévy density $f_\theta(z)$, that is,

$$p(dt, dz) = \sum_{s \geq 0} \mathbf{1}_{\{|\Delta z_s| > 0\}} \mathbf{1}_{(s, \Delta z_s)}(dt, dz),$$

where the process z is independent of a Wiener process w .

In estimation problems of Lévy measures, it is crucial that what kinds of singularities Lévy measures have. In Chapter 3, we studied the case where the Lévy density decreased around the origin in the order of some polynomial and had the finite total mass. In this chapter, however, we would like to treat more general models, that is, $\lambda_\theta = \int_{\mathbb{R}^k} f_\theta(z) dz$ may be possibly finite or possibly infinite. If λ_θ is finite then z is a compound Poisson process with a distribution of jumps $(\lambda_\theta)^{-1} f_\theta(z) dz$. This is a finite activity model dealt with in Chapter 3. If λ_θ is infinite then z is a Lévy process which has infinitely many jumps even in any finite time interval. This is an infinite activity model. In the latter model, however, the process $z_t^{(\varepsilon_n)} = \sum_{s \leq t} \Delta z_s \mathbf{1}_{\{|\Delta z_s| > \varepsilon_n\}}$ (This is always a finite sum for fixed t and fixed n .) for some $\varepsilon_n > 0$ is a compound Poisson process with a distribution of jumps $(\lambda_\theta^{(\varepsilon_n)})^{-1} f_\theta^{(\varepsilon_n)}(z) dz$, where $\lambda_\theta^{(\varepsilon_n)} = \int f_\theta^{(\varepsilon_n)}(z) dz$ and $f_\theta^{(\varepsilon_n)} = f_\theta \mathbf{1}_{\{|z| > \varepsilon_n\}}$; see Chapter 2. Thanks to this fact, it is convenient to split the jump term of X into two parts by a sequence $\varepsilon_n \downarrow 0$ as follows:

$$\int_{\mathbb{R}^k} c(X_{t-}, z, \theta) r_\theta(dt, dz) =: dB_n^\theta(t) + dJ_n^\theta(t),$$

and

$$\begin{aligned} dB_n^\theta(t) &:= \int_{|z| \leq \varepsilon_n} c(X_{t-}, z, \theta) r_\theta(dt, dz), \\ dJ_n^\theta(t) &:= \int_{\mathbb{R}^k} c(X_{t-}, z, \theta) \{p^{(\varepsilon_n)}(dt, dz) - q_\theta^{(\varepsilon_n)}(dt, dz)\}, \end{aligned}$$

where $p^{(\varepsilon_n)}$ is a random measure generated by $z^{(\varepsilon_n)}$ and $q_\theta^{(\varepsilon_n)}$ is its intensity measure, that is, $q_\theta^{(\varepsilon_n)}(dt, dz) = f_\theta^{(\varepsilon_n)}(z) dz dt$. When n is sufficiently large, so ε_n is sufficiently small, then it may be possible to regard $B_n^\theta(t)$ as a small diffusion approximately.

Using these notations, and assuming that $\int_{\mathbb{R}^k} c(x, z, \theta) f_\theta^{(\varepsilon_n)}(z) dz < \infty$ for each n , we can rewrite the first model (4.1) in the form

$$dX_t = a_n(X_t, \theta) dt + b(X_t, \sigma) dw_t + dB_n^\theta(t) + \int_{\mathbb{R}^k} c(X_{t-}, z, \theta) p^{(\varepsilon_n)}(dt, dz), \quad (4.3)$$

where $a_n(x, \theta) = a(x, \theta) - \int_{\mathbb{R}^k} c(x, z, \theta) f_\theta^{(\varepsilon_n)}(z) dz$.

To discuss the asymptotic theory, we have to suppose the existence of the limit of a_n , that is, $a_n(x, \theta) \rightarrow \tilde{a}(x, \theta) := a(x, \theta) - \int_{\mathbb{R}^k} c(x, z, \theta) f_\theta(z) dz$ should exist. In order to ensure these conditions, we will assume the following assumption.

A 5 For any $x \in \mathbb{R}^d$, $\theta \in \Theta$ and any $p \geq 1$,

$$\int_{\mathbb{R}^k} |z|^p \partial_\theta^v f_\theta(z) dz < \infty,$$

where $v = 0, 1$.

It seems that this assumption is too restrictive, but this can be satisfied in a very wide class of Lévy measures. Actually, many practical Lévy measures in applications such as gamma, inverse Gaussian, variance gamma, normal inverse Gaussian or some generalized tempered stable processes satisfy that

$$|\partial_\theta^v f_\theta(z)| \leq L|z|^{-\alpha} e^{-C|z|}$$

for $\alpha < 2$ and some positive constants L and C , although stable processes with Lévy measures of the form $f_\theta(z) = \gamma|z|^{-(\alpha+1)}$ ($0 < \alpha < 2$) do not satisfy these assumptions.

Under Assumption A5, the above $a_n(x, \theta)$ and $\tilde{a}(x, \theta)$ are well-defined since the integrals $\int_{\mathbb{R}^k} c(x, z, \theta) g_\theta^{(n)}(z) dz$ exist for $g_\theta^{(n)} = f_\theta$ or $f_\theta^{(\varepsilon_n)}$.

Definition 4.1 *Under Conditions A1 and A5, we put*

$$a_n(x, \theta) := a(x, \theta) - \int_{\mathbb{R}^k} c(x, z, \theta) f_\theta^{(\varepsilon_n)}(z) dz, \quad (4.4)$$

$$\tilde{a}(x, \theta) := a(x, \theta) - \int_{\mathbb{R}^k} c(x, z, \theta) f_\theta(z) dz, \quad (4.5)$$

where $f_\theta^{(\varepsilon_n)}(z) = f_\theta(z) \mathbf{1}_{\{|z| > \varepsilon_n\}}$.

4.2 Selection of data

As we have described in the previous section, we deal with the process $B_n^\theta(t)$ in (4.3) as if the *small* Brownian shocks. Therefore, we can regard the model (4.3) as a finite activity model, then the same idea as in Chapter 3 may be applied to this model, that is, if $|\Delta_i X^n| \leq h_n^\rho$, then we can judge that the jumps by $\int_0^t \int c(X_{s-}, z, \theta) p^{(\varepsilon_n)}(ds, dz)$ did not occur. The next lemma justifies this treatment.

Let $\tau_i^n(\varepsilon_n)$ be stopping times defined by $\tau_i^n(\varepsilon_n) := \inf \{t \in [t_{i-1}^n, t_i^n]; |\Delta z_t| > \varepsilon_n\}$ for fixed n , where z is a Lévy process which generates the random measure p as already mentioned. Then the following equality is valid.

Lemma 4.1 *Let $\rho \in (0, 1/2)$ and $\varepsilon_n = h_n^{\rho'}$ with $\rho < \rho'$. Then it follows for any $p \geq 1$ and $i = 1, \dots, n$ that*

$$P_{i-1}^n \left\{ \sup_{t \in [t_{i-1}^n, \tau_i^n(\varepsilon_n))} |X_t - X_{t_{i-1}^n}| > h_n^\rho \right\} = R(\alpha, h_n^\rho, X_{t_{i-1}^n}). \quad (4.6)$$

This lemma implies that if there does not exist any jump with $|\Delta z_t| > \varepsilon_n$ in the interval $[t_{i-1}^n, t_i^n)$ then the increment of the path of X in that interval is smaller than h_n^ρ with a large probability. Let us prove this lemma.

Proof . Let $\|Y\|_{i,n} := \sup_{u \in (t_{i-1}^n, t_i^n]} |Y_u|$ for a process Y . On the interval $[t_{i-1}^n, \tau_i^n(\varepsilon_n))$, the solution process X to (4.1) satisfies a stochastic differential equation $dX_t = a_n(X_t) dt + b(X_t) dw_t + dB_n(t)$. Therefore, it follows from the Lipschitz continuity of a_n and Gronwall's inequality that, for sufficiently large n ,

$$|X_t - X_{t_{i-1}^n}| \leq C \left(h_n |a_n(X_{t_{i-1}^n})| + \left\| \int_{t_{i-1}^n}^{\cdot} b(X_{s-}) dw_s \right\|_{i,n} + \|B_n(\cdot) - B_n(t_{i-1}^n)\|_{i,n} \right).$$

We will only show $Q_i^n := P_{i-1}^n \{C\|B_n(\cdot) - B_n(t_{i-1}^n)\|_{i,n} > h_n^\rho\} = R(\alpha, h_n^p, X_{t_{i-1}^n})$ since

$$\begin{aligned} P_{i-1}^n \left\{ Ch_n |a_n(X_{t_{i-1}^n})| > h_n^\rho \right\} &= R(\alpha, h_n^p, X_{t_{i-1}^n}) \\ P_{i-1}^n \left\{ \left\| \int_{t_{i-1}^n}^{\cdot} b(X_{s-}) dw_s \right\|_{i,n} > h_n^\rho \right\} &= R(\alpha, h_n^p, X_{t_{i-1}^n}) \end{aligned}$$

were shown in Lemma 3.1 in Chapter 3.

According to Markov's inequality and Lemma 4.2, for $p' = 2^q$ ($q \in \mathbb{N}$),

$$\begin{aligned} Q_i^n &\leq Ch_n^{-p'\rho} E_{i-1}^n \left[\|B_n(\cdot) - B_n(t_{i-1}^n)\|_{i,n}^{p'} \right] \\ &\leq Ch_n^{-p'\rho} E_{i-1}^n \left[\sup_{u \in (t_{i-1}^n, t_i^n]} \left| \int_{t_{i-1}^n}^u \int_{|z| \leq \varepsilon_n} c(X_{s-}, z) r(ds, dz) \right|^{p'} \right] \\ &\leq Ch_n^{-p'\rho} E_{i-1}^n \left[\sup_{u \in (t_{i-1}^n, t_i^n]} \int_{t_{i-1}^n}^u \int_{|z| \leq \varepsilon_n} |c(X_{s-}, z)|^{p'} q(ds, dz) \right]. \end{aligned}$$

Noticing that $|c(x, z, \theta)| \leq \zeta(z)(1 + |x|)$ and $|\zeta(z)| \mathbf{1}_{\{|z| \leq 1\}} \leq C|z|$, we have

$$\begin{aligned} Q_i^n &\leq Ch_n^{-p'\rho} E_{i-1}^n \left[\sup_{u \in (t_{i-1}^n, t_i^n]} \int_{t_{i-1}^n}^u \int_{|z| \leq \varepsilon_n} |z|^{p'-2} |z|^2 (1 + |X_{s-}|)^{p'} q(ds, dz) \right] \\ &= R(\alpha, h_n^{p'(\rho'-\rho)+1-2\rho'}, X_{t_{i-1}^n}). \end{aligned}$$

Since we can take p' arbitrary large, $Q_i^n = R(\alpha, h_n^p, X_{t_{i-1}^n})$ for any $p \geq 1$ when $\rho' > \rho$. This completes the proof. \square

Lemma 4.2 For $p = 2^q$, $q \in \mathbb{N}$, $t_{i-1}^n \leq t \leq t_i^n$, and any Borel subset B of \mathbb{R}^k ,

$$E_{i-1}^n \left[\left| \int_B \int_{t_{i-1}^n}^t c(X_{s-}, z) r(ds, dz) \right|^p \right] \leq C_p E_{i-1}^n \left[\int_B \int_{t_{i-1}^n}^t |c(X_{s-}, z)|^p q(ds, dz) \right]. \quad (4.7)$$

Proof . This is the direct result from Lemma 3.4, in which we replace $c(X_{s-}, z)$ with $c(X_{s-}, z) \mathbf{1}_B(z)$ for any Borel set B of \mathbb{R}^k . \square

Lemma 4.3 Suppose that the process \tilde{X} satisfies the stochastic differential equation

$$d\tilde{X}_t = a_n(\tilde{X}_t) dt + b(\tilde{X}_t) dw_t + d\tilde{B}_n(t),$$

where $\tilde{B}_n(t) = \int_0^t \int_{|z| \leq \varepsilon_n} c(\tilde{X}_{s-}, z) r(ds, dz)$. Then, for $k \geq 2$, $k \in \mathbb{N}$ and $t \in [t_{i-1}^n, t_i^n)$,

$$E_{i-1}^n \left[|\tilde{X}_t - \tilde{X}_{t_{i-1}^n}^n|^k \right] \leq C_k |t - t_{i-1}^n|^{k/2 \wedge \{1 + \rho'(k-2)\}} (1 + |\tilde{X}_{t_{i-1}^n}^n|)^k. \quad (4.8)$$

Proof . We suppose that $k = 2^q$ ($q \in \mathbb{N}$). Applying Jensen's inequality, Burkholder-Davis-Gundy's inequality and Lemma 4.2, we see that

$$\begin{aligned} & E_{i-1}^n \left[|\tilde{X}_t - \tilde{X}_{t_{i-1}^n}^n|^k \right] \\ & \leq C_k E_{i-1}^n \left[\left| \int_{t_{i-1}^n}^t a_n(\tilde{X}_s) ds \right|^k + \left| \int_{t_{i-1}^n}^t b(\tilde{X}_{s-}) dw_s \right|^k \right. \\ & \quad \left. + \left| \int_{t_{i-1}^n}^t \int_{|z| \leq \varepsilon_n} c(\tilde{X}_{s-}, z) r(ds, dz) \right|^k \right] \\ & \leq C_k \left\{ h_n^{k-1} \int_{t_{i-1}^n}^t E_{i-1}^n \left[|a_n(\tilde{X}_s)|^k \right] ds + \left| \int_{t_{i-1}^n}^t E_{i-1}^n \left[|b(\tilde{X}_{s-})|^2 \right] ds \right|^{k/2} \right. \\ & \quad \left. + E_{i-1}^n \left[\left| \int_{t_{i-1}^n}^t \int_{|z| \leq \varepsilon_n} |c(\tilde{X}_{s-}, z)|^2 p^{(\varepsilon_n)}(ds, dz) \right|^{k/2} \right] ds \right\} \\ & \leq C_k \left\{ h_n^{k-1} \int_{t_{i-1}^n}^t E_{i-1}^n \left[|a_n(\tilde{X}_s)|^k \right] ds + h_n^{k/2-1} \int_{t_{i-1}^n}^t E_{i-1}^n \left[|b(\tilde{X}_s)|^k \right] ds \right. \\ & \quad \left. + \varepsilon_n^{(k-2)} \int_{t_{i-1}^n}^t \int_{|z| \leq \varepsilon_n} E_{i-1}^n \left[|c(\tilde{X}_{s-}, z)|^2 \right] q^{(\varepsilon_n)}(ds, dz) \right\}. \end{aligned}$$

Noticing that, on the set $\{|z| \leq \varepsilon_n\}$,

$$|c(\tilde{X}_{s-}, z)|^2 \leq \zeta^2(z)(1 + |\tilde{X}_{s-}|)^2 \leq C|z|^2(1 + |\tilde{X}_{s-}|)^k$$

since $k \geq 2$, and using the linear growthness of a and b , we obtain that

$$\begin{aligned} E_{i-1}^n \left[|\tilde{X}_t - \tilde{X}_{t_{i-1}^n}^n|^k \right] & \leq C_k \left\{ |t - t_{i-1}^n|^{k/2} + |t - t_{i-1}^n|^{1+\rho'(k-2)} \right\} (1 + |\tilde{X}_{t_{i-1}^n}^n|)^k \\ & \quad + C_k \int_{t_{i-1}^n}^t E_{i-1}^n \left[|\tilde{X}_s - \tilde{X}_{t_{i-1}^n}^n|^k \right] ds. \end{aligned}$$

Gronwall's inequality leads the consequence for $k = 2^q$.

It is easy to extend the above result to the case of arbitrary $k \geq 2$ by using the binary expansion of k and Cauchy-Schwarz's inequality repeatedly. \square

Corollary 4.1 *Let $\rho \in (0, 1/2)$ and $\varepsilon_n = h_n^{\rho'}$ with $\rho < \rho'$. Under the same assumptions as in Lemma 4.3,*

$$E \left[\sup_{s, t \in [t_{i-1}^n, t_i^n]} |\tilde{X}_t - \tilde{X}_s|^k \right] = O \left(h_n^{k/2 \wedge \{1 + \rho'(k-2)\}} \right). \quad (4.9)$$

Proof . By the same argument as in the proof of Lemma 4.3, it follows that

$$\begin{aligned} & E \left[\sup_{s, t \in [t_{i-1}^n, t_i^n]} |\tilde{X}_t - \tilde{X}_s|^k \right] \\ & \leq C_k \left\{ h_n^{k-1} \int_{t_{i-1}^n}^{t_i^n} E \left[|a_n(\tilde{X}_s)|^k \right] ds + h_n^{k/2-1} \int_{t_{i-1}^n}^{t_i^n} E \left[|b(\tilde{X}_s)|^k \right] ds \right. \\ & \quad \left. + \varepsilon_n^{(k-2)} \int_{t_{i-1}^n}^{t_i^n} \int_{|z| \leq \varepsilon_n} E \left[|c(\tilde{X}_{s-}, z)|^2 \right] q^{(\varepsilon_n)}(ds, dz) \right\} \\ & = O \left(h_n^{k/2 \wedge \{1 + \rho'(k-2)\}} \right). \quad \square \end{aligned}$$

Let $J_i^n(\varepsilon_n)$ be the number of *large* jumps of z ; $|\Delta z| > \varepsilon_n$ in the interval $[t_{i-1}^n, t_i^n]$, and we set $\{|\Delta_i X^n| \leq h_n^\rho\} = C_{i,0}^n(\varepsilon_n) \cup C_{i,1}^n(\varepsilon_n)$ and $\{|\Delta_i X^n| > h_n^\rho\} = D_{i,0}^n(\varepsilon_n) \cup D_{i,1}^n(\varepsilon_n)$, where

$$\begin{aligned} C_{i,0}^n(\varepsilon_n) &= \{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) = 0\}, \\ C_{i,1}^n(\varepsilon_n) &= \{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) \geq 1\}, \\ D_{i,0}^n(\varepsilon_n) &= \{|\Delta_i X^n| > h_n^\rho, J_i^n(\varepsilon_n) = 0\}, \\ D_{i,1}^n(\varepsilon_n) &= \{|\Delta_i X^n| > h_n^\rho, J_i^n(\varepsilon_n) \geq 1\}. \end{aligned}$$

Lemma 4.1 immediately yields the next lemma.

Lemma 4.4 *Let $\rho \in (0, 1/2)$ and $\varepsilon_n = h_n^{\rho'}$ with $\rho < \rho'$. Assume that $\lambda_0^{(\varepsilon_n)} h_n = O(1)$ as $n \rightarrow \infty$. Then, for any $p \geq 1$,*

$$\begin{aligned} P_{i-1}^n \{C_{i,0}^n(\varepsilon_n)\} &= e^{-\lambda_0^{(\varepsilon_n)} h_n} R(\alpha, 1, X_{t_{i-1}^n}), \\ P_{i-1}^n \{D_{i,0}^n(\varepsilon_n)\} &= e^{-\lambda_0^{(\varepsilon_n)} h_n} R(\alpha, h_n^p, X_{t_{i-1}^n}), \\ P_{i-1}^n \{C_{i,1}^n(\varepsilon_n)\} &= \lambda_0^{(\varepsilon_n)} h_n e^{-\lambda_0^{(\varepsilon_n)} h_n} R(\alpha, 1, X_{t_{i-1}^n}), \\ P_{i-1}^n \{D_{i,1}^n(\varepsilon_n)\} &= \lambda_0^{(\varepsilon_n)} h_n e^{-\lambda_0^{(\varepsilon_n)} h_n} R(\alpha, 1, X_{t_{i-1}^n}), \end{aligned}$$

where $\lambda_0^{(\varepsilon_n)} = \int_{|z| > \varepsilon_n} f_{\theta_0}(z) dz$.

Proof . It is obvious that $P_{i-1}^n\{D_{i,0}^n(\varepsilon_n)\} = e^{-\lambda_0^{(\varepsilon_n)}h_n}R(\alpha, h_n^p, X_{t_{i-1}^n})$ by Lemma 4.1 since $P_{i-1}^n\{J_i^n(\varepsilon_n) = 0\} = e^{-\lambda_0^{(\varepsilon_n)}h_n}$. Hence $P_{i-1}^n\{C_{i,0}^n(\varepsilon_n)\} = e^{-\lambda_0^{(\varepsilon_n)}h_n}R(\alpha, 1, X_{t_{i-1}^n})$ since $P_{i-1}^n\{C_{i,0}^n(\varepsilon_n)\} = P_{i-1}^n\{J_i^n(\varepsilon_n) = 0\} - P_{i-1}^n\{D_{i,0}^n(\varepsilon_n)\}$. The other proofs are easy since $P_{i-1}^n\{J_i^n(\varepsilon_n) = 1\} = \lambda_0^{(\varepsilon_n)}h_n e^{-\lambda_0^{(\varepsilon_n)}h_n}$. \square

In Chapter 3, we decomposed the events $\{|\Delta_i X^n| \leq h_n^\rho\}$ and $\{|\Delta_i X^n| > h_n^\rho\}$ into three events, that is, for no jump, for a single jump and for more than two jumps. However, it is difficult to discriminate between the event $\{|\Delta_i X^n| > h_n^\rho, J_i^n(\varepsilon_n) = 1\}$ and the event $\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) = 1\}$ in the case where $\int f_\theta(z) dz = \infty$. In the above lemma, we take ρ' with $\rho' > \rho$ in order to ignore the probability $P_{i-1}^n\{D_{i,0}^n(\varepsilon_n)\}$, otherwise, the probability $P_{i-1}^n\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) = 1\}$ can not be ignored. On the other hand, if we take ρ' with $\rho' \leq \rho$, then the probability $P_{i-1}^n\{D_{i,0}^n(\varepsilon_n)\}$ can not be ignored although the probability $P_{i-1}^n\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) = 1\}$ can be ignored. That is a tradeoff, so we can not identify the event of a single jump by such a filter, therefore we could not expect an efficient estimator for the jump part.

We assumed in Chapter 3 the condition of a nondegeneracy of the coefficient $c(x, z)$: $\inf_x |c(x, z, \theta_0)| > c_0|z|$ for a positive constant c_0 near the origin in order to discriminate the event of no jump and the event of a single jump. Of course, it is not an essential assumption but it seems much trouble without such an assumption when you compute an estimator explicitly. However, we do not need such an assumption in our method since we do not demand the efficiency for parameters in jump part.

Remark 4.1 One can easily find that it is possible to make the above filter $\{|\Delta_i X^n| \leq Lh_n^\rho\}$ for any constant $L > 0$. As a matter of fact, the value of L has sometimes a great influence on the performance of estimation as we show in Section 4.4. However, it is difficult to select the optimal L in practice in our setting where infinitely many jumps occur. We will discuss the selection problems of such a filter elsewhere. Throughout this chapter, we suppose that $L = 1$ although the following all results are valid for any $L > 0$.

4.3 Estimating functions and asymptotic results

4.3.1 Moment type estimating functions

In this section, we propose estimating functions. In the following discussion, we assume that $\rho \in (0, 1/2)$ and $\varepsilon_n = h_n^{\rho'}$ with $\rho < \rho'$, and we set $\bar{X}_{i,n}(\theta) = \Delta_i X^n - h_n a_{n,i-1}(\theta)$. Recall that $a_n(x, \theta) = a(x, \theta) - \int_{\mathbb{R}^k} c(x, z, \theta) f_\theta^{(\varepsilon_n)}(z) dz$ and $a_{n,i-1}(\theta) = a_n(X_{t_{i-1}^n}, \theta)$.

In Chapter 3, a natural contrast function of MLE-type was proposed under the condition that $\int_{\mathbb{R}^k} f_\theta(z) dz < \infty$. To estimate the drift and the diffusion parameters, they picked up the data $X_{t_{i-1}^n}$ and $\Delta_i X^n$ with $\{|\Delta_i X^n| \leq h_n^\rho\}$ and made use of a contrast of an usual diffusion process:

$$\frac{1}{2h_n} \sum_{i=1}^n (\bar{X}_{i,n})^*(\theta) \beta_{i-1}^{-1}(\sigma) \bar{X}_{i,n}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} + \sum_{i=1}^n \frac{1}{2} \log \det \beta_{i-1}(\sigma) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}. \quad (4.10)$$

It would be natural to make use of the derivatives with respect to parameters of this contrast function as an estimating function since we can judge again that there does not exist a *large* jump in an interval with $\{|\Delta_i X^n| \leq h_n^\rho\}$. On the other hand, we only infer that at least a single *large* jump occurred from the fact $\{|\Delta_i X^n| > h_n^\rho\}$ thanks to that $P_{i-1}^n\{D_{i,0}^n(\varepsilon_n)\} = R(\alpha, h_n^\rho, X_{t_{i-1}^n})$, and we can not identify the number of jumps of $\int_0^t \int_{\mathbb{R}^k} c(X_{s-}, z, \theta) p^{(\varepsilon_n)}(ds, dz)$. Therefore, it may be better to approximate the moments

$$E \left[\int_0^t \int_{\mathbb{R}^k} \prod_{j=1}^q c^{(l_j)}(X_{s-}, z, \theta) p^{(\varepsilon_n)}(ds, dz) \right] \quad (q \geq 1) \quad (4.11)$$

by the data $X_{t_{i-1}^n}$ and $\Delta_i X^n$ with $\{|\Delta_i X^n| > h_n^\rho\}$, where $l_j = 1, 2, \dots, d$ and we allow the case where $l_i = l_j$ for $i \neq j$.

We use the following notations to define an estimating function. Let $\{G_n^{(j)}(x, \alpha)\}_j$ be a sequence of \mathbb{R} -valued functions defined on $\mathbb{R}^d \times \Xi$ which satisfy the following conditions:

$$|G_n^{(j)}(x, \alpha)| \leq L(x, \alpha), \quad |\partial_\alpha G_n^{(j)}(x, \alpha)|, |\partial_x G_n^{(j)}(x, \alpha)| \leq C(1 + |x|)^C \quad (4.12)$$

for fixed j and all n , where $L^2(x, \alpha)$ is a π -integrable function for all α and there exists a function $G^{(j)}(x, \alpha)$ which is differentiable with respect to α such that for $k = 0, 1$,

$$\partial_\alpha^k G_n^{(j)}(x, \alpha) \rightarrow \partial_\alpha^k G^{(j)}(x, \alpha) \quad \pi\text{-a.s.} \quad (4.13)$$

Notice that $\partial_\alpha^k G_n^{(j)}(x, \alpha)$ is π -integrable for all α since $\{\partial_\alpha^k G_n^{(j)}\}_n$ is uniformly integrable. Moreover, let $H_{n,Q}(x, y)$ be \mathbb{R} -valued functions defined on $\mathbb{R}^d \times \mathbb{R}^d$ such that

$$H_{n,Q}(x, y) = \sum_{Q'=2}^Q G_n^{(Q')}(x, \alpha) \prod_{j=1}^{Q'} y^{(l_j)} \quad (4.14)$$

for an integer $Q \geq 2$, and we write $H_Q(x, y) = \sum_{Q'=2}^Q G^{(Q')}(x, \alpha) \prod_{j=1}^{Q'} y^{(l_j)}$.

Definition 4.2 We define the estimating function

$$\psi_n(\alpha) = \sum_{i=1}^n \left(\psi_{i,n}^{(1)}(\alpha) + \psi_{i,n}^{(2)}(\alpha), \psi_{i,n}^{(3)}(\alpha) \right)^*$$

as follows:

$$\psi_{i,n}^{(1)}(\alpha) = \left(\psi_{i,n,q}^{(1)} \right)_{1 \leq q \leq m_1}, \quad \psi_{i,n}^{(2)}(\alpha) = \left(\psi_{i,n,q}^{(2)} \right)_{1 \leq q \leq m_1}, \quad \psi_{i,n}^{(3)}(\alpha) = \left(\psi_{i,n,r}^{(3)} \right)_{1 \leq r \leq m_2}$$

and

$$\begin{aligned} \psi_{i,n,q}^{(1)}(\alpha) &= H_{n,Q_q} \left(X_{t_{i-1}^n}, \bar{X}_{i,n} \right) \mathbf{1}_{\{|\Delta_i X^n| > h_n^p\}} - h_n \int_{\mathbb{R}^k} H_{n,Q_q} (X_{t_{i-1}^n}, c_{i-1}(z, \theta)) f_\theta(z) dz, \\ \psi_{i,n,q}^{(2)}(\alpha) &= \partial_{\theta_q} a_{n,i-1}^*(\theta) \beta_{i-1}^{-1}(\sigma) \bar{X}_{i,n}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^p\}}, \\ \psi_{i,n,r}^{(3)}(\alpha) &= \left\{ \frac{1}{h_n} \bar{X}_{i,n}^*(\theta) \partial_{\sigma_r} \beta_{i-1}^{-1}(\sigma) \bar{X}_{i,n}(\theta) + \frac{\partial_{\sigma_r} \det \beta_{i-1}(\sigma)}{\det \beta_{i-1}(\sigma)} \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^p\}}, \end{aligned}$$

where $Q_q \geq 2$ and $\bar{X}_{i,n}(\theta) = \Delta_i X^n - h_n a_{n,i-1}(\theta)$. On l_j in (4.14), $l_j = 1, 2, \dots, d$ and we allow the case where $l_i = l_j$ for $i \neq j$.

The summation $\sum_{i=1}^n \psi_{i,n}^{(1)}$ corresponds to the discretization of a weighted sum for $q = 2, \dots, Q_q$ of the expectation (4.11). The summation $\sum_{i=1}^n \psi_{i,n}^{(2)}$ and $\sum_{i=1}^n \psi_{i,n}^{(3)}$ correspond to the first derivatives of (4.10) with respect to θ and σ respectively.

Definition 4.3 We define the estimator of α_0 as a solution $\hat{\alpha}_n$ to the equation

$$\psi_n(\alpha) = 0.$$

This implies that we should determine the function H_{n,Q_q} (so also Q_q) so that there exists a solution $\hat{\alpha}_n$.

In applications, it may be better to choose H_{n,Q_q} so that $G_n^{(j)}(x, \alpha) \equiv 0$ for $1 \leq j \leq Q_q - 1$ for simplicity of computations, where Q_q should be determined so that $\hat{\alpha}_n$ is well-defined (see Section 5 for examples). Generally, $G_n^{(j)}$'s play the roles of the weights to add the two functions $\psi^{(1)}$ and $\psi^{(2)}$, and they should be selected in an optimal way. However, it may not be so easy to find the optimal weights. That is a problem for the future.

This estimating function gives partially efficient estimators, that is, the asymptotic variances of the estimators for the parameters in the coefficient a are efficient in the sense that they attain the asymptotic variances of the estimators from continuous observations; see Remark 3.4 on this discussion, although the estimators for jump part

are not efficient in this sense. However, it may be actually convenient to use this estimating function because of the simplicity of both the selecting method of data and the forms of the estimating function.

In order to obtain the consistency result, we assume the following A6: the identifiability conditions. Below, we use the notation

$$U_q(x, \theta, \sigma) = \int \{H_{Q_q}(x, c(x, z, \theta_0))f_{\theta_0}(z) - H_{Q_q}(x, c(x, z, \theta))f_{\theta}(z)\} dz \\ + \partial_{\theta_q} \tilde{a}^*(x, \theta) \beta^{-1}(x, \sigma) \{\tilde{a}(x, \theta_0) - \tilde{a}(x, \theta)\}.$$

The integral $\int U_q(x, \theta, \sigma) d\pi$ will appear later as the limit of

$$\frac{1}{nh_n} \sum_{i=1}^n \left\{ \psi_{i,n,q}^{(1)} + \psi_{i,n,q}^{(2)} \right\} (\alpha).$$

A 6 For π -almost all $x \in \mathbb{R}^d$, $\sigma = \sigma_0$ if and only if $\beta(x, \sigma) = \beta(x, \sigma_0)$. For all $q = 1, \dots, m_1$ and π -almost all $x \in \mathbb{R}^d$, $\theta = \theta_0$ if and only if $U_q(x, \theta, \sigma_0) = U_q(x, \theta_0, \sigma_0)$.

4.3.2 Consistency and asymptotic normality

Now we present the result on the asymptotic behavior of $\hat{\alpha}_n$.

Theorem 4.1 Suppose Conditions A1 - A6, and that $h_n \rightarrow 0$, $nh_n \rightarrow \infty$, $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ as $n \rightarrow \infty$ with $\rho' > \rho$. Then the estimator $\hat{\alpha}_n$ has the consistency to the true α_0 :

$$\hat{\alpha}_n \xrightarrow{P} \alpha_0 \quad (n \rightarrow \infty).$$

What singularities of Lévy measures around the origin are admitted under the assumption $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ as $n \rightarrow \infty$? For example, let us consider the following one dimensional stochastic differential equation with one dimensional parameter $\theta^* \geq 0$:

$$dX_t = a(X_t) dt + b(X_t) dw_t + c(X_{t-}) dz_t^{\theta^*}, \quad (4.15)$$

and suppose that

$$f_{\theta^*}(z) = \tilde{f}_{\theta^*} \mathbf{1}_{\{|z| \leq 1\}} + \bar{f}_{\theta^*}(z) \mathbf{1}_{\{|z| > 1\}}, \quad |\tilde{f}_{\theta^*}(z)| \sim C|z|^{-\theta^*} \quad (z \rightarrow 0). \quad (4.16)$$

We call such a parameter θ^* the *shape parameter*. Originally, we need

$$\int_{|z| < \varepsilon_n} |z|^2 f_{\theta^*}(z) dz < \infty,$$

that is, $0 \leq \theta_0^* < 3$. Under this setting, $\lambda^{(\varepsilon_n)} h_n^\rho = O\left(h_n^{\rho'(1-\theta_0^*)+\rho}\right) + O(h_n^\rho)$ if $\theta_0^* \neq 1$, and $\lambda^{(\varepsilon_n)} h_n^\rho = O(h_n^\rho \log h_n) + O(h_n^\rho)$ if $\theta_0^* = 1$, hence we need that $0 \leq \theta_0^* < 1 + \rho/\rho'$. Therefore, the models with $0 \leq \theta_0^* < 2$ are allowed since we can take ρ' ($\rho' > \rho$) to be sufficiently near to ρ .

Theorem 4.2 *Suppose Conditions A1 - A6 and α_0 is in the interior of Ξ , and that $h_n \rightarrow 0$, $nh_n \rightarrow \infty$, $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ and $n(\lambda_0^{(\varepsilon_n)})^2 h_n^{4\rho} \rightarrow 0$ as $n \rightarrow \infty$. Moreover, we take $\rho'(> \rho)$ so that $\int_{|z| \leq \varepsilon_n} |z|^2 f_{\theta_0}(z) dz = o(n^{-1/2})$. Then*

$$M_n^{1/2}(\hat{\alpha}_n - \alpha_0) \xrightarrow{d} \mathcal{N}_m(0, K^{-1}),$$

where $M_n = \begin{pmatrix} nh_n I_{m_1} & 0 \\ 0 & n I_{m_2} \end{pmatrix}$, I_m is the m -dimensional identity matrix and $K = \begin{pmatrix} K_1 & 0 \\ 0 & K_2 \end{pmatrix}$ with $K_1^{-1} = (K_1')^{-1} K_1'' (K_1')^{-1}$,

$$\begin{aligned} (K_1')^{(p,q)} &= \int (\partial_{\theta_p} \tilde{a})^* \beta^{-1} (\partial_{\theta_q} \tilde{a})(x, \alpha_0) d\pi \\ &\quad - \sum_{Q'=2}^{Q_q} \iint \partial_{\theta_p} G^{(Q')}(x, \alpha_0) \prod_{j=1}^{Q'} c^{(l_j)}(x, z, \theta_0) f_{\theta_0}(z) dz d\pi \\ &\quad + \sum_{Q'=2}^{Q_q} \int \left\{ G^{(Q')}(x, \alpha_0) \partial_{\theta_p} \int_{\mathbb{R}^k} \prod_{j=1}^{Q'} c^{(l_j)}(x, z, \theta_0) f_{\theta_0}(z) dz \right\} d\pi, \\ (K_1'')^{(p,q)} &= \int (\partial_{\theta_p} \tilde{a})^* \beta^{-1} (\partial_{\theta_q} \tilde{a})(x, \alpha_0) d\pi \\ &\quad + \sum_{Q=2}^{Q_q} \sum_{Q'=2}^{Q'_q} \iint G^{(Q)} G^{(Q')}(x, \alpha_0) \left\{ \prod_{j=1}^{Q+Q'} c^{(l_j)}(x, z, \theta_0) \right\} f_{\theta_0}(z) dz d\pi, \\ (K_2)^{(p,q)} &= \frac{1}{2} \int \text{tr} [(\partial_{\sigma_p} \beta) \beta^{-1} (\partial_{\sigma_q} \beta) \beta^{-1}] (x, \sigma_0) d\pi. \end{aligned}$$

Remark 4.2 Suppose that X follows the SDE (4.15) with the Lévy measure (4.16). If $\theta_0^* \neq 1$,

$$n(\lambda_0^{(\varepsilon_n)})^2 h_n^{4\rho} = O\left(nh_n^{4\rho+2\rho'(1-\theta_0^*)}\right), \quad \int_{|z| \leq \varepsilon_n} z^2 f_{\theta_0^*}(z) dz = O\left(\sqrt{h_n^{2\rho'(3-\theta_0^*)}}\right).$$

Hence we admit $\theta_0^* \leq 1 + (4\rho - 1 - \nu)/2\rho'$ under the experimental design as $nh_n^{1+\nu} \rightarrow 0$ ($0 < \nu < 1$), that is, the model with $\theta^* < 2$ can be admitted for suitable ρ, ρ' and ν . Moreover, when $\theta_0^* = 1$, we need

$$n(\lambda_0^{(\varepsilon_n)})^2 h_n^{4\rho} = O(nh_n^{4\rho} (\log h_n)^2) \rightarrow 0,$$

and it is possible for a proper choice of ρ . Hence $\theta_0^* = 1$ is also admitted.

In the inference for infinite activity models, it is a big problem what the rate of convergence is. If the parameter θ in the Lévy characteristic has the finite information, that is, $\int \dot{f}_\theta^2 / f_\theta(z) dz < \infty$, where \dot{f}_θ stands for the first derivative with respect to θ , the maximum rate of convergence should be $\sqrt{nh_n}$ since, even in the continuously observed model up to time $T(\rightarrow \infty)$, the rate becomes \sqrt{T} which is identified with $\sqrt{nh_n}$; see Akritas and Johnson [2] for Lévy processes and Sørensen [99] for jump-diffusions. Our method attains the rate of convergence $\sqrt{nh_n}$, therefore it is sometimes a good rate. However the finiteness of the information is not so general, for example, if z is a stable process with the Lévy density $f_\theta(z) = \theta_1 |z|^{-1-\theta_2}$ ($\theta_2 \in (0, 2)$), then the information of $\theta = (\theta_1, \theta_2)$ becomes infinite: $\int \dot{f}_{\theta_i}^2 / f_{\theta_i}(z) dz = \infty$, and the maximum rate of convergence would not be $\sqrt{nh_n}$ any more in this case. Actually, for the inference of discretely observed Lévy processes, Woerner [113] showed the LAN results for the *scale parameter* (θ_1) of a stable process with the convergence rate \sqrt{n} under $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$. Hence the better rate than $\sqrt{nh_n}$ may be demanded. However, as Woerner [113] also showed, the maximal rates of convergence for the many important examples are $\sqrt{nh_n}$; see Section 3.1 in Woerner [113]. Therefore, it would not be so pessimistic for the rate $\sqrt{nh_n}$.

4.4 Examples and simulation study

We give some examples of X and show their simulation results.

Example 4.1 We consider the following one-dimensional SDE:

$$dX_t = (\theta_1 - \theta_2 X_t) dt + \sqrt{\sigma} dw_t + dz_t,$$

where $\theta_2 > 0$ and z_t is a compound Poisson process with $\int_{|z| > \varepsilon_n} z f(z) dz = 0$ for large n . As we have described in Example 2.2, this model is ergodic.

$$\begin{aligned} \psi_n(\alpha) &= (\psi_n^{(1)}(\alpha), \psi_n^{(2)}(\alpha), \psi_n^{(3)}(\alpha))^*, \\ \psi_n^{(1)}(\alpha) &= \sum_{i=1}^n \sigma^{-1} \left\{ \Delta_i X^n - h_n (\theta_1 - \theta_2 X_{t_{i-1}^n}) \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ \psi_n^{(2)}(\alpha) &= \sum_{i=1}^n X_{t_{i-1}^n} \sigma^{-1} \left\{ \Delta_i X^n - h_n (\theta_1 - \theta_2 X_{t_{i-1}^n}) \right\} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ \psi_n^{(3)}(\alpha) &= - \sum_{i=1}^n \bar{X}_{i,n}^2 (\theta_1, \theta_2) h_n^{-1} \sigma^{-2} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} + \sigma^{-1} \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \end{aligned}$$

where $\bar{X}_{i,n}(\theta_1, \theta_2) = \Delta_i X^n - (\theta_1 - \theta_2 X_{t_{i-1}^n}) h_n$.

Let

$$\begin{aligned} S_0 &= h_n \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ S_1 &= h_n \sum_{i=1}^n X_{t_{i-1}^n} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ S_2 &= h_n \sum_{i=1}^n X_{t_{i-1}^n}^2 \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ S_3 &= \sum_{i=1}^n \Delta_i X^n \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ S_4 &= \sum_{i=1}^n \Delta_i X^n X_{t_{i-1}^n} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}. \end{aligned}$$

Then we have the following estimating equation:

$$\theta_1 S_0 = S_3 + \theta_2 S_1, \quad \theta_1 S_1 = S_4 + \theta_2 S_2, \quad \sigma S_0 = \sum_{i=1}^n \bar{X}_{i,n}^2(\theta_1, \theta_2) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}.$$

Hence we have

$$\begin{aligned} \hat{\theta}_{1,n} &= \frac{S_2 S_3 - S_1 S_4}{S_1^2 - S_0 S_2}, \quad \hat{\theta}_{2,n} = \frac{S_1 S_3 - S_0 S_4}{S_1^2 - S_0 S_2}, \\ \hat{\sigma}_n &= S_0^{-1} \sum_{i=1}^n \bar{X}_{i,n}^2(\hat{\theta}_{1,n}, \hat{\theta}_{2,n}) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}. \end{aligned}$$

Example 4.2 We consider the following one-dimensional SDE:

$$dX_t = -\mu X_t dt + \sqrt{\sigma} dw_t + dz_t^\theta,$$

where $\mu > 0$, z_t^θ is a compound Poisson process with the Lévy density

$$f_\theta(z) = \frac{\lambda}{\sqrt{2\pi\nu}} \exp\left(-\frac{z^2}{2\nu}\right)$$

and $\theta = (\mu, \nu, \lambda)$. We set $\alpha = (\theta, \sigma)$. First, we suppose that either λ_0 or ν_0 is known. Then one of the simplest estimating functions is as follows:

$$\begin{aligned} \psi_n(\alpha) &= (\psi_n^{(j)}(\alpha))_{1 \leq j \leq 3}, \\ \psi_n^{(1)}(\alpha) &= \sum_{i=1}^n \bar{X}_{i,n}^2(\mu) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} - nh_n \int z^2 f_\theta(z) dz, \end{aligned}$$

$$\begin{aligned}\psi_n^{(2)}(\alpha) &= \sum_{i=1}^n -X_{t_{i-1}^n} \sigma^{-1} \bar{X}_{i,n}(\mu) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\ \psi_n^{(3)}(\alpha) &= -\sum_{i=1}^n \bar{X}_{i,n}^2(\mu) h_n^{-1} \sigma^{-2} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} + \sigma^{-1} \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}},\end{aligned}$$

where $\bar{X}_{i,n}(\theta) = \Delta_i X^n + \mu X_{t_{i-1}^n} h_n$ since $\int_{|z| > \varepsilon_n} z f_\theta(z) dz = 0$, and all the conditions for h_n^ρ and ρ' in Theorem 4.2 are satisfied if $nh_n^{4\rho} \rightarrow 0$ for any ρ' with $\rho < \rho'$. This estimating functions satisfy the identifiability condition A6.

The estimators of μ and σ are similarly obtained as in Example 4.1, and

$$\hat{\mu}_n = -\frac{S_4}{S_2}, \quad \hat{\sigma}_n = S_0^{-1} \sum_{i=1}^n \bar{X}_{i,n}^2(\hat{\mu}_n) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}. \quad (4.17)$$

Moreover, noticing that $\int z^{2k} f_\theta(z) dz = \frac{(2k)!}{2^k k!} \lambda \nu^k$, the estimator of ν or λ becomes

$$\hat{\nu}_n^{(0)} = \lambda_0^{-1} T_2(\hat{\mu}_n), \quad \hat{\lambda}_n^{(0)} = \nu_0^{-1} T_2(\hat{\mu}_n), \quad (4.18)$$

where $T_k(\mu) = \frac{1}{nh_n} \sum_{i=1}^n \bar{X}_{i,n}^k(\mu) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}$.

When we estimate (λ, ν) jointly, we should add, for example, an estimating function

$$\psi_n^{(4)}(\alpha) = \sum_{i=1}^n \bar{X}_{i,n}^4(\mu) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} - nh_n \int z^4 f_\theta(z) dz. \quad (4.19)$$

Then, the estimators become

$$\hat{\nu}_n = \frac{T_4(\hat{\mu}_n)}{3T_2(\hat{\mu}_n)}, \quad \hat{\lambda}_n = \frac{3T_2^2(\hat{\mu}_n)}{T_4(\hat{\mu}_n)}. \quad (4.20)$$

Example 4.3 We consider the same SDE as in the above Example 4.2, but z_t^θ is a two-sided gamma process with $\theta = (\alpha, \beta)$, where α is the shape parameter and β is the scale parameter, that is, the Lévy measure of the form $f_\theta(z) = 2^{-1} \alpha e^{-\beta|z|} |z|^{-1}$:

$$dX_t = -\mu X_t dt + \sqrt{\sigma} dw_t + dz_t^{(\alpha, \beta)}.$$

This is the case where infinitely many jumps occur in any finite time interval since

$$\int_{\mathbb{R}^k} f_\theta(z) dz = \infty.$$

Again, we suppose that either α_0 or β_0 is known. The estimators for (μ, σ) are the same as in Example 4.2 (4.17). The estimating functions for (α, β) are also the same as in Example 4.2, but $\int_{\mathbb{R}^k} z^2 f_\theta(z) dz = \alpha \beta^{-2}$. Therefore

$$\hat{\alpha}_n^{(0)} = \beta_0^2 T_2(\hat{\mu}_n), \quad \hat{\beta}_n^{(0)} = \sqrt{\frac{\alpha_0}{T_2(\hat{\mu}_n)}}. \quad (4.21)$$

The conditions for h_n^ρ and ρ' in Theorem 4.2 are satisfied if $nh_n^{4\rho}(\log h_n)^2 \rightarrow 0$. For example, if $h_n = n^{-\delta}$ and $nh_n^{1+\nu} \rightarrow 0$, then the last condition is satisfied for $\delta \in ((1+\nu)^{-1} \vee (4\rho)^{-1}, 1)$ and any ρ' with $\rho < \rho'$. Notice that $\int z^4 f_\theta(z) dz = 6\alpha\beta^{-4}$. When we estimate (α, β) jointly, we obtain the following estimators by adding (4.19) to the estimating functions in Example 4.2:

$$\hat{\alpha}_n = \frac{6T_2^2(\hat{\mu}_n)}{T_4(\hat{\mu}_n)}, \quad \hat{\beta}_n = \sqrt{\frac{6T_2(\hat{\mu}_n)}{T_4(\hat{\mu}_n)}}. \quad (4.22)$$

Now, let us show some simulation results for the above examples. In order to simulate the discrete sampling from the continuous path, we picked up the each observation every 100 sample points that were generated by the Euler-Maruyama discretization scheme in each h_n -interval. Each experiment was repeated 3000 times and the estimators were averaged out through those experiments. The number of simulated observations n is 500, 1000, or 3000.

First, we consider the model in Example 4.2 with the true values $(\mu_0, \lambda_0, \nu_0, \sigma_0) = (0.1, 0.3, 1.0, 0.05)$. A sample path of this model is shown in Fig. 4.1. Intuitively speaking, this model may be comparatively easy to judge whether a jump occurred or did not. Because, if a jump occurred then it should be large compared with the increments by diffusions only. However, we can obtain only a few samples to estimate the jump part since the intensity λ_0 is small, so it may not be so good for accuracy of estimation.

Below, we show the result of the case where we use (4.17) and (4.18) as the estimators. When we used (4.20) instead of (4.18), the result sometimes showed a bad performance because of a large variance by using the higher order moment T_2^2 and T_4 . We denote by J_n the mean of $\sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}$ in 3000 experiments. This is the mean of the number of samples which is used for estimating the parameters of the jump part. Moreover, we choose $h_n = n^{-0.6}$ and $\rho = 0.49$, which satisfy the desired conditions $nh_n \rightarrow \infty$ and $nh_n^{4\rho} \rightarrow 0$.

The result of this simulation is shown in Tab. 4.1. The mean of the number of jumps are expected as $nh_n\lambda_0$, so J_n should be fewer than $nh_n\lambda_0$ on average since *small* jumps are cut. However, it seems that the filter misunderstood a few *large* increments by the diffusion part for the true jumps in this example. On the other hand, the parameters in continuous part are estimated relatively good. A few misunderstandings as above do not influence on estimating the parameters in this part.

Second, we consider the same model as above but the parameters are different. We choose the true values as $(\mu_0, \lambda_0, \nu_0, \sigma_0) = (0.1, 5.0, 0.25, 0.2)$. It may be more difficult

to judge between the increments by Brownian shocks and the real discontinuity of the path than the preceding example, however the samples to estimate the jump part can be obtained more than the preceding one since λ_0 is large. The sample path is shown in Fig. 4.2. We choose the same h_n and ρ as the preceding one. The result is in Tab. 4.2. In this simulation, the number of jumps is overestimated much more than the previous one, so the diffusion parameter σ is underestimated. To avoid such a trouble, we need to choose L for a threshold Lh_n^ρ appropriately. Just for your information, we will show the result with $L = 1.5$ in the last example. The result is shown in Tab. 4.3. The estimation of σ, λ and ν are improved although the parameters in the jump part are still overestimated. This may be because the number of samples whose mean is J_n is too small to estimate by the method of moment.

The performance can be often improved for an appropriate L , and it can be sometimes terribly bad for an inappropriate L . However, as described in Remark 4.1, we can not easily determine the optimal coefficient L . The selection problem of L (or ρ), or more generally, the one of the better filter is important in practice.

Finally, we consider the model in Example 4.3 with the true values $(\mu_0, \sigma_0, \alpha_0, \beta_0) = (0.3, 0.2, 0.5, 1.8)$. A sample path of this model is shown in Fig. 4.3.

For the simulation, we need the condition $nh_n^{4\rho}(\log h_n)^2 \rightarrow 0$, so we choose $h_n = n^{-0.6}$ and $\rho = 0.48$. We again assume that α_0 is known to estimate β , or β_0 is known to estimate α , so we used the estimator (4.21) to estimate the jump part instead of (4.22). The result is shown in Tab. 4.4. As we have already described, such a model has the infinitely many jumps in any finite time interval and it has been difficult to estimate the parameters from only discrete data. Therefore, our method is worth using although it is not asymptotically efficient.

n (nh_n)	500 (12.01)	1000 (15.85)	3000 (24.60)	True value
$\hat{\mu}_n$	0.1318	0.1190	0.1110	0.1
$\hat{\sigma}_n$	0.0501	0.0500	0.0499	0.05
$\hat{\lambda}_n^{(0)}$	0.4398	0.4070	0.3748	0.3
$\hat{\nu}_n^{(0)}$	1.4661	1.3560	1.2411	1.0
J_n	4.02	5.17	7.71	—

Tab. 4.1: Result for Example 4.2. (μ, σ, λ) or (μ, σ, ν) are estimated jointly.

n (nh_n)	500 (12.01)	1000 (15.85)	3000 (24.60)	True value
$\hat{\mu}_n$	0.1004	0.0995	0.0999	0.1
$\hat{\sigma}_n$	0.1790	0.1774	0.1770	0.2
$\hat{\lambda}_n^{(0)}$	7.5168	6.9917	6.2219	5.0
$\hat{\nu}_n^{(0)}$	0.3758	0.3495	0.3114	0.25
J_n	52.6	79.92	158.1	—

Tab. 4.2: Result for different values of the parameters with Tab 4.1. (μ, σ, λ) or (μ, σ, ν) are estimated jointly.

n (nh_n)	500 (12.01)	1000 (15.85)	3000 (24.60)	True value
$\hat{\mu}_n$	0.1115	0.1112	0.1025	0.1
$\hat{\sigma}_n$	0.2230	0.2122	0.2041	0.2
$\hat{\lambda}_n^{(0)}$	7.3301	6.8551	6.1509	5.0
$\hat{\nu}_n^{(0)}$	0.3695	0.3392	0.3054	0.25
J_n	37.4	54.4	95.5	—

Tab. 4.3: Result for $L = 1.5$ with the same parameters as in Tab 4.2. (μ, σ, λ) or (μ, σ, ν) are estimated jointly.

n (nh_n)	500 (12.01)	1000 (15.85)	3000 (24.60)	True value
$\hat{\mu}_n$	0.3509	0.3302	0.3125	0.3
$\hat{\sigma}_n$	0.1800	0.1805	0.1921	0.2
$\hat{\alpha}_n^{(0)}$	0.5579	0.5558	0.5574	0.5
$\hat{\beta}_n^{(0)}$	2.0863	1.9759	1.9001	1.8
J_n	13.8	24.0	58.1	—

Tab. 4.4: Result for Example 4.3. (μ, σ, α) or (μ, σ, β) are estimated jointly.

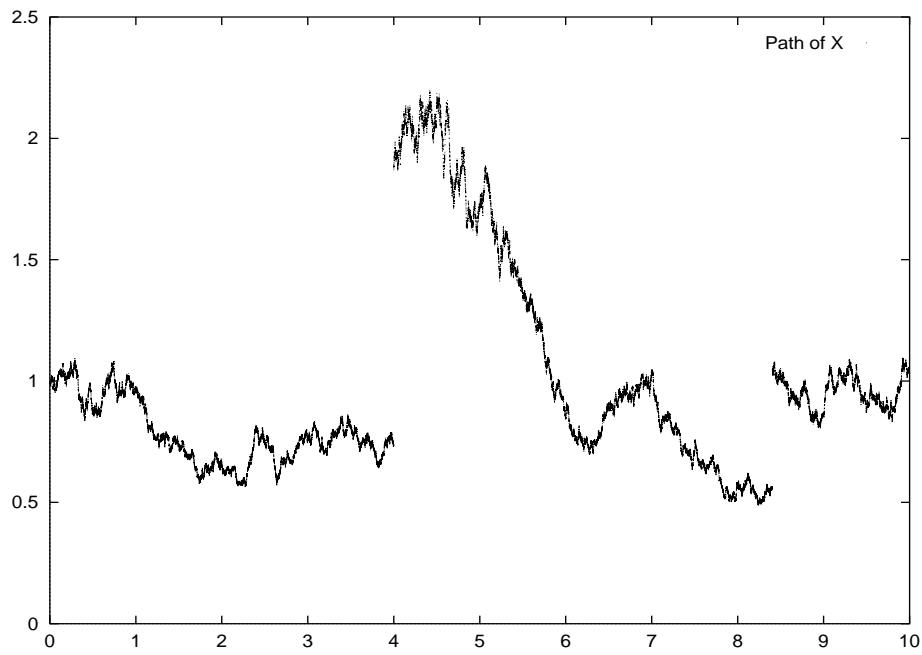


Fig. 4.1: A sample path of X : $dX_t = -\mu_0 X_t dt + \sqrt{\sigma_0} dw_t + dz_t^{(\lambda_0, \nu_0)}$, where z is a compound Poisson process with $f_\theta(z) = \frac{\lambda}{\sqrt{2\pi\nu}} \exp\left(-\frac{z^2}{2\nu}\right)$ and $(\mu_0, \lambda_0, \nu_0, \sigma_0) = (0.1, 0.3, 1.0, 0.05)$.

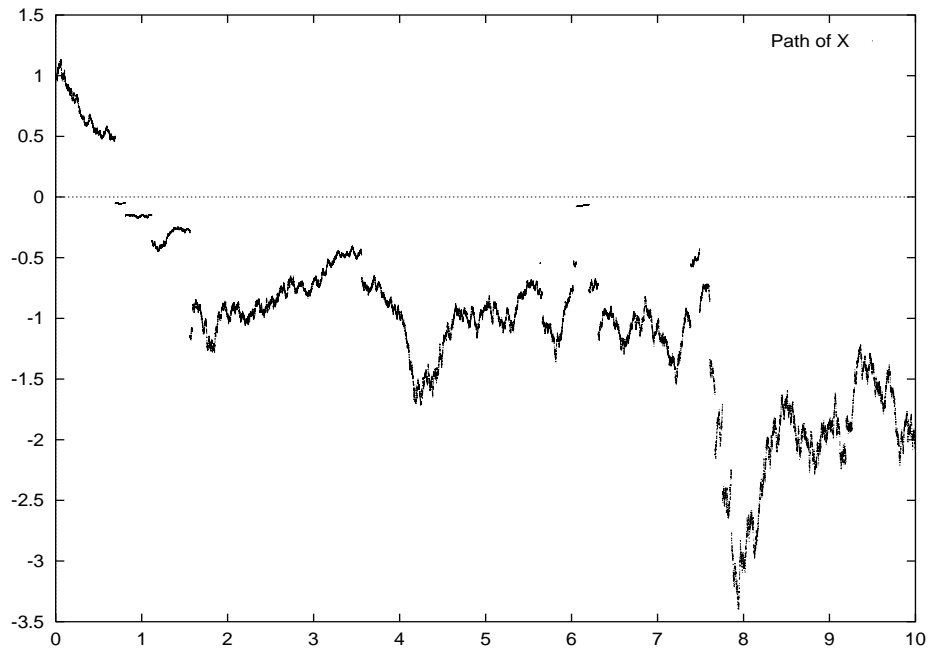


Fig. 4.2: A sample path of X : $dX_t = -\mu_0 X_t dt + \sqrt{\sigma_0} dw_t + dz_t^{(\lambda_0, \nu_0)}$, where z is a compound Poisson process with $f_\theta(z) = \frac{\lambda}{\sqrt{2\pi\nu}} \exp\left(-\frac{z^2}{2\nu}\right)$ and $(\mu_0, \lambda_0, \nu_0, \sigma_0) = (0.5, 4.0, 0.25, 0.2)$.

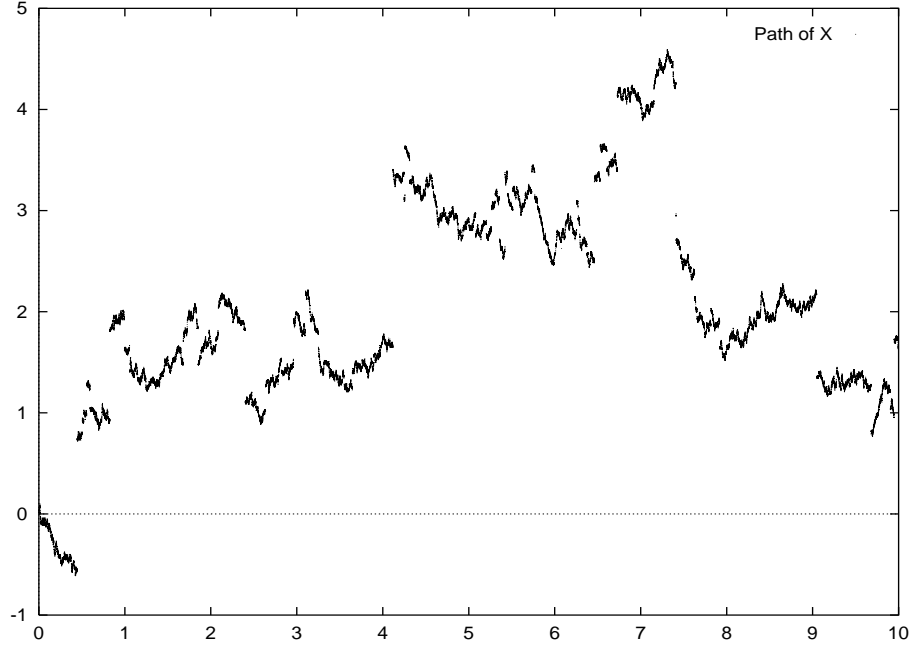


Fig. 4.3: A sample path of X : $dX_t = -\mu_0 X_t dt + \sqrt{\sigma_0} dw_t + dz_t^{(\alpha_0, \beta_0)}$, where z is a gamma process with $f_\theta(z) = 2^{-1} \alpha e^{-\beta|z|} |z|^{-1}$ and $(\mu_0, \sigma_0, \alpha_0, \beta_0) = (0.1, 0.2, 0.5, 1.8)$.

4.5 Moment estimates in infinite activity case

In this section, we shall give some lemmas and their proofs which are needed to show the asymptotic properties of estimators. First, we shall prepare some notations.

Let $L_{n,\alpha}$ and L_α be the following operators: for $g \in C^2$,

$$\begin{aligned} L_{n,\alpha} g(x) &= \sum_{i=1}^d a_n^{(i)}(x, \theta) \frac{\partial g}{\partial x_i}(x) + \frac{1}{2} \sum_{j,k=1}^d \beta^{(j,k)}(x, \sigma) \frac{\partial^2 g}{\partial x_j \partial x_k}(x) \\ &\quad + \int_{|z| \leq \varepsilon_n} \left\{ g(x + c(x, z, \theta)) - g(x) - \sum_{i=1}^d c^{(i)}(x, z, \theta) \frac{\partial g}{\partial x_i}(x) \right\} f_\theta(z) dz, \end{aligned}$$

and

$$\begin{aligned} L_\alpha g(x) &= \sum_{i=1}^d a^{(i)}(x, \theta) \frac{\partial g}{\partial x_i}(x) + \frac{1}{2} \sum_{j,k=1}^d \beta^{(j,k)}(x, \sigma) \frac{\partial^2 g}{\partial x_j \partial x_k}(x) \\ &\quad + \int \left\{ g(x + c(x, z, \theta)) - g(x) - \sum_{i=1}^d c^{(i)}(x, z, \theta) \frac{\partial g}{\partial x_i}(x) \right\} f_\theta(z) dz. \end{aligned}$$

Notice that $L_{n,\alpha}$ is the infinitesimal generator of $dX_t = a_n(X_t, \theta) dt + b(X_t, \sigma) dw_t + dB_n(t)$, and L_α is the one of SDE (4.1).

Lemma 4.5 *Suppose Conditions A1, A4, A5 and (4.2). Then it follows for any integer $k \geq 2$ and $t \in [t_{i-1}^n, t_i^n)$ that*

$$E_{i-1}^n \left[|X_t - X_{t_{i-1}^n}|^k \right] \leq C_k |t - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^k. \quad (4.23)$$

Let g be a function defined on $\mathbb{R}^d \times \Xi$ and is of polynomial growth in x uniformly in α then

$$E[|g(X_t, \alpha)|] \leq C(1 + |X_{t_{i-1}^n}|)^C. \quad (4.24)$$

Proof . The result follows by the same argument as in Lemma 3.5. \square

Lemma 4.6 *Suppose Conditions A1, A4, A5 and (4.2). Then*

$$E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \right] = h_n \int_{|z| > \varepsilon_n} c_{i-1}^{(l_1)}(z) f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right), \quad (4.25)$$

$$\begin{aligned} E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)} \right] &= h_n \left(\beta_{i-1}^{(l_1, l_2)} + \int_{\mathbb{R}^k} c_{i-1}^{(l_1)} c_{i-1}^{(l_2)}(z) f(z) dz \right) \\ &\quad + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right), \end{aligned} \quad (4.26)$$

$$E_{i-1}^n \left[\prod_{j=1}^3 \bar{X}_{i,n}^{(l_j)} \right] = h_n \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^3 c_{i-1}^{(l_j)}(z) \right\} f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right), \quad (4.27)$$

$$E_{i-1}^n \left[\prod_{j=1}^4 \bar{X}_{i,n}^{(l_j)} \right] = h_n \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^4 c_{i-1}^{(l_j)}(z) \right\} f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right). \quad (4.28)$$

Proof . Let $g_p^{(l_1, \dots, l_p)}(y, x) = \prod_{j=1}^p (y - x)^{(l_j)}$ where $l_j = 1, 2, \dots, d$. According to Proposition 2.6 and Lemma 4.5 (4.24),

$$\begin{aligned} E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \right] &= E_{i-1}^n \left[g_1^{(l_1)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] - h_n a_{n,i-1}^{(l_1)} \\ &= h_n L_{\alpha_0} g_1^{(l_1)}(y, X_{t_{i-1}^n}) \big|_{y=X_{t_{i-1}^n}} - h_n a_{i-1}^{(l_1)} \\ &\quad + h_n \int_{|z| > \varepsilon_n} c_{i-1}^{(l_1)}(z) f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right), \end{aligned}$$

where $L_{\alpha_0} g_1^{(l_1)}(y, X_{t_{i-1}^n}) \big|_{y=X_{t_{i-1}^n}} = a_{i-1}^{(l_1)}$. Hence

$$E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \right] = h_n \int_{|z| > \varepsilon_n} c_{i-1}^{(l_1)}(z) f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right).$$

The equation (4.26) and (4.27) are similarly proved as above, so we shall show (4.28) only.

$$E_{i-1}^n \left[\prod_{j=1}^4 \bar{X}_{i,n}^{(l_j)} \right]$$

$$\begin{aligned}
&= E_{i-1}^n \left[g_4^{(l_1, \dots, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] - h_n a_{n,i-1}^{(l_1)} E_{i-1}^n \left[g_3^{(l_2, l_3, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] \\
&\quad - h_n a_{n,i-1}^{(l_2)} E_{i-1}^n \left[g_3^{(l_1, l_3, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] - h_n a_{n,i-1}^{(l_3)} E_{i-1}^n \left[g_3^{(l_1, l_2, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] \\
&\quad - h_n a_{n,i-1}^{(l_4)} E_{i-1}^n \left[g_3^{(l_1, l_2, l_3)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right) \\
&= E_{i-1}^n \left[g_4^{(l_1, \dots, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] \\
&\quad + h_n^2 \sum_{k=1}^4 a_{n,i-1}^{(l_k)} \int_{\mathbb{R}^k} \left\{ \prod_{j=1, j \neq k}^4 c^{(l_j)}(x, z) \right\} f_\theta(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right).
\end{aligned}$$

We used (4.27) in the last equality. Applying Proposition 2.6 and Lemma 4.5 (4.24) to the first term, we have

$$E_{i-1}^n \left[g_4^{(l_1, \dots, l_4)}(X_{t_i^n}, X_{t_{i-1}^n}) \right] = h_n L g_4^{(l_1, \dots, l_4)}(y, X_{t_{i-1}^n})|_{y=X_{t_{i-1}^n}} + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right),$$

and it is easy to see that

$$L g_4^{(l_1, \dots, l_4)}(y, X_{t_{i-1}^n})|_{y=X_{t_{i-1}^n}} = \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^4 c_{i-1}^{(l_j)}(z) \right\} f(z) dz. \quad \square$$

By the same argument as in the above proof, we can obtain the next result.

Remark 4.3 Generally it follows for $p \geq 3$ that

$$E_{i-1}^n \left[\prod_{j=1}^p \bar{X}_{i,n}^{(l_j)} \right] = h_n \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^p c_{i-1}^{(l_j)}(z) \right\} f(z) dz + R \left(\alpha, h_n^2, X_{t_{i-1}^n} \right). \quad (4.29)$$

Proof . Obvious. \square

Lemma 4.7 Suppose Conditions A1, A4, A5 and (4.2). Then

$$E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] = R \left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+\rho}, X_{t_{i-1}^n} \right), \quad (4.30)$$

$$\begin{aligned}
E_{i-1}^n \left[\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] &= h_n \left(\beta_{i-1}^{(l_1, l_2)} + \int_{|z| \leq \varepsilon_n} c_{i-1}^{(l_1)} c_{i-1}^{(l_2)}(z) f(z) dz \right) \\
&\quad + R \left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+2\rho}, X_{t_{i-1}^n} \right), \quad (4.31)
\end{aligned}$$

$$\begin{aligned}
E_{i-1}^n \left[\prod_{j=1}^3 \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] &= h_n \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^3 c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\
&\quad + R \left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+3\rho} \vee h_n^2, X_{t_{i-1}^n} \right), \quad (4.32)
\end{aligned}$$

$$E_{i-1}^n \left[\prod_{j=1}^4 \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] = h_n \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^4 c_{i-1}^{(l_j)}(z) \right\} f(z) dz$$

$$\begin{aligned}
& + h_n^2 J_n^{(l_1, l_2, l_3, l_4)}(X_{t_{i-1}^n}, \alpha_0) \\
& + R\left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+4\rho} \vee h_n^3, X_{t_{i-1}^n}\right), \tag{4.33}
\end{aligned}$$

where $J_n^{(l_1, l_2, l_3, l_4)}$ is a polynomial growth function such that

$$\lim_{n \rightarrow \infty} J_n^{(l_1, l_2, l_3, l_4)} = \beta^{(l_1, l_2)} \beta^{(l_3, l_4)} + \beta^{(l_1, l_3)} \beta^{(l_2, l_4)} + \beta^{(l_1, l_4)} \beta^{(l_2, l_3)}.$$

Moreover, for $p \geq 3$, we can also write that

$$\begin{aligned}
E_{i-1}^n \left[\prod_{j=1}^p \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] & = h_n \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^p c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\
& + R\left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+p\rho} \vee h_n^2, X_{t_{i-1}^n}\right). \tag{4.34}
\end{aligned}$$

Proof . We shall show only (4.33). The proofs for others are done similarly.

Let $M_{i,n,p}(\theta) = \prod_{j=1}^p \bar{X}_{i,n}^{(l_j)}$.

$$\begin{aligned}
E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] & = E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{C_{i,0}^n(\varepsilon_n)} \right] + E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{C_{i,1}^n(\varepsilon_n)} \right] \\
& = E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{C_{i,0}^n(\varepsilon_n)} \right] + R\left(\alpha, \lambda^{(\varepsilon_n)} h_n^{1+4\rho}, X_{t_{i-1}^n}\right)
\end{aligned}$$

since $|\bar{X}_{i,n}^{(l_j)}| \mathbf{1}_{C_{i,1}^n(\varepsilon_n)} \leq h_n^\rho \mathbf{1}_{C_{i,1}^n(\varepsilon_n)}$ and $P_{i-1}^n\{C_{i,1}^n(\varepsilon_n)\} = R\left(\alpha, \lambda^{(\varepsilon_n)} h_n, X_{t_{i-1}^n}\right)$. On the other hand, we shall notice that the diffusion with jumps X follows the following stochastic differential equation

$$dX_t = a_n(X_t) dt + b(X_t) dw_t + dB_n(t), \tag{4.35}$$

on the set $\{J_i^n(\varepsilon_n) = 0\}$. Therefore, we can replace X in $E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{\{J_i^n(\varepsilon_n)=0\}} \right]$ with Y which follows the differential equation (4.35) and which is independent of $J_i^n(\varepsilon_n)$, that is,

$$\begin{aligned}
E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{C_{i,0}^n(\varepsilon_n)} \right] & = E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{\{J_i^n(\varepsilon_n)=0\}} \right] - E_{i-1}^n \left[M_{i,n,4} \mathbf{1}_{D_{i,0}^n(\varepsilon_n)} \right] \\
& = E_{i-1}^n \left[\prod_{j=1}^4 \bar{Y}_{i,n}^{(l_j)} \right] P_{i-1}^n\{J_i^n(\varepsilon_n) = 0\} + R\left(\alpha, h_n^q, X_{t_{i-1}^n}\right)
\end{aligned}$$

for any $q \geq 1$. The last equality is obtained by Lemma 4.4. We can calculate the expectation $E_{i-1}^n \left[\prod_{j=1}^4 \bar{Y}_{i,n}^{(l_j)} \right]$ by applying Proposition 2.6 and Lemma 4.5 (4.24) with $L_{n,\alpha}$ to

$$E_{i-1}^n \left[g_4^{(l_1, \dots, l_4)}(Y_{t_i^n}, Y_{t_{i-1}^n}) \right],$$

that is, by the same argument as in Lemma 4.6, we have

$$\begin{aligned} E_{i-1}^n \left[\prod_{j=1}^4 \bar{Y}_{i,n}^{(l_j)} \right] &= h_n \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^p c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\ &\quad + h_n^2 J_n^{(l_1, l_2, l_3, l_4)}(X_{t_{i-1}^n}, \alpha_0) + R\left(\alpha, h_n^3, X_{t_{i-1}^n}\right), \end{aligned}$$

where

$$\begin{aligned} J_n^{(l_1, l_2, l_3, l_4)}(x, \alpha) &= \sum_{k=1}^4 a_{n, i-1}^{(l_k)}(\theta) \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1, j \neq k}^4 c^{(l_j)}(x, z, \theta) \right\} f_\theta(z) dz \\ &\quad + \frac{1}{2} L_{n, \alpha}^2 g_4^{(l_1, \dots, l_4)}(y, x) \Big|_{y=x}, \end{aligned}$$

and it is easy to see by simple computation that

$$\begin{aligned} \lim_{n \rightarrow \infty} J_n^{(l_1, l_2, l_3, l_4)}(x, \alpha_0) &= \frac{1}{2} \lim_{n \rightarrow \infty} L_{n, \alpha_0}^2 g_4^{(l_1, \dots, l_4)}(y, x) \Big|_{y=x} \\ &= \beta^{(l_1, l_2)} \beta^{(l_3, l_4)} + \beta^{(l_1, l_3)} \beta^{(l_2, l_4)} + \beta^{(l_1, l_4)} \beta^{(l_2, l_3)} \end{aligned}$$

since $\int_{|z| \leq \varepsilon_n} g_n(x, z) f(z) dz \rightarrow 0$ for $g_n \rightarrow g$ which is integrable with respect to the measure $f(z) dz$. \square

Corollary 4.2 For $p \geq 1$,

$$\begin{aligned} E_{i-1}^n \left[\prod_{j=1}^p \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^p\}} \right] &= h_n \int_{|z| > \varepsilon_n} \left\{ \prod_{j=1}^p c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\ &\quad + R\left(\alpha, \lambda_0^{(\varepsilon_n)} h_n^{1+p\rho} \vee h_n^2, X_{t_{i-1}^n}\right). \end{aligned} \quad (4.36)$$

Proof . These are immediate consequences from Lemmas 4.6 and 4.7. \square

4.6 Limit theorems

Proposition 4.1 Assume Conditions A1, A2, and that $\lambda_0^{(\varepsilon_n)} h_n \rightarrow 0$ as $n \rightarrow \infty$. We denote by $g(x, \alpha) : \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ a function which satisfies the following conditions:

$$|g^{(n)}(x, \alpha)| \leq L(x, \alpha), \quad |\partial_\alpha g^{(n)}|, \quad |\partial_x g^{(n)}| \leq C(1 + |x|)^C,$$

where $L(x, \alpha)$ is a π -integrable function for all α , and there exists a function $g(x, \alpha)$ such that

$$g^{(n)}(x, \alpha) \rightarrow g(x, \alpha) \quad \pi\text{-a.s.}$$

as $n \rightarrow \infty$. Then $g(x, \alpha)$ is a π -integrable for all α and

$$\sup_{\alpha \in \Xi} \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \int g(x, \alpha) d\pi(x) \right| \xrightarrow{P} 0, \quad (4.37)$$

$$\sup_{\alpha \in \Xi} \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} - \int g(x, \alpha) d\pi(x) \right| \xrightarrow{P} 0. \quad (4.38)$$

Proof . The π -integrability of $g(x, \alpha)$ is the immediate result from the uniform integrability of $\{g^{(n)}(x, \alpha)\}_n$.

Let us show the convergences for fixed α . On (4.37),

$$\begin{aligned} P \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \int g(x, \alpha) \pi(dx) \right| > \epsilon \right\} \\ \leq P \left\{ \left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds \right| > \frac{\epsilon}{3} \right\} \\ + P \left\{ \left| \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds - \frac{1}{nh_n} \int_0^{nh_n} g(X_s, \alpha) ds \right| > \frac{\epsilon}{3} \right\} \\ + P \left\{ \left| \frac{1}{nh_n} \int_0^{nh_n} g(X_s, \alpha) ds - \int g(x, \alpha) \pi(dx) \right| > \frac{\epsilon}{3} \right\}. \end{aligned}$$

The third term converges to zero by Assumption A2. Let us call the first and second terms P_n^1 and P_n^2 respectively, then

$$\begin{aligned} P_n^1 &\leq \frac{3}{\epsilon} E \left[\left| \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) - \frac{1}{nh_n} \int_0^{nh_n} g^{(n)}(X_s, \alpha) ds \right| \right] \\ &\leq \frac{3}{\epsilon} E \left[\frac{1}{nh_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} |g^{(n)}(X_s, \alpha) - g_{i-1}^{(n)}(\alpha)| ds \right]. \end{aligned}$$

Applying Taylor's formula and Schwarz's inequality, we have

$$\begin{aligned} P_n^1 &\leq \frac{3}{nh_n \epsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left(E \left[|X_s - X_{t_{i-1}^n}|^2 \right] \right)^{\frac{1}{2}} \\ &\quad \times \left(E \left[\left(\int_0^1 \partial_x g^{(n)}(X_s + u(X_s - X_{t_{i-1}^n})) du \right)^2 \right] \right)^{\frac{1}{2}} ds. \end{aligned}$$

Lemma 4.5 yields that

$$P_n^1 \leq \frac{3}{nh_n \epsilon} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \left(E \left[C |s - t_{i-1}^n| (1 + |X_{t_{i-1}^n}|)^C \right] \right)^{\frac{1}{2}} \left(E \left[(1 + |X_{t_{i-1}^n}|)^C \right] \right)^{\frac{1}{2}} ds$$

$$\begin{aligned}
&\leq \frac{1}{nh_n\epsilon} \sum_{i=1}^n \left(\int_{t_{i-1}^n}^{t_i^n} |s - t_{i-1}^n|^{1/2} ds \right) \\
&\leq O(\sqrt{h_n})
\end{aligned}$$

and

$$\begin{aligned}
P_n^2 &\leq \frac{3}{\epsilon nh_n} \int_0^{nh_n} E|g^{(n)}(X_t, \alpha) - g(X_t, \alpha)| dt \\
&= \frac{3}{\epsilon} \int |g^{(n)}(x, \alpha) - g(x, \alpha)| d\pi.
\end{aligned}$$

This converges to zero by Lebesgue's convergence theorem.

The convergence (4.38) is immediately deduced from the fact that

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \left(1 - \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right) &= \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\
&= o_p \left(\sqrt{\lambda_0^{(\varepsilon_n)} h_n} \right).
\end{aligned}$$

Finally we have to show the uniformity of the convergence in α . We will only show (4.37). The one for (4.38) can be proved similarly. Let $s_n(\alpha) = \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha)$, and we regard this as a random element taking values in $(C(\Xi), \|\cdot\|_\infty)$ and we will check the tightness of this sequence; see Remark B.1 in Appendix B. Since we already showed the convergence of the marginal distribution of $s_n(\alpha)$, the tightness is implied by $\sup_n E \left[\sup_\alpha |\partial_\alpha s_n(\alpha)| \right] < \infty$; see Corollary B.1, and it is clear from Assumption A2. \square

Proposition 4.2 *Assume Conditions A1, A2, A4 and A5, and that $\lambda_0^{(\varepsilon_n)} h_n^{2\rho} \rightarrow 0$ and $\lambda_0^{(\varepsilon_n)} h_n^{4\rho-1} = O(1)$ as $n \rightarrow \infty$. We denote by $g(x, \alpha) : \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ a function which satisfies the following conditions:*

$$|g^{(n)}(x, \alpha)| \leq L(x, \alpha), \quad |\partial_\alpha g^{(n)}|, \quad |\partial_x g^{(n)}| \leq C(1 + |x|)^C,$$

where $L^2(x, \alpha)$ is a π -integrable function for all α , and there exists a function $g(x, \alpha)$ such that

$$g^{(n)}(x, \alpha) \rightarrow g(x, \alpha) \quad \pi\text{-a.s.}$$

as $n \rightarrow \infty$. Then $g(x, \alpha)$ is a π -integrable for all α and

$$\frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \xrightarrow{P} \int g(x, \alpha) \beta^{(l_1, l_2)}(x) d\pi \quad (4.39)$$

uniformly in α .

Proof . The π -integrability of g is shown similarly as in the proof of the preceding Proposition.

On the convergence of (4.39) for fixed α , we set

$$\xi_i^n(\alpha) := \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}.$$

We show that

$$A_n = \sum_{i=1}^n E_{i-1}^n [\xi_i^n(\alpha)] \xrightarrow{P} \int g(x, \alpha) \beta^{(l_1, l_2)}(x) d\pi, \quad B_n = \sum_{i=1}^n E_{i-1}^n [(\xi_i^n(\alpha))^2] \xrightarrow{P} 0.$$

Applying Lemma 4.7 and Proposition 4.1 to A_n , we have

$$\begin{aligned} A_n &= \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)} \left(\beta_{i-1}^{(l_1, l_2)} + \int_{|z| \leq \varepsilon_n} c_{i-1}^{(l_1)} c_{i-1}^{(l_2)}(z) f(z) dz + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n^\rho, X_{t_{i-1}^n} \right) \right) \\ &\xrightarrow{P} \int g(x, \alpha) \beta^{(l_1, l_2)}(x) d\pi \end{aligned}$$

since $\int_{|z| \leq \varepsilon_n} c_{i-1}^{(l_1)} c_{i-1}^{(l_2)}(z) f(z) dz \rightarrow 0$ according to Assumption A5. Similarly, applying Lemma 4.7 and Proposition 4.1 again to B_n , we see that

$$\begin{aligned} B_n &= \frac{1}{nh_n} \cdot \frac{1}{n} \sum_{i=1}^n \left(g_{i-1}^{(n)} \right)^2 \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^2 \left(c_{i-1}^{(l_j)} \right)^2(z) \right\} f(z) dz \\ &\quad + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n^{2\rho}, X_{t_{i-1}^n} \right) \\ &\xrightarrow{P} 0. \end{aligned}$$

This ends the proof of (4.39) for fixed α .

The uniformity in α of the convergence (4.37) is shown by checking the tightness criterion $\sup_n E [\sup_\alpha |\partial_\alpha \sum_{i=1}^n \xi_i^n(\alpha)|] < \infty$.

By using the equality (4.33),

$$\begin{aligned} &E \left[\sup_\alpha \left| \partial_\alpha \sum_{i=1}^n \xi_i^n(\alpha) \right| \right] \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n \left\{ E \left[\sup_\alpha |\partial_\alpha g_{i-1}(\alpha)|^2 \right] \right\}^{1/2} \left\{ E \left[|\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)}|^2 \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \right\}^{1/2}. \end{aligned}$$

Here, using Corollary 4.1, we have

$$\begin{aligned} &E \left[|\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)}|^2 \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\ &= E \left[|\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)}|^2 \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) = 0\}} \right] + E \left[|\bar{X}_{i,n}^{(l_1)} \bar{X}_{i,n}^{(l_2)}|^2 \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) \geq 1\}} \right] \end{aligned}$$

$$= O(h_n^2) + O\left(\lambda_0^{(\varepsilon_n)} h_n^{4\rho+1}\right).$$

Hence we find that $E[\sup_\alpha |\partial_\alpha \sum_{i=1}^n \xi_i^n(\alpha)|] = O(1)$ as $n \rightarrow \infty$. This completes the proof. \square

Proposition 4.3 *Assume Conditions A1, A2, A4 and A5, and that $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ as $n \rightarrow \infty$. We denote by $g(x, \alpha) : \mathbb{R}^d \times \Xi \rightarrow \mathbb{R}$ a function which satisfies the following conditions:*

$$|g^{(n)}(x, \alpha)| \leq L(x, \alpha), \quad |\partial_\alpha g^{(n)}| + |\partial_x g^{(n)}| \leq C(1 + |x|)^C,$$

where C is a constant, $L^2(x, \alpha)$ is a π -integrable function for all α , and there exists a function $g(x, \alpha)$ such that

$$g^{(n)}(x, \alpha) \rightarrow g(x, \alpha) \quad \pi\text{-a.s.}$$

as $n \rightarrow \infty$. Then $g(x, \alpha)$ is a π -integrable for all α and, for $p \geq 1$,

$$\begin{aligned} \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \prod_{j=1}^p \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\ \xrightarrow{P} \iint g(x, \alpha) \prod_{j=1}^p c^{(l_j)}(x, z) f(z) dz d\pi, \end{aligned} \quad (4.40)$$

uniformly in α .

Proof . We set

$$\eta_i^n(\alpha) := \frac{1}{nh_n} g_{i-1}(\alpha) \prod_{j=1}^p \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}$$

for $p \geq 1$. We shall also show the following.

$$\begin{aligned} A'_n &= \sum_{i=1}^n E_{i-1}^n [\eta_i^n(\alpha)] \xrightarrow{P} \iint g(x, \alpha) \prod_{j=1}^p c^{(l_j)}(x, z) f(z) dz d\pi, \\ B'_n &= \sum_{i=1}^n E_{i-1}^n [(\eta_i^n(\alpha))^2] \xrightarrow{P} 0. \end{aligned}$$

Applying Corollary 4.2 and Proposition 4.1 to A'_n , we see that

$$A'_n = \frac{1}{n} \sum_{i=1}^n g_{i-1}^{(n)} \int_{|z| > \varepsilon_n} \prod_{j=1}^p c_{i-1}^{(l_j)}(z) f(z) dz + R\left(\alpha, \lambda^{(\varepsilon_n)} h_n^{\rho\rho}, X_{t_{i-1}^n}\right)$$

$$\xrightarrow{P} \iint g(x, \alpha) \prod_{j=1}^p c^{(l_j)}(x, z) f(z) dz d\pi,$$

and, similarly,

$$\begin{aligned} B'_n &= \frac{1}{nh_n} \cdot \frac{1}{n} \sum_{i=1}^n \left(g_{i-1}^{(n)} \right)^2 \int_{|z| > \varepsilon_n} \left\{ \prod_{j=1}^p \left(c_{i-1}^{(l_j)} \right)^2(z) \right\} f(z) dz \\ &\quad + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n^{2p\rho}, X_{t_{i-1}^n} \right) \\ &\xrightarrow{P} 0. \end{aligned}$$

This is the proof of the convergence (4.2) for fixed α .

Let $K(x, y) = \sup_{\alpha} |\partial_{\alpha} g(x)| \prod_{j=1}^p |y^{(l_j)-} - h_n a_n^{(l_j)}(x)|$. Let us show the following tightness criterion:

$$\sup_n \frac{1}{nh_n} \sum_{i=1}^n E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^{\rho}\}} \right] < \infty.$$

For fixed i , notice the decomposition

$$E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^{\rho}\}} - h_n \iint_{|z| > \varepsilon_n} K(x, c(x, z)) f(z) dz d\pi \right] = \sum_{k=1}^5 I_k,$$

where

$$\begin{aligned} I_1 &= E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^{\rho}\}} - K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{J_i^n(\varepsilon_n)=1\}} \right], \\ I_2 &= E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{J_i^n(\varepsilon_n)=1\}} - K(X_{t_{i-1}^n}, X_{\tau_i^n(\varepsilon_n)}) \mathbf{1}_{\{J_i^n(\varepsilon_n)=1\}} \right], \\ I_3 &= E \left[K(X_{t_{i-1}^n}, X_{\tau_i^n(\varepsilon_n)}) \mathbf{1}_{\{J_i^n(\varepsilon_n)=1\}} - \int_{t_{i-1}^n}^{t_i^n} \int K(X_{t_{i-1}^n}, c(X_{t_{i-1}^n}, z)) p^{(\varepsilon_n)}(dz, ds) \right], \\ I_4 &= E \left[\int_{t_{i-1}^n}^{t_i^n} \int K(X_{t_{i-1}^n}, c(X_{t_{i-1}^n}, z)) (p^{(\varepsilon_n)} - q^{(\varepsilon_n)})(dz, ds) \right], \\ I_5 &= E \left[\int_{t_{i-1}^n}^{t_i^n} \int K(X_{t_{i-1}^n}, c(X_{t_{i-1}^n}, z)) q^{(\varepsilon_n)}(dz, ds) \right. \\ &\quad \left. - h_n \iint_{|z| > \varepsilon_n} K(x, c(x, z)) f(z) dz d\pi \right]. \end{aligned}$$

See Lemma 4.1 and 4.4 on the definitions of $\tau_i^n(\varepsilon_n)$ and $J_i^n(\varepsilon_n)$.

On I_2 and I_3 , using Corollary 4.1 (4.9) and taking the same argument as in Proposition 3.4, we can obtain that $I_2 + I_3 = O \left(\lambda_0^{(\varepsilon_n)} h_n^{3/2} \right)$. Moreover, it is easy to see that

$I_4 + I_5 = 0$. Let us estimate I_1 :

$$I_1 = E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \left(\mathbf{1}_{D_{i,0}^n(\varepsilon_n)} + \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n)=1\}} + \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho, J_i^n(\varepsilon_n) \geq 2\}} \right) \right].$$

Noticing that

$$\begin{aligned} & E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n)=1\}} \right] \\ & \leq E \left[\sup_{\alpha} \partial_a |g_{i-1}(\alpha)| \prod_{j=1}^p (h_n^\rho + h_n |a_{n,i-1}|) \mathbf{1}_{\{J_i^n(\varepsilon_n)=1\}} \right] \\ & = E \left[\sup_{\alpha} \partial_a |g_{i-1}(\alpha)| \prod_{j=1}^p (h_n^\rho + h_n |a_{n,i-1}|) P_{i-1}^n \{J_i^n(\varepsilon_n) = 1\} \right] \\ & = O(\lambda_0^{(\varepsilon_n)} h_n^{p\rho+1}), \end{aligned}$$

and that $P_{i-1}^n \{J_i^n(\varepsilon_n) \geq 2\} = O_p \left((\lambda_0^{(\varepsilon_n)} h_n)^2 \right)$ and Hölder's inequality, we see, for arbitrary $0 < \delta < 1$,

$$\begin{aligned} I_1 &= O(h_n^q) + O(\lambda_0^{(\varepsilon_n)} h_n^{p\rho+1}) + O \left((\lambda_0^{(\varepsilon_n)} h_n)^{2\delta} \right) \\ &= O(\lambda_0^{(\varepsilon_n)} h_n^{p\rho+1}) + O \left(\left(\lambda_0^{(\varepsilon_n)} h_n^{1-(2\delta)^{-1}} \right)^{2\delta} \right) \rightarrow 0 \end{aligned}$$

if we take δ as $2^{-1}(1 - \rho)^{-1} \leq \delta < 1$. Therefore,

$$\frac{1}{nh_n} \sum_{i=1}^n E \left[K(X_{t_{i-1}^n}, \Delta_i X^n) \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right] \longrightarrow \iint K(x, c(x, z)) f(z) dz d\pi < \infty.$$

This completes the proof. \square

Proposition 4.4 *Assume Conditions A1, A2, A4 and A5, and that $\lambda_0^{(\varepsilon_n)} h_n^{2\rho} \rightarrow 0$ as $n \rightarrow \infty$. Then*

$$\sup_{\alpha \in \Xi} \left| \frac{1}{nh_n} \sum_{i=1}^n g_{i-1}^{(n)}(\alpha) \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right| \xrightarrow{P} 0, \quad (4.41)$$

where $l = 1, 2, \dots, d$, and a function $g^{(n)}(x, \alpha)$ is given in Proposition 4.3.

Proof . Set

$$\zeta_i^n(\alpha) = \bar{\zeta}_i^n(\alpha) + \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta X_i^n| > h_n^\rho\}},$$

where $\bar{\zeta}_i^n(\alpha) = \frac{1}{nh_n} g_{i-1}(\alpha) \bar{X}_{i,n}^{(l)}$. The convergence of $\sum_{i=1}^n \zeta_i^n(\alpha)$ for each α is similar to the proof of Proposition 4.2.

Let us show the tightness of $\{\sum_{i=1}^n \zeta_i^n(\alpha)\}_n$. The tightness of the second term in the last right-hand side has been proved in Proposition 4.3. Therefore, we only show the tightness of $\{\sum_{i=1}^n \bar{\zeta}_i^n(\alpha)\}_{n \in \mathbb{N}}$. The proof is similar to the one of Proposition 3.3; see Theorem B.8 in Appendix B. We show, for any $N \in \mathbb{N}$ and a constant $H > 0$, that

$$E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] \leq H, \quad (4.42)$$

$$E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha_1) - \sum_{i=1}^n \bar{\zeta}_i^n(\alpha_2) \right)^{2N} \right] \leq H |\alpha_1 - \alpha_2|^{2N}. \quad (4.43)$$

We only show the inequality (4.42). The inequality (4.43) is similarly proved.

Setting $H(s, \alpha) = \sum_{i=1}^n g_{i-1}(\alpha) \mathbf{1}_{[t_{i-1}^n, t_i^n)}(s)$, we have

$$\begin{aligned} \sum_{i=1}^n \bar{\zeta}_i^n(\alpha) &= \frac{1}{nh_n} \left\{ \int_0^{nh_n} H(s, \alpha) a^{(l)}(X_s) ds + \sum_{j=1}^r \int_0^{nh_n} H(s, \alpha) b^{(l,j)}(X_s) dW_s^{(j)} \right. \\ &\quad \left. + \int_0^{nh_n} \int H(s, \alpha) c^{(l)}(X_{s-}, z) r(ds, dz) - \sum_{i=1}^n g_{i-1} a_{n,i-1}^{(l)} h_n \right\}. \end{aligned}$$

Therefore,

$$\begin{aligned} E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] &\leq C_N \left\{ E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} H(s, \alpha) a^{(l)}(X_s) ds \right)^{2N} \right] \right. \\ &\quad + \sum_{j=1}^r E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} H(s, \alpha) b^{(l,j)}(X_s) dW_s^{(j)} \right)^{2N} \right] \\ &\quad + E \left[\left(\frac{1}{nh_n} \int_0^{nh_n} \int H(s, \alpha) c^{(l)}(X_{s-}, z) r(ds, dz) \right)^{2N} \right] \\ &\quad \left. + E \left[\left(\frac{1}{n} \sum_{i=1}^n g_{i-1} a_{n,i-1}^{(l)} \right)^{2N} \right] \right\}. \end{aligned}$$

Applying Jensen's inequality and Burkholder-Davis-Gundy's inequality, we see that

$$\begin{aligned} E \left[\left(\sum_{i=1}^n \bar{\zeta}_i^n(\alpha) \right)^{2N} \right] &\leq C_N \left\{ \frac{1}{nh_n} \int_0^{nh_n} E[H^{2N}(s, \alpha) (a^{(l)}(X_s))^{2N}] ds \right. \\ &\quad \left. + \frac{1}{(nh_n)^{N+1}} \sum_{j=1}^r \int_0^{nh_n} E[H^{2N}(s, \alpha) (b^{(l,j)}(X_s))^{2N}] ds \right. \end{aligned}$$

$$\begin{aligned}
& + \frac{1}{(nh_n)^{N+1}} \int_0^{nh_n} \int E[H^{2N}(s, \alpha)(c^{(l)}(X_{s-}, z))^{2N}] q(ds, dz) \\
& + E \left[\left(\frac{1}{n} \sum_{i=1}^n g_{i-1} a_{n,i-1}^{(l)} \right)^{2N} \right] \Bigg\}.
\end{aligned}$$

One can see that these all are bounded because of Condition A2 . \square

4.7 Proofs of the main theorems

4.7.1 Proof of consistency

We show the proof of Theorem 4.1.

We set $M_n = \begin{pmatrix} nh_n I_{m_1} & 0 \\ 0 & n I_{m_2} \end{pmatrix}$, where I_m is the m -dimensional identity matrix, and we set $\Psi_n(\alpha) = M_n^{-1} \psi_n(\alpha)$. Then, there exists a function $\Psi(\alpha)$ such that

$$\sup_{\alpha} |\Psi_n(\alpha) - \Psi(\alpha)| \xrightarrow{P} 0,$$

since, by Proposition 4.1 - 4.4, we can obtain the following convergence under Conditions A1 - A6 and $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ as $n \rightarrow 0$:

$$\begin{aligned}
& \frac{1}{nh_n} \sum_{i=1}^n \left\{ \psi_{i,n,q}^{(1)} + \psi_{i,n,q}^{(2)} \right\} (\alpha) \\
& \xrightarrow{P} \int \left\{ H_{Q_q}(x, c(x, z, \theta_0)) f_{\theta_0}(z) - H_{Q_q}(x, c^{(l_j)}(x, z, \theta)) f_{\theta}(z) \right\} dz \\
& + \sum_{k,l=1}^d \int_{\mathbb{R}^d} \partial_{\theta_q} \tilde{a}^{(k)}(x, \theta) (\beta^{-1}(x, \sigma))^{(k,l)} \left\{ \tilde{a}^{(l)}(x, \theta_0) - \tilde{a}^{(l)}(x, \theta) \right\} d\pi(x) \\
& = \int U_q(x, \theta, \sigma) d\pi(x),
\end{aligned}$$

and

$$\frac{1}{n} \sum_{i=1}^n \psi_{i,n,r}^{(3)}(\alpha) \xrightarrow{P} \int_{\mathbb{R}^d} \left\{ \text{tr} [\partial_{\sigma_r} \beta^{-1}(x, \sigma) \beta(x, \sigma_0)] + \frac{\partial_{\sigma_r} \det \beta(x, \sigma)}{\det \beta(x, \sigma)} \right\} d\pi(x).$$

These convergences are uniform in α and $\Psi(\alpha)$ is this limits. Therefore, the similar argument as in Section 3.5.1 yields the consistency of $\hat{\alpha}_n$ to α_0 . We omit the details.

4.7.2 Proof of asymptotic normality

According to Taylor's formula and the definition of $\hat{\alpha}_n$,

$$\int_0^1 \partial_\alpha \psi_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) du (\hat{\alpha}_n - \alpha_0) = -\psi_n(\alpha_0).$$

By multiplying $M_n^{-1/2}$ from the left in the both sides, we obtain that

$$\int_0^1 C_n(\alpha_0 + u(\hat{\alpha}_n - \alpha_0)) du \cdot M_n^{-1/2}(\hat{\alpha}_n - \alpha_0) = -L_n(\alpha_0), \quad (4.44)$$

where

$$C_n(\alpha) = \begin{pmatrix} \frac{1}{nh_n} \partial_\theta \psi_n^{(1)}(\alpha) & \frac{1}{n\sqrt{h_n}} \partial_\sigma \psi_n^{(1)}(\alpha) \\ \frac{1}{n\sqrt{h_n}} \partial_\theta \psi_n^{(2)}(\alpha) & \frac{1}{n} \partial_\sigma \psi_n^{(2)}(\alpha) \end{pmatrix}, \quad L_n(\alpha) = \begin{pmatrix} \frac{1}{\sqrt{nh_n}} \psi_n^{(1)}(\alpha) \\ \frac{1}{\sqrt{n}} \psi_n^{(2)}(\alpha) \end{pmatrix}.$$

If we notice the relation (4.44), the following two lemmas easily deduce the desired results.

Lemma 4.8 *As $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ and $\lambda_0^{(\varepsilon_n)} h_n^{4\rho-1} \rightarrow 0$, we have*

$$(i) \ C_n(\alpha_0) \xrightarrow{P} B, \text{ where } B = \begin{pmatrix} B^{(1,1)} & B^{(1,2)} \\ B^{(2,1)} & B^{(2,2)} \end{pmatrix}, \ B^{(1,1)} = -K_1', \ B^{(2,2)} = 2K_2 \text{ and } B^{(1,2)} = B^{(2,1)} = 0.$$

$$(ii) \ \sup_{|\alpha| \leq \delta_n} |C_n(\alpha + \alpha_0) - C_n(\alpha_0)| \xrightarrow{P} 0 \text{ for any positive sequences } \delta_n \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Lemma 4.9 *Assume that $\lambda_0^{(\varepsilon_n)} h_n^\rho \rightarrow 0$ and $n(\lambda_0^{(\varepsilon_n)})^2 h_n^{4\rho} \rightarrow 0$. Moreover, we take $\rho'(> \rho)$ so that $\int_{|z| \leq \varepsilon_n} |z|^2 f_{\theta_0}(z) dz = o(n^{-1/2})$. Then*

$$L_n(\alpha_0) \xrightarrow{d} \mathcal{N}_m(0, \bar{B}),$$

$$\text{where } \bar{B} = \begin{pmatrix} \bar{B}^{(1,1)} & \bar{B}^{(1,2)} \\ \bar{B}^{(2,1)} & \bar{B}^{(2,2)} \end{pmatrix} \text{ and } \bar{B}^{(1,1)} = K_1'', \ \bar{B}^{(2,2)} = 4K_2 \text{ and } \bar{B}^{(1,2)} = \bar{B}^{(2,1)} = 0.$$

We shall prove above lemmas in the subsections below. This ends the proof. \square

4.7.3 Proof of Lemma 4.8

By simple computation, we have the following first derivatives of ψ_n : for $p = 1, \dots, m_1$, $s = 1, \dots, m_2$,

$$\begin{aligned}
\partial_{\theta_p} \psi_{i,n,q}^{(1)}(\alpha) &= \sum_{Q'=2}^{Q_q} \partial_{\theta_p} G_{n,i-1}^{(Q')}(\alpha) \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\
&\quad + \sum_{Q'=2}^{Q_q} G_{n,i-1}^{(Q')}(\alpha) \partial_{\theta_p} \left\{ \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \right\} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\
&\quad - h_n \sum_{Q'=2}^{Q_q} \partial_{\theta_p} \int_{\mathbb{R}^k} \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z, \theta) f_\theta(z) dz, \\
\partial_{\theta_p} \psi_{i,n,q}^{(2)}(\alpha) &= \sum_{k,l=1}^d \partial_{\theta_p} \partial_{\theta_q} a_{n,i-1}^{(k)}(\theta) (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \\
&\quad - h_n \sum_{k,l=1}^d \partial_{\theta_p} a_{n,i-1}^{(k)}(\theta) (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \partial_{\theta_q} a_{n,i-1}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\
\partial_{\sigma_r} \psi_{i,n,q}^{(1)}(\alpha) &= \sum_{Q'=2}^{Q_q} \left(\partial_{\sigma_r} G_{n,i-1}^{(Q')}(\alpha) \right) \left\{ \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \right\} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \\
&\quad - h_n \sum_{Q'=2}^{Q_q} \left(\partial_{\sigma_r} G_{n,i-1}^{(Q')}(\alpha) \right) \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z, \theta) \right\} f_\theta(z) dz, \\
\partial_{\sigma_r} \psi_{i,n,q}^{(2)}(\alpha) &= \sum_{k,l=1}^n \partial_{\theta_q} a_{n,i-1}(\theta) (\partial_{\sigma_r} \beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}, \\
\partial_{\theta_p} \psi_{i,n,r}^{(3)}(\alpha) &= \partial_{\sigma_r} \psi_{i,n,q}^{(2)}(\alpha), \\
\partial_{\sigma_s} \psi_{i,n,r}^{(3)}(\alpha) &= \sum_{k,l=1}^d \frac{1}{h_n} \partial_{\sigma_s} \partial_{\sigma_r} (\beta_{i-1}^{-1})^{(k,l)}(\sigma) \bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \\
&\quad + \sum_{k,l=1}^d \partial_{\sigma_s} \partial_{\sigma_r} \log \det \beta_{i-1} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}}.
\end{aligned}$$

Applying Proposition 4.1 and 4.3, we easily obtain that

$$\begin{aligned}
&\lim_{n \rightarrow \infty} \frac{1}{nh_n} \sum_{i=1}^n \partial_{\theta_p} \psi_{i,n,q}^{(1)}(\alpha) \\
&= \lim_{n \rightarrow \infty} \frac{1}{nh_n} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} \partial_{\theta_p} G_{n,i-1}^{(Q')}(\alpha) \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}}
\end{aligned}$$

$$\begin{aligned}
& + \lim_{n \rightarrow \infty} \frac{1}{nh_n} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} G_{n,i-1}^{(Q')}(\alpha) \partial_{\theta_p} \left\{ \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \right\} \mathbf{1}_{\{|\Delta_i X^n| > h_n^p\}} \\
& - \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} G_{n,i-1}^{(Q')}(\alpha) \partial_{\theta_p} \int_{\mathbb{R}^k} \prod_{j=1}^{Q'} c_{i-1}^{(l_j)} f_{\theta}(z) dz \\
& = \sum_{Q'=2}^{Q_q} \iint \partial_{\theta_p} G^{(Q')}(x, \alpha) \prod_{j=1}^{Q'} c^{(l_j)}(x, z, \theta_0) f_{\theta_0}(z) dz d\pi \\
& - \sum_{Q'=2}^{Q_q} \int \left\{ G^{(Q')}(x, \alpha) \partial_{\theta_p} \int_{\mathbb{R}^k} \prod_{j=1}^{Q'} c^{(l_j)}(x, z, \theta_0) f_{\theta_0}(z) dz \right\} d\pi.
\end{aligned}$$

This is the convergence in probability uniformly in α . Moreover, by Proposition 4.1 and 4.4,

$$\begin{aligned}
\frac{1}{nh_n} \sum_{i=1}^n \partial_{\theta_p} \psi_{i,n,q}^{(2)}(\alpha) & \xrightarrow{P} - \sum_{k,l=1}^d \int \partial_{\theta_p} a^{(k)} (\beta^{-1})^{(k,l)} \partial_{\theta_q} \tilde{a}^{(l)}(x, \alpha) d\pi \\
& = - \int (\partial_{\theta_p} \tilde{a})^* \beta^{-1} (\partial_{\theta_q} \tilde{a})(x, \alpha) d\pi,
\end{aligned}$$

$$\frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \left\{ \partial_{\sigma_r} \psi_{i,n,q}^{(1)}(\alpha) + \partial_{\sigma_r} \psi_{i,n,q}^{(2)}(\alpha) \right\} \xrightarrow{P} 0,$$

$$\frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \partial_{\theta_p} \psi_{i,n,r}^{(3)}(\alpha) = \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \partial_{\sigma_r} \psi_{i,n,q}^{(2)}(\alpha) \xrightarrow{P} 0.$$

These convergences are also uniformly in α . Furthermore, applying Proposition 4.1 and 4.2, we have

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \partial_{\sigma_s} \psi_{i,n,r}^{(3)}(\alpha) & \xrightarrow{P} \sum_{k,l=1}^d \int \partial_{\sigma_s} \partial_{\sigma_r} (\beta^{-1}(x, \sigma))^{(k,l)} \beta^{(k,l)}(x, \sigma_0) d\pi \\
& + \int \partial_{\sigma_s} \partial_{\sigma_r} \log \det \beta(x, \sigma) d\pi
\end{aligned}$$

uniformly in α . Hence

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \partial_{\sigma_s} \psi_{i,n,r}^{(3)}(\alpha_0) & = - \int \text{tr} [\partial_{\sigma_r} \beta^{-1} \partial_{\sigma_s} \beta] (x, \sigma_0) d\pi \\
& = \int \text{tr} [(\partial_{\sigma_r} \beta) \beta^{-1} (\partial_{\sigma_s} \beta) \beta^{-1}] (x, \sigma_0) d\pi. \quad \square
\end{aligned}$$

This ends the proof of (i).

The proof of (ii) is very easy. Notice that the convergence of C_n is uniform in the parameters and that the uniform limit of C_n is continuous with respect to the parameters. Denoting by $B(\alpha)$ the uniform limit of $C_n(\alpha)$, we have that

$$\begin{aligned} \sup_{|\alpha| \leq \varepsilon_n} |C_n(\alpha + \alpha_0) - C_n(\alpha_0)| \\ \leq 2 \sup_{|\alpha| \leq \varepsilon_n} |C_n(\alpha + \alpha_0) - B(\alpha + \alpha_0)| + \sup_{|\alpha| \leq \varepsilon_n} |B(\alpha + \alpha_0) - B|, \end{aligned}$$

where $B = B(\alpha_0)$. The last side tends to zero as n tends to infinity. \square

4.7.4 Proof of Lemma 4.9

According to the central limit theorem for a triangular array; see Theorem A.3 and its remark, it suffices to show the following conditions:

$$\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(v)} \right] \right| \xrightarrow{P} 0 \quad (v = 1, 2), \quad (4.45)$$

$$\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{n}} \psi_{i,n,q}^{(3)} \right] \right| \xrightarrow{P} 0, \quad (4.46)$$

$$\sum_{i=1}^n \sum_{v=1,2} E_{i-1}^n \left[\frac{1}{nh_n} \psi_{i,n,q}^{(v)} \psi_{i,n,q'}^{(v)} \right] \xrightarrow{P} (K_1'')^{(q,q')} \quad (v = 1, 2), \quad (4.47)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n} \psi_{i,n,r}^{(3)} \psi_{i,n,r'}^{(3)} \right] \xrightarrow{P} 4K_2^{(r,r')}, \quad (4.48)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \psi_{i,n,q}^{(1)} \psi_{i,n,q'}^{(2)} \right] \xrightarrow{P} 0, \quad (4.49)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \psi_{i,n,q}^{(v)} \psi_{i,n,q'}^{(3)} \right] \xrightarrow{P} 0 \quad (v = 1, 2), \quad (4.50)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(v)} \right|^{2+\nu} \right] \xrightarrow{P} 0 \quad (v = 1, 2, \nu > 0), \quad (4.51)$$

$$\sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n}} \psi_{i,n,r}^{(3)} \right|^{2+\nu} \right] \xrightarrow{P} 0 \quad (\nu > 0), \quad (4.52)$$

where $q, q' = 1, 2, \dots, m_1$, and $r, r' = 1, 2, \dots, m_2$. It is not necessary that ν for $v = 1, 2$ in (4.51) and one for (4.52) are the same value. The same argument was done in the proof of Lemma 3.9.

Proof of (4.45).

For $v = 1$, applying Corollary 4.2 (4.36), we have

$$\begin{aligned}
\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(1)} \right] \right| &= \frac{-h_n}{\sqrt{nh_n}} \sum_{i=1}^n \left| \sum_{Q'=2}^{Q_q} \left[G_{n,i-1}^{(Q')} \int_{|z| \leq \varepsilon_n} \left\{ \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z) \right\} f(z) dz \right. \right. \\
&\quad \left. \left. + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n^{\rho Q'} \vee h_n^2, X_{t_{i-1}^n} \right) \right] \right| \\
&= O_p \left(\sqrt{nh_n \varepsilon_n^2} \right) + O_p \left(\sqrt{nh_n^{1+4\rho} (\lambda^{(\varepsilon_n)})^2} \vee \sqrt{nh_n^5} \right) \\
&\xrightarrow{P} 0.
\end{aligned}$$

For $v = 2$, applying Lemma 3.38 (4.30), we have

$$\begin{aligned}
\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(1)} \right] \right| &= \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| \sum_{k,l=1}^d \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} E_{i-1}^n \left[\bar{X}_{i,n} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \right| \\
&= O_p \left(\sqrt{nh_n^{1+2\rho} (\lambda^{(\varepsilon_n)})^2} \right) \\
&\xrightarrow{P} 0.
\end{aligned}$$

Proof of (4.46).

Applying Lemma 3.38 (4.31) and Lemma 4.4, we have

$$\begin{aligned}
\sum_{i=1}^n \left| E_{i-1}^n \left[\frac{1}{\sqrt{n}} \psi_{i,n,q}^{(3)} \right] \right| &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left| \left[\text{tr}[\partial_{\theta_r} \beta_{i-1}^{-1} \beta_{i-1}] + \int_{|z| \leq \varepsilon_n} c_{i-1}^{(l_1)} c_{i-1}^{(l_2)} f(z) dz + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n^{2\rho}, X_{t_{i-1}^n} \right) \right] \right. \\
&\quad \left. + \left[\frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} + R \left(\alpha, \lambda^{(\varepsilon_n)} h_n, X_{t_{i-1}^n} \right) \right] \right|.
\end{aligned}$$

Noticing that $\text{tr}[\partial_{\sigma_r} \beta_{i-1}^{-1} \beta_{i-1}] = -\frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}}$, and $\int_{|z| \leq \varepsilon_n} |z|^2 f(z) dz = o(n^{-1/2})$, we have

$$\begin{aligned}
\sum_{i=1}^n \left| E \left[\frac{1}{\sqrt{n}} \psi_{i,n,q}^{(3)} \right] \right| &= O_p \left(\sqrt{n(\lambda^{(\varepsilon_n)})^2 h_n^{4\rho}} \right) + O_p \left(\sqrt{n(\lambda^{(\varepsilon_n)} h_n)^2} \right) + o_p(1) \\
&\xrightarrow{P} 0.
\end{aligned}$$

Proof of (4.47).

Applying Corollary 4.2 (4.36), we obtain

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \psi_{i,n,q}^{(1)} \psi_{i,n,q'}^{(1)} \right] \\
&= \frac{1}{nh_n} \sum_{i=1}^n \sum_{Q,Q'=2}^{Q_q, Q'_q} G_{n,i-1}^{(Q)} G_{n,i-1}^{(Q')} E_{i-1}^n \left[\prod_{j=1}^{Q+Q'} \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right] \\
&\quad - \frac{2}{n} \sum_{i=1}^n \sum_{Q,Q'=2}^{Q_q, Q'_q} G_{n,i-1}^{(q)} G_{n,i-1}^{(q')} \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^Q c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\
&\quad \times E_{i-1}^n \left[\prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right] \\
&\quad + \frac{h_n}{n} \sum_{i=1}^n \sum_{Q,Q'=2}^{Q_q, Q'_q} \prod_{q=Q,Q'} G_{n,i-1}^{(q)} \left(\int_{\mathbb{R}^k} \left\{ \prod_{j=1}^q c_{i-1}^{(l_j)}(z) \right\} f(z) dz \right) \\
&\xrightarrow{P} \sum_{Q,Q'=2}^{Q_q, Q'_q} \iint G^{(Q)} G^{(Q')}(x) \left\{ \prod_{j=1}^{Q+Q'} c^{(l_j)}(x, z) \right\} f(z) dz d\pi.
\end{aligned}$$

Moreover, applying Lemma 3.38 (4.31), we have

$$\begin{aligned}
\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \psi_{i,n,q}^{(2)} \psi_{i,n,q'}^{(2)} \right] &= \frac{1}{nh_n} \sum_{i=1}^n E_{i-1}^n \left[\left(\sum_{k,l=1}^d \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \bar{X}_{i,n}^{(l)} \right) \right. \\
&\quad \times \left. \left(\sum_{k,l=1}^d \partial_{\theta_{q'}} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \bar{X}_{i,n}^{(l)} \right) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\xrightarrow{P} \int (\partial_{\theta_q} \tilde{a})^* \beta^{-1} (\partial_{\theta_{q'}} \tilde{a})(x) d\pi.
\end{aligned}$$

Proof of (4.48).

Applying Lemma 3.38 (4.31), (4.33) and Lemma 4.4, and noticing that

$$\int_{|z| \leq \varepsilon_n} \prod_{j=1}^4 c_{i-1}^{(l_j)}(z) f(z) dz = o_p \left(h_n^{2\rho'} n^{-1/2} \right),$$

since $|c(x, z)| \leq C|z|(1 + |x|)^C$ on the set $\{|z| \leq 1\}$ and $\int_{|z| \leq \varepsilon_n} |z|^2 f(z) dz = o(n^{-1/2})$, we have

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n} \psi_{i,n,r}^{(3)} \psi_{i,n,r'}^{(3)} \right]$$

$$\begin{aligned}
&= \frac{1}{nh_n^2} \sum_{i=1}^n \sum_{k,l,k',l'=1}^d (\partial_{\sigma_r} \beta_{i-1}^{-1})^{(k,l)} (\partial_{\sigma_{r'}} \beta_{i-1}^{-1})^{(k',l')} \\
&\quad \times E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \bar{X}_{i,n}^{(k')} \bar{X}_{i,n}^{(l')} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\quad + \frac{1}{nh_n} \sum_{i=1}^n \sum_{k,l=1}^d \frac{\partial_{\sigma_{r'}} \det \beta_{i-1}}{\det \beta_{i-1}} (\partial_{\sigma_r} \beta_{i-1}^{-1})^{(k,l)} E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\quad + \frac{1}{nh_n} \sum_{i=1}^n \sum_{k',l'=1}^d \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} (\partial_{\sigma_{r'}} \beta_{i-1}^{-1})^{(k',l')} E_{i-1}^n \left[\bar{X}_{i,n}^{(k')} \bar{X}_{i,n}^{(l')} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} \frac{\partial_{\sigma_{r'}} \det \beta_{i-1}}{\det \beta_{i-1}} P_{i-1}^n \{|\Delta_i X^n| \leq h_n^\rho\} \\
&= \frac{1}{n} \sum_{i=1}^n \sum_{k,l,k',l'=1}^d (\partial_{\sigma_r} \beta_{i-1}^{-1})^{(k,l)} (\partial_{\sigma_{r'}} \beta_{i-1}^{-1})^{(k',l')} J_n^{(k,l,k',l')}(X_{t_{i-1}^n}, \alpha) \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{\partial_{\sigma_{r'}} \det \beta_{i-1}}{\det \beta_{i-1}} \text{tr} [\partial_{\sigma_r} \beta_{i-1}^{-1} \beta_{i-1}] + \frac{1}{n} \sum_{i=1}^n \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} \text{tr} [\partial_{\sigma_{r'}} \beta_{i-1}^{-1} \beta_{i-1}] \\
&\quad + \frac{1}{n} \sum_{i=1}^n \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} \frac{\partial_{\sigma_{r'}} \det \beta_{i-1}}{\det \beta_{i-1}} + O_p \left(\lambda_0^{(\varepsilon_n)} h_n^{4\rho-1} \right).
\end{aligned}$$

Hence $\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n} \psi_{i,n,r}^{(3)} \psi_{i,n,r'}^{(3)} \right] \xrightarrow{P} 4K_2^{(r,r')}$ by Proposition 4.1 (4.37).

Proof of (4.49).

By Lemma 3.38 (4.30), we see that

$$\begin{aligned}
\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{nh_n} \psi_{i,n,q}^{(1)} \psi_{i,n,q'}^{(2)} \right] &= -\frac{1}{n} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} G_{n,i-1}^{(Q')} \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z) \right\} f(z) dz \\
&\quad \times \sum_{k,l=1}^d \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
&= O_p \left(\lambda_0^{(\varepsilon_n)} h_n^{1+\rho} \right) \xrightarrow{P} 0.
\end{aligned}$$

Proof of (4.50).

For $v = 1$, applying Lemma 4.4 and Lemma 3.38 (4.31), we have

$$\sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \psi_{i,n,q}^{(1)} \psi_{i,n,q'}^{(3)} \right] = -\frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} G_{n,i-1}^{(Q')} \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z) \right\} f(z) dz$$

$$\begin{aligned}
& \times \left[\sum_{k,l=1}^d \partial_{\sigma_r} (\beta_{i-1}^{-1})^{(k,l)} E_{i-1}^n \left[\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \right. \\
& \quad \left. + h_n \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} P_{i-1}^n \{|\Delta_i X^n| \leq h_n^\rho\} \right] \\
& = O_p \left(\sqrt{h_n} \right).
\end{aligned}$$

For $v = 2$, applying Lemma 3.38 (4.30) and (4.32), we have

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\frac{1}{n\sqrt{h_n}} \psi_{i,n,q}^{(2)} \psi_{i,n,q'}^{(3)} \right] \\
& = \frac{\sqrt{h_n}}{n} \sum_{i=1}^n \sum_{k,l,k',l'=1}^d \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \partial_{\sigma_r} (\beta_{i-1}^{-1})^{(k',l')} \\
& \quad \times E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \bar{X}_{i,n}^{(k')} \bar{X}_{i,n}^{(l')}(\theta) \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
& \quad + \frac{1}{n\sqrt{h_n}} \sum_{i=1}^n \sum_{k,l=1}^d \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}} E_{i-1}^n \left[\bar{X}_{i,n}^{(l)} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\
& = O_p(\lambda_0^{(\varepsilon_n)} h_n^{\rho+1/2}).
\end{aligned}$$

Proof of (4.51).

For $v = 1$, applying Lemma 4.2 (4.36), we have

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(1)} \right|^{2+\nu} \right] \\
& \leq \frac{C}{(nh_n)^{1+\nu/2}} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} \left(G_{n,i-1}^{(Q')} \right)^{2+\nu} E_{i-1}^n \left[\left| \prod_{j=1}^{Q'} \bar{X}_{i,n}^{(l_j)} \right|^{2+\nu} \mathbf{1}_{\{|\Delta_i X^n| > h_n^\rho\}} \right] \\
& \quad + \frac{Ch_n^{2+\nu}}{(nh_n)^{1+\nu/2}} \sum_{i=1}^n \sum_{Q'=2}^{Q_q} E_{i-1}^n \left[\left| G_{n,i-1}^{(Q')} \int_{\mathbb{R}^k} \left\{ \prod_{j=1}^{Q'} c_{i-1}^{(l_j)}(z) \right\} f(z) dz \right|^{2+\nu} \right] \\
& = O_p \left(\frac{1}{n^{\nu/2} h_n^{\nu/2}} \right) + O_p \left(\frac{h_n^{1+\nu}}{n^{\nu/2} h_n^{\nu/2}} \right).
\end{aligned}$$

For $v = 2$, applying Lemma 3.38 (4.34), we have

$$\begin{aligned}
& \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{nh_n}} \psi_{i,n,q}^{(2)} \right|^{2+\nu} \right] \\
& \leq \frac{C}{(nh_n)^{1+\nu/2}} \sum_{k,l=1}^d \left| \partial_{\theta_q} a_{n,i-1}^{(k)} (\beta_{i-1}^{-1})^{(k,l)} \right|^{2+\nu} E_{i-1}^n \left[\left| \bar{X}_{i,n}^{(l)} \right|^{2+\nu} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right]
\end{aligned}$$

$$= O_p \left(\frac{1}{n^{\nu/2} h_n^{\nu/2}} \right).$$

Proof of (4.52).

Notice that, from Lemma 4.3 (4.8),

$$E_{i-1}^n \left[|\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)}|^{2+\nu} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n)=0\}} \right] = R \left(\alpha, h_n^{(2+\nu)/2 \wedge (1+\rho'\nu)}, X_{t_{i-1}^n} \right),$$

and that $|\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)}| = R \left(\alpha, h_n^{2\rho}, X_{t_{i-1}^n} \right)$ on the set $\{|\Delta_i X^n| \leq h_n^\rho, J_i^n(\varepsilon_n) \geq 1\}$. Then we see that

$$\begin{aligned} & \sum_{i=1}^n E_{i-1}^n \left[\left| \frac{1}{\sqrt{n}} \psi_{i,n,r}^{(3)} \right|^{2+\nu} \right] \\ & \leq \frac{1}{n^{1+\nu/2} h_n^{2+\nu}} \sum_{i=1}^n \sum_{k,l=1}^d \partial_{\sigma_r} (\beta_{i-1}^{-1})^{(k,l)} E_{i-1}^n \left[|\bar{X}_{i,n}^{(k)} \bar{X}_{i,n}^{(l)}|^{2+\nu} \mathbf{1}_{\{|\Delta_i X^n| \leq h_n^\rho\}} \right] \\ & \quad + \frac{1}{n^{1+\nu/2}} \sum_{i=1}^n \left| \frac{\partial_{\sigma_r} \det \beta_{i-1}}{\det \beta_{i-1}}(\sigma) \right| P_{i-1}^n \{|\Delta_i X^n| \leq h_n^\rho\} \\ & = O_p \left(\frac{h_n^{(2+\nu)/2 \wedge (1+\rho'\nu)}}{n^{\nu/2} h_n^{2+\nu}} \right) + O_p \left(\frac{\lambda^{(\varepsilon_n)} h_n^{1+2\rho(2+\nu)}}{n^{\nu/2} h_n^{2+\nu}} \right) + O_p \left(\frac{1}{n^{\nu/2}} \right) \\ & = O_p \left((n h_n)^{-\nu/2} h_n^{-1} \right) + O_p \left(\frac{\lambda^{(\varepsilon_n)} h_n^{4\rho-4(1-2\rho)/\nu}}{n h_n} \right) + O_p \left(\frac{1}{n^{\nu/2}} \right) \end{aligned}$$

The last right-hand side tends to zero as $n \rightarrow \infty$ if we take ν to be sufficiently large. This completes the proof. \square

Chapter 5

Nonparametric estimation of Lévy densities

This chapter concentrates on the density estimation of Lévy measures, which are one of the most important characteristics in jump-processes. We regard a *large* increment as a jump size approximately, and apply the idea of the usual kernel density estimation. We prove that our kernel estimator has the consistency in the sense of the mean squared error (MSE). In Section 5.4, we show some simulation results and point out a practical problem on our asymptotic filter. We shall give some intuitive solution for that problem meanwhile, and more theoretical method is discussed in the next chapter.

5.1 Nonparametric framework

Up to the previous chapter, we have discussed the parametric inference for diffusion processes with jumps. Parametric models can often be powerful tools to predict the future's phenomena once the true model is specified. However, in applications, we sometimes face the trouble of a parametrization of a model since, for example, we might get little prior information about the true model of the corresponding phenomena, and we could not determine the specific parametric model. Therefore we might need a rich parametrization which includes sufficiently many parameters. However it could then cause some statistical problems such as identifiabilities, or the optimization of estimating functions. On the other hand, too simple parametrization would cause misspecifications of the model. To break out of this dilemma, a nonparametric framework is useful.

As we have already seen, in the parametric inference treated in Chapter 3 and

4, the ergodicity of the model was essential to obtain some convergence theorems of estimating functions. However, in the practical case, especially in the financial modeling, non-ergodic models are well used, and we need to prepare the inference under the non-ergodic framework.

In this chapter we again return to a Poisson type jump-diffusion X :

$$dX_t = a(X_t) dt + b(X_t) dw_t + \int_{\mathcal{E}} c(X_{t-}, z) (p - q)(dt, dz),$$

where a, b and c are unknown Borel functions and $q(dt, dz) = f(z) dz dt$ is an unknown measure with $\int_{\mathcal{E}} f(z) dz < \infty$. Suppose that the observations consist of the same type of one as in Chapter 3 and 4; $\{X_{t_i^n}\}_{i=0}^n$, $t_i^n = ih_n$ with $h_n \rightarrow 0$ and $nh_n \rightarrow \infty$.

There are only a few previous works about the nonparametric inference under such a setting. Recently, Bandi and Nguyen [5] proposed some functional estimators of higher order infinitesimal moments conditional on $X_{t-} = x$ at any time $t > 0$:

$$a(x), \sigma^2(x) + \int_{\mathcal{E}} c^2(x, z) f(z) dz, \text{ and } \int_{\mathcal{E}} c^k(x, z) f(z) dz \ (k > 2),$$

by way of the local time estimates. However their method could not generally give the separate estimator of the diffusion part and the quadratic moment of jump size. Shimizu [90] gave an estimator of only the diffusion coefficient separately under the different sampling scheme; the terminal nh_n is fixed, by using the same type of the filter as in Chapter 3. Although we would not describe this method here since the experimental design is different from ours in this thesis, it could be easily imagined from the facts (4.31) and (4.36) in Chapter 4 that such a separation could be possible. By the same idea as above, Shimizu [93] also proposed a kernel type estimator of the Lévy density $f(z)$ under our sampling scheme. This kernel estimator also enables us to obtain the separate estimator of the quadratic moment of jumps; e.g. Shimizu [95]. In this chapter, we describe the density estimation of Lévy measures from discrete observations.

Let us explain the idea of a kernel density method used in this chapter. If we can obtain continuous data of X then it would be easy to construct consistent estimators of Lévy measures by the analogy of the case where samples are independent and identically distributed since we know all of amplitudes of jumps exactly. For example, we consider an 1-dimensional model:

$$dX_t = a(X_t) dt + b(X_t) dw_t + dz_t,$$

where a, b are unknown functions, w is a Wiener process, and z is a compound Poisson process with the Poisson intensity 1 and an unknown distribution of jumps $F(z) dz$. If

we can observe whole the path on $[0, T]$, so we can observe the first n jumps $(\Delta z_i)_{1 \leq i \leq n}$ on $[0, T]$, then we can estimate the density $F(z)$ by

$$F_n(z) = \frac{1}{n\delta_n} \sum_{i=1}^n K\left(\frac{z - \Delta z_i}{\delta_n}\right) \quad (5.1)$$

with a suitable kernel K and a sequence δ_n , which converges to the true F in the MSE sense at the rate $n\delta_n$; see Masry [65] for details. A more general type of inference from randomly sampled data is discussed in Prakasa Rao [78, 79].

If we can observe discrete data only, it is natural to substitute the increments of neighboring data for Δz_i 's in above $F_n(z)$ according to the context of Chapter 3. Though we consider more general types of stochastic differential equations, the same type of estimator can be available.

5.2 Notations and assumptions

We consider the following k -dimensional stochastic differential equation on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$:

$$\begin{cases} dX_t = a(X_t) dt + b(X_t) dw_t + \int_{\mathcal{E}} c(X_{t-}, z) (p - q)(dt, dz), \\ X_0 = x_0, \end{cases} \quad (5.2)$$

where x_0 is a random variable on \mathbb{R}^d , $\mathcal{E} = \mathbb{R}^d \setminus \{0\}$, $(w_t)_{t \geq 0}$ is a k' -dimensional Wiener process, $a(x)$ is an \mathbb{R}^k -valued Borel function defined on \mathbb{R}^k , $b(x)$ is an $\mathbb{R}^k \otimes \mathbb{R}^{k'}$ -valued Borel function defined on \mathbb{R}^k , $c(x, z)$ is an \mathbb{R}^k -valued Borel function defined on $\mathbb{R}^k \times \mathcal{E}$, $p(dt, dz)$ is a homogeneous Poisson random measure on $\mathbb{R}_+ \times \mathcal{E}$, and $q(dt, dz)$ is its intensity measure, that is, $E[p(dt, dz)] = q(dt, dz)$. We suppose that q has the expression $q(dt, dz) = f(z) dz dt$, that is, f is the Lévy density. Hereafter, it would also be convenient to regard p as a random measure associated with a compound Poisson process z of the form $z_t = \sum_{i=1}^{N_t} \varepsilon_i$ with the Lévy density $f(z) = \lambda F(z)$.

We suppose that the process X_t is observed at each time point $t_i^n = ih_n$ ($i = 1, 2, \dots, n$) in the time interval $[0, T_n]$, where $T_n = nh_n$ with the asymptotics $h_n \rightarrow 0$, $T_n \rightarrow \infty$ as $n \rightarrow \infty$. Our goal is to estimate the Lévy density $f(z)$ from such discrete observations.

Let $C_r^m(\ell)$ ($r = m + l$, $0 < l \leq 1$, $m \in \mathbb{N} \cup \{0\}$) be the space of real valued bounded functions g , defined on \mathbb{R}^d , which are m times differentiable and such that

$$\left| \frac{\partial g^{(m)}}{\partial z_1^{j_1} \dots \partial z_d^{j_d}}(z) - \frac{\partial g^{(m)}}{\partial z_1^{j_1} \dots \partial z_d^{j_d}}(z') \right| \leq \ell |z - z'|^l$$

for every $z, z' \in \mathbb{R}^d$; $j_1 + \dots + j_d = m$.

Throughout this chapter, we make the following four assumptions.

A 1 $f(z) \in C_r^m(\ell)$ for some $r > 0$ and $m \in \mathbb{N} \cup \{0\}$. Moreover $0 < \int_{\mathcal{E}} f(z) dz < \infty$.

A 2 There exists a positive constant L such that $|a(x) - a(y)| + |b(x) - b(y)| \leq L|x - y|$. Moreover, the coefficient $c(x, z)$ is known, and there exists a function $\zeta(z)$ with $\int_{\mathcal{E}} \zeta^2(z) f(z) dz < \infty$ such that $|c(x, z) - c(y, z)| \leq \zeta(z)|x - y|$ and $|c(x, z)| \leq \zeta(z)(1 + |x|)$.

A 3 $\inf_x |c(x, z)| \geq c_0|z|$ for a constant $c_0 > 0$, and $y = c(x, z)$ has an inverse $z = c^{-1}(x, y)$ such that $\partial_y c^{-1}(x, y)$ is bounded, and that $\partial_x c^{-1}(x, y)|leC|y|$ uniformly in x .

A 4 $\sup_{t \geq 0} E[|X_t|^p] < \infty$ for arbitrary $p \geq 0$.

Remark 5.1 We admit the case where the coefficients $a(x)$ and $b(x)$ are unknown. The assumption that $c(x, z)$ is known; A2, seems to be very restrictive. However if we did not know $c(x, z)$ then it should not be possible to identify Δz , therefore we can not estimate the distribution of Δz . Nevertheless the readers may also be interested in considering the case where c is unknown. In this setting, for example, as $k = 1$, an integral such as $\int_{\mathcal{E}} c^p(X_{t-}, z) f(z) dz$ or $E \left[\int_{\mathcal{E}} c^p(X_{t-}, z) f(z) dz \right]$ can be estimated; see Bandi and Nguyen [5], Shimizu [92].

5.3 Density estimation and the optimal rate

5.3.1 Continuously observed case

Before considering the estimation from sampled data, let us consider the kernel density estimation for continuously observed diffusion processes with jumps.

In the following discussion, we consider the space $H_{m,l}$ of kernels of order (m, l) ($m \in \mathbb{N} \cup \{0\}$, $0 < l \leq 1$), that is, the space of mappings $K : \mathbb{R}^d \rightarrow \mathbb{R}$ bounded, integrable, with the bounded first derivative such that $\int_{\mathbb{R}^d} K(u) du = 1$ and satisfying the following conditions:

$$\int_{\mathbb{R}^d} |(u_1, \dots, u_d)|^l |u_1|^{\alpha_1} \dots |u_d|^{\alpha_d} |K(u_1, \dots, u_d)| du_1 \dots du_d < \infty \quad (5.3)$$

for $\alpha_1, \dots, \alpha_d \in \mathbb{N} \cup \{0\}$; $0 \leq \alpha_1 + \dots + \alpha_d \leq m$, and

$$\int_{\mathbb{R}^d} u_1^{\tilde{\alpha}_1} \dots u_d^{\tilde{\alpha}_d} K(u_1, \dots, u_d) du_1 \dots du_d = 0 \quad (5.4)$$

for $\tilde{\alpha}_1, \dots, \tilde{\alpha}_d \in \mathbb{N} \cup \{0\}$; $1 \leq \tilde{\alpha}_1 + \dots + \tilde{\alpha}_d \leq m$.

Suppose that we can observe whole the path of X in $[0, T]$, and let us denote by τ_i the time point of i th jump of X . In the interval $[0, T]$, we can observe the samples $(\Delta z_{\tau_i})_{1 \leq i \leq N_T}$, where $\Delta z_s = c^{-1}(X_{s-}, \Delta X_s)$, and we can regard them as the randomly observed samples, whose number N_T follows a Poisson distribution with the intensity λT , from a stationary process with the marginal distribution F . Our interest is to estimate the function $f(z) = \lambda F(z)$ from the above samples. A possible estimator is given by the following intuitive discussion: as well as the i.i.d. case, a kernel estimator of F may be given by

$$\begin{aligned} F_T(z) &:= \frac{1}{N_T \delta_T^d} \sum_{i=1}^{N_T} K\left(\frac{z - \Delta z_{\tau_i}}{\delta_T}\right) \\ &= \frac{1}{N_T \delta_T^d} \int_0^T K\left(\frac{z - \Delta z_t}{\delta_T}\right) dN_t, \end{aligned}$$

where $K \in H_{m,l}$ and δ_T is a real sequence which satisfies some conditions. Since $N_T/T \rightarrow \lambda$ ($T \rightarrow \infty$) with probability one, and the amplitude of each jump is independent of the Poisson process N ,

$$f_T(z) := \frac{N_T}{T} F_T(z) = \frac{1}{T \delta_T^d} \int_0^T K\left(\frac{z - \Delta z_t}{\delta_T}\right) dN_t \quad (5.5)$$

may be expected to become an L^2 -consistent estimator of $f(z)$. Actually, Prakasa Rao [78, 79] studied the same problem as $d = 1$. He studied some asymptotic properties of the same type of estimators made by *delta-type* kernels. As $d = 1$, our kernel corresponds to his special case but we deal with the multidimensional case. The following theorem shows the consistency in the MSE sense and the optimal rate of convergence.

Theorem 5.1 *Assume that there exists a real-valued sequence $\{\delta_T\}$ indexed by T such that $\eta_T := \delta_T T^{\frac{1}{2r+d}} \rightarrow \eta$ for a positive constant η . Then, the following inequality is valid for the estimator $f_T(z)$ given in (5.5):*

$$\limsup_{T \rightarrow \infty} \sup_{z \in \mathbb{R}^d} T^{\frac{2r}{2r+d}} E|f_T(z) - f(z)|^2 \leq C(\eta, f, K), \quad (5.6)$$

where

$$\begin{aligned} C(\eta, f, K) &= \eta^{2r} \left(\sum_{j_1 + \dots + j_d = m} \frac{\ell}{j_1! \dots j_d!} \int_{\mathbb{R}^d} |u|^l |u_1|^{j_1} \dots |u_d|^{j_d} |K(u)| du \right)^2 \\ &\quad + \eta^{-d} \sup_{z \in \mathbb{R}^d} |f(z)| \int_{\mathbb{R}^d} K^2(u) du. \end{aligned} \quad (5.7)$$

See Section 5.7.1 on this proof.

Remark 5.2 The convergence rate $T^{\frac{2r}{2r+d}}$ is natural as an analogy to the i.i.d. case; see Ibragimov and Has'minskii [41], Chapter IV. If $d = 1$ then this rate is consistent with the case of the delta-kernel $\delta_T^{-1}K(z\delta_T^{-1})$ as in Prakasa Rao [79], and this is a natural extension to the multidimensional case.

5.3.2 Discretely observed case

We use the judgment proposed in Chapter 3 to discriminate the continuity from the discontinuity of the path in each h_n -time interval. The following lemma which justifies the judgment is a corollary of the Lemma 3.3 in Chapter 3, and the proof is almost the same as that. Therefore we omit the proof.

Lemma 5.1 *Let J_i^n be the number of jumps in an interval $[t_i^n, t_{i-1}^n)$ and we set*

$$\{|\Delta X_i^n| \leq Lh_n^\rho\} = \bigcup_{j=0}^2 C_{i,j}^n, \quad \{|\Delta X_i^n| > Lh_n^\rho\} = \bigcup_{j=0}^2 D_{i,j}^n$$

for constants $L > 0$ and $\rho \in [0, 1/2)$, where

$$\begin{aligned} C_{i,0}^n &= \{J_i^n = 0, |\Delta X_i^n| \leq Lh_n^\rho\}, & D_{i,0}^n &= \{J_i^n = 0, |\Delta X_i^n| > Lh_n^\rho\}, \\ C_{i,1}^n &= \{J_i^n = 1, |\Delta X_i^n| \leq Lh_n^\rho\}, & D_{i,1}^n &= \{J_i^n = 1, |\Delta X_i^n| > Lh_n^\rho\}, \\ C_{i,2}^n &= \{J_i^n \geq 2, |\Delta X_i^n| \leq Lh_n^\rho\}, & D_{i,2}^n &= \{J_i^n \geq 2, |\Delta X_i^n| > Lh_n^\rho\}. \end{aligned}$$

Then, for any $p \geq 1$,

$$\begin{aligned} P\{C_{i,0}^n | \mathcal{F}_{i-1}^n\} &= e^{-\lambda h_n} \tilde{R}(z, h_n^p, X_{t_{i-1}^n}^p), & P\{D_{i,0}^n | \mathcal{F}_{i-1}^n\} &= e^{-\lambda h_n} R(z, h_n^p, X_{t_{i-1}^n}^p), \\ P\{C_{i,1}^n | \mathcal{F}_{i-1}^n\} &= R(z, h_n^{\rho+1}, X_{t_{i-1}^n}^p), & P\{D_{i,1}^n | \mathcal{F}_{i-1}^n\} &= \lambda h_n e^{-\lambda h_n} \tilde{R}(z, h_n^p, X_{t_{i-1}^n}^p), \\ P\{C_{i,2}^n | \mathcal{F}_{i-1}^n\} &\leq \lambda^2 h_n^2, & P\{D_{i,2}^n | \mathcal{F}_{i-1}^n\} &\leq \lambda^2 h_n^2. \end{aligned}$$

The order of the conditional probability of $C_{i,1}^n$ in the above lemma is different from the one of order h_n^3 in Chapter 3. We assumed there that a Lévy density $f(z)$ satisfies $|f(z)| \leq c|z|^\gamma$ ($\gamma > 3$) around the origin for estimation under the asymptotics $nh_n^2 \rightarrow 0$. However, we only assume the boundedness of f around the origin. Therefore, it is possible in our setting that small jumps occur more frequently than the setting in Chapter 3, and it would be more difficult to identify such small jumps. That is why the order of the conditional probability of $C_{i,1}^n$ is smaller than the one in Chapter 3. On the other hand, we have to demand a more rapid experimental design such that $nh_n^{1+\delta} \rightarrow 0$ for a constant $\delta \in (0, 1/2]$ instead of relaxing the assumption on the Lévy density.

Remark 5.3 One might think that the constant L in the filter is redundant since it could be included in the sequence h_n by regarding h_n as $L^{-1/\rho}h_n$. Nevertheless it is convenient to leave this constant L since h_n is the observation interval, which is given in practical data, and we should choose L in accordance with h_n . Actually we encounter a problem that the filter does not work well when n is fixed in dealing with the practical data. Selecting L suitably, we can improve the performance of estimation; see Section 5.4 and Chapter 6.

According to the context in Chapter 3, the amplitude of jump of X in the interval $[t_{i-1}^n, t_i^n)$ can be approximated by the increment $\Delta_i X^n$. Therefore, we can estimate an unobservable underlying jump of z as $\Delta_i z^n := c^{-1}(X_{t_{i-1}^n}, \Delta_i X^n)$, and it would be natural to use the following $f_n(z)$ as an estimator of the Lévy density:

$$f_n(z) := \frac{1}{T_n \delta_n^d} \sum_{i=1}^n K \left(\frac{z - \Delta_i z^n}{\delta_n} \right) \mathbf{1}_{\{|\Delta X_i^n| > L h_n^\rho\}}, \quad (5.8)$$

where $K \in H_{m,l}$ and $\{\delta_n\}_{n \in \mathbb{N}}$ is a real sequence such that $T_n \delta_n^d \rightarrow \infty$. This is a straight discretization of (5.5).

Our main theorem is the following.

Theorem 5.2 *Assume that there exists a constant $\nu \in (0, 2^{-1}(2r+d)(2r+d+1)^{-1})$ such that $T_n h_n^\nu = O(1)$ as $n \rightarrow \infty$, and that there exist a constant $\rho \in [0, 1/2)$ and a real-valued sequence $\{\delta_n\}_{n \in \mathbb{N}}$ such that $\eta_n := \delta_n T_n^{\frac{1}{2r+d}} \rightarrow \eta$ for a positive constant η , $\delta_n h_n^{\rho-1/2} = O(1)$ and $T_n h_n^{1/2} \delta_n^{r-1} = o(1)$ as $n \rightarrow \infty$. Then the following inequality is valid for the estimator $f_n(z)$ given in (5.8):*

$$\limsup_{n \rightarrow \infty} \sup_{z \in \mathbb{R}^d} T_n^{\frac{2r}{2r+d}} E |f_n(z) - f(z)|^2 \leq C(\eta, f, K), \quad (5.9)$$

where $C(\eta, f, K)$ is given in Theorem 5.1 (5.7).

See Section 5.7.2 on this proof.

Remark 5.4 The above convergence rate $T_n^{\frac{2r}{2r+d}}$ is natural since this attains the optimal rate for continuous case. However, without being particular about the optimality, it is possible to improve the order of the experimental design, that is, it is also possible to estimate under the assumption, for example, $T_n h_n^\nu \rightarrow \infty$ but $T_n h_n^{\tilde{\nu}} \rightarrow 0$ for some $\tilde{\nu} \geq 1/2$; see the proof in Section 5.7.2 and the expression (5.32).

Remark 5.5 We sometimes can not check the condition A4. However this condition is one of the sufficient conditions for the above theorem and the estimator $f_n(z)$ sometimes

can work well without this condition if we construct the filter suitably. In Section 5.4.4 below, we describe how to construct the another filter under the case where the diffusion coefficient b is known, and show the performance for that filter.

5.4 Simulation study

For simulation studies, we use the following one dimensional SDE:

$$dX_t = \mu X_t dt + b(X_t, \sigma) dw_t + dz_t, \quad z_t = \sum_{j=1}^{N_t} \varepsilon_j, \quad (5.10)$$

where N_t is a Poisson process with the intensity λ and ε_j 's are i.i.d. r.v's with a density $F(z)$ which satisfies A1.

In Sections 5.4.1 - 5.4.3 below, we consider the case where $b(x, \sigma) = \sigma$. This process is called a Lévy driven Ornstein-Uhlenbeck (O-U) process, in which X is ergodic if $\mu < 0$, and there exists an invariant measure π . It is known that $\sup_t E[|X_t|^p] < \infty$ if $\int_E |z|^p f(z) dz < \infty$ for any $p > 0$ and $\int_{\mathbb{R}^d} |x|^p \eta(dx) < \infty$ for η which is the distribution of X_0 ; see Masuda [67], therefore, A4 are satisfied in this model. Of course, Conditions A2 and A3 are satisfied.

In Section 5.4.4 below, we set $b(x, \sigma) = \sigma x$ in which X is non-ergodic. It is not so easy to check A4, and the condition would be probably unsatisfied. However, we show that our method is robust without this condition if we choose a suitable filter. In this example, we assume that the diffusion coefficient is known, and we propose another filter which is data adaptive.

In each simulation below, we computed the estimated value $f_n(z)$ as a pointwise sample mean based on 500 times experiments. We set $X_0 = 1.0$, $h_n = n^{-0.8}$ and $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ as a kernel throughout this section.

5.4.1 Simple examples

First, let us consider the case where $F(z)$ is the density of Gaussian distribution $\mathcal{N}(0, \nu)$ and $(\mu, \sigma, \lambda, \nu) = (-0.01, 0.01, 15.0, 1.0)$. We choose $\rho = 0.49$, $\delta_n = n^{-0.1}$ and $L = 1.0$. The estimation results are presented in Fig. 5.1. It will be comparatively easy to judge whether a jump occurred or did not in this model; if a jump occurred then it should be large compared with Brownian shocks. We see from Fig. 5.1 that the variance of jumps seems to be overestimated since the small jumps are cut by the filter.

Second, we give another example for the model (5.10) with different parameters such that $(\mu, \sigma, \lambda) = (-0.03, 0.01, 1.0)$, $\rho = 0.49$, but $F(z)$ is a density of the two-sided

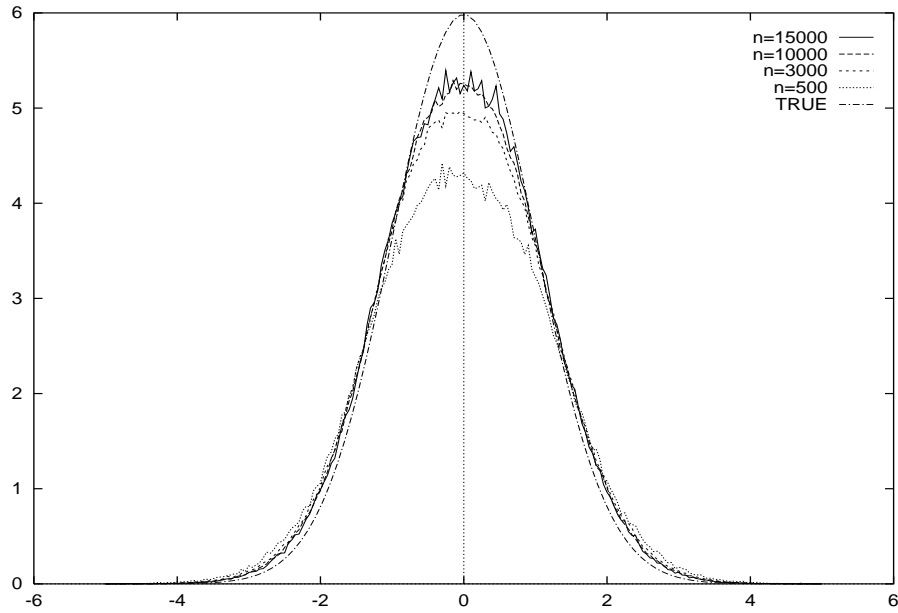


Fig. 5.1: Estimation of the Lévy density $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \sigma, \lambda, \nu) = (-0.01, 0.01, 15.0, 1.0)$, $\rho = 0.49$, $L = 1.0$, $h_n = n^{-0.8}$, $\delta_n = n^{-0.1}$ for sample size 500, 3000, 10000 and 15000 respectively.

gamma distribution $\Gamma(\alpha, \beta)$:

$$F(z) = \frac{1}{2}\beta^\alpha |x|^{\alpha-1} \Gamma^{-1}(\alpha) \exp(-\beta|x|) \quad (5.11)$$

with $\alpha = 4.0$ and $\beta = 5.0$. We show the results with the sample sizes $n = 500, 3000, 10000$, and 15000 in Fig. 5.2. In this simulation, we set $L = 0.2$ in the filter and this seems to discriminate the pure jumps from the diffusion shocks so well.

5.4.2 Some troubles in the estimation

In the first two examples, we took $L = 1.0$ in the former example and $L = 0.2$ in the latter, and that $\rho = 0.49$ in common. How should we choose the constant L and ρ ?

Thanks to Lemma 5.1, it may be better to choose $\rho \in [0, 1/2)$ as large as possible since the larger ρ becomes, the more easily the filter can judge a single jump. Hence we consider the choice of L with ρ fixed as large as possible; in the sequel, we fix $\rho = 0.49$.

On the constant L , we could not choose L by the asymptotic theory. However, if the constant L is chosen unsuitably then the performance of estimation can get worse.

See Fig. 5.3, in which we chose the true value of the diffusion parameter $\sigma (= 0.5)$ larger than the one in the first example, and the other setting is the same. In this

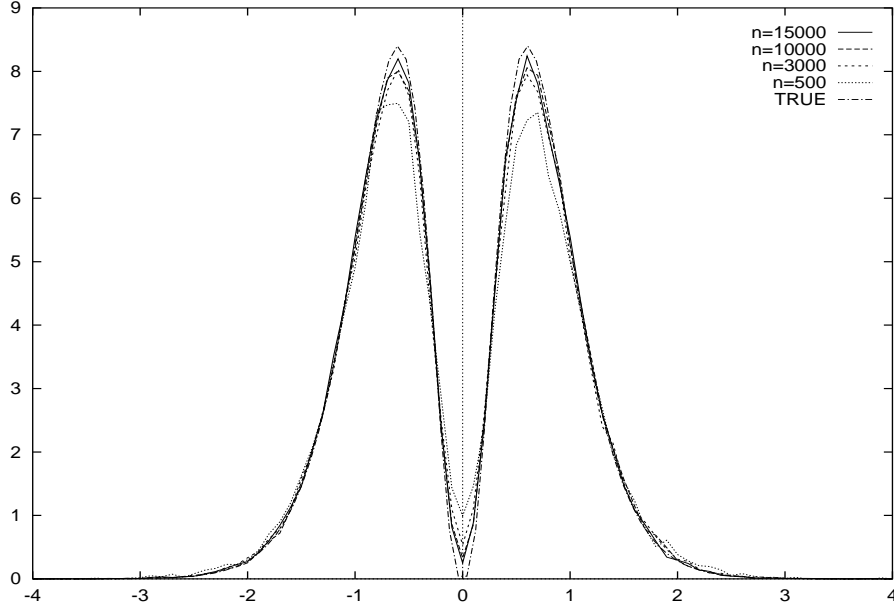


Fig. 5.2: Estimation of the Lévy density $f(z) = \lambda\beta^\alpha|x|^{\alpha-1}\exp(-\beta|x|)/2\Gamma(\alpha)$ with $(\mu, \sigma, \lambda, \alpha, \beta) = (-0.03, 0.01, 15.0, 4.0, 5.0)$, $\rho = 0.49$, $L = 0.2$, $h_n = n^{-0.8}$, $\delta_n = n^{-0.1}$ for sample size 500, 3000, 10000 and 15000 respectively.

case, the estimated densities are awfully overestimated, particularly around the origin. It seems that the filter misunderstands the *large* increments by Brownian shocks with jumps, so the Poisson intensity is overestimated. This is a trouble caused by the observation number n stopped.

Our estimator $f_n(z)$ is theoretically L^2 -consistent estimator and Theorem 5.2 ensures that our filter asymptotically works well for any selection of L . However, the sufficient sample number can depend on the structure of the true model, therefore we have to choose the suitable constant L according to n in each model. Of course, this does not imply that some choices of L possibly lead to the inconsistency of estimators; this can be confirmed, for example, in Fig. 5.6 as mentioned later.

5.4.3 Selection problem of the filter

According to the discussion in Chapter 3, if the process X is ergodic, the estimator

$$\hat{\lambda}_n(L) = \frac{1}{nh_n} \sum_{j=1}^n \mathbf{1}_{\{|\Delta_i X^n| > Lh_n^\rho\}} \quad (5.12)$$

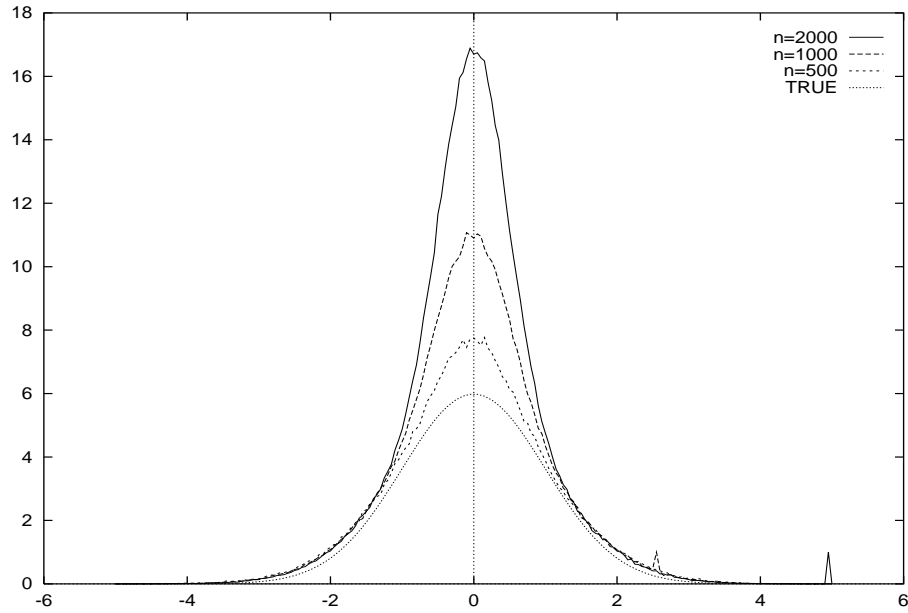


Fig. 5.3: Estimation of the Lévy density $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \sigma, \lambda, \nu) = (-0.01, 0.5, 15.0, 1.0)$, $\rho = 0.49$, $L = 1.0$, $h_n = n^{-0.8}$, $\delta_n = n^{-0.1}$ for sample size 500, 1000 and 2000 respectively.

can be one of estimators of the intensity λ . Fig. 5.4 shows the graphs of the L -pointwise sample mean of $\hat{\lambda}_n(L)$ based on 1000 simulations for $n = 500, 3000$ and 10000 ; it corresponds to the approximation of the curve of $E[\hat{\lambda}_n(L)]$. It is natural that $\hat{\lambda}_n(L)$ decreases rapidly in *small* L , and slowly in *large* L since the filter with *small* L captures the increments by diffusions, whose order of numbers is $O(n)$, as well as by the true jumps, and the filter with *large* L hardly captures the increments by diffusions but some *large* jumps, whose order of numbers is $O(nh_n)$. Therefore, intuitively speaking, it might be better to choose the *smallest* possible L at which the curve becomes *nearly* flat. It implies that the filter excluded influences by diffusions as much as possible, but does not exclude too much the true jumps. To see this, we presented in Fig. 5.5 the graphs of the numerical derivatives in L of the estimated intensity curve. In order to avoid blurring of the pointwise calculated numerical derivatives, we plotted the five-points moving average curves.

From the intuitive point of view as above, it is better to choose L at which the derivative is *nearly* zero visually. In fact, by evaluating $E[\hat{\lambda}_n(L)]$ analytically, we find that such derivatives can approximately decrease in the same way as the tail of the Lévy density $f(z)$ after that the filter excludes the influences by Brownian shocks:

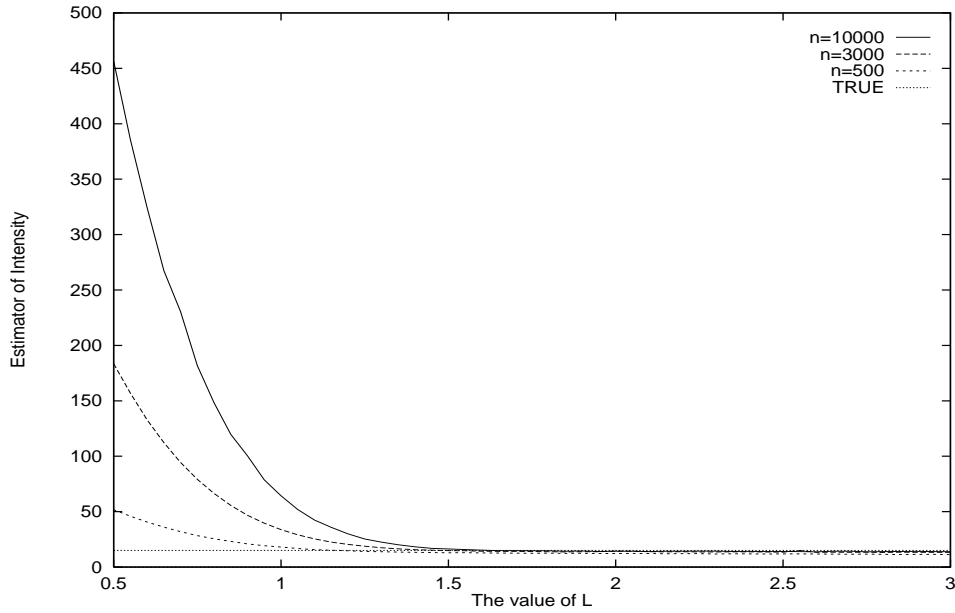


Fig. 5.4: Estimation of the intensity λ as $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \sigma, \lambda, \nu) = (-0.01, 0.5, 15.0, 1.0)$ for sample size 500, 3000 and 10000 respectively.

$\frac{d}{dL}E[\hat{\lambda}_n(L)] \approx 2e^{-\lambda h_n}f(Lh_n^\rho)$ for large L ; Lemma 6.1 in Chapter 6. Therefore the derivatives can not be zero generally. However, if $F(z)$ has the light tail as the normal density and $\lambda \ll n$, then the curve of the derivatives look like zero compared with the enormous influences by diffusions.

Remark 5.6 Though we calculated the L -pointwise sample mean of $\hat{\lambda}_n(L)$ based on 1000 simulations in Fig. 5.4, we could not calculate such a sample mean from the real data which is obtained from an one sample path. However, if n is sufficiently large, then we can use the estimator $\hat{\lambda}_n(L)$ from one sample path instead of $E[\hat{\lambda}_n(L)]$ since $|\hat{\lambda}_n(L) - E[\hat{\lambda}_n(L)]| \rightarrow 0$ ($n \rightarrow \infty$) in probability for each L .

We can choose, for example, that $(n, L) = (500, 1.2)$, $(3000, 1.6)$ and $(10000, 1.7)$ in view of Fig. 5.5, and the results are shown in Fig. 5.7. We also show the case with $(n, L) = (15000, 1.7)$ in that figure for the purpose of reference. Though we might not point out the optimal L exactly, the performances are dramatically improved compared with in Fig. 5.3. It would be useful in some practical situations although this is completely intuitive method. However, the constant L should be chosen from more statistical point of view. A suitable method has never established yet, and it is a problem to be studied for the future. However, in a simple model, we will discuss a

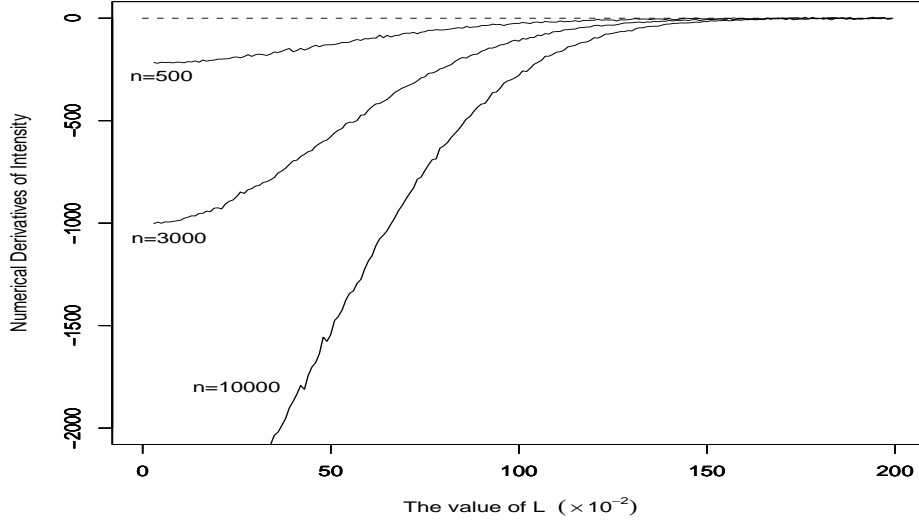


Fig. 5.5: The five-points moving average curves of the numerical derivatives of $\hat{\lambda}_n(L)$ in L for sample size 500, 3000 and 10000 respectively.

statistical filter selection in Chapter 6.

For the purpose of reference, we presented the graphs of $\hat{\lambda}_n(L)$ for the model (5.11) in Fig. 5.6. We will be able to guess that it would suffice to choose, for example, $0.15 \leq L \leq 0.2$ for each sample size. As already mentioned, $L = 0.2$ was enough.

The view of Fig. 5.6 also implies that our filter works asymptotically well for much *smaller* L than 0.2. Look at the curves around the interval $[0.06, 0.07]$ for each n . The intensity for $n = 3000$ is estimated larger than for $n = 500$ by some misjudgments, however, it appears to be closer to the true value as $n = 10000$ than the case where $n = 500$. This implies that the consistency is true for the different selection of L . In the example of Fig. 5.4, we could not show such phenomenon visually because of the too slow convergence of estimator. Intuitively speaking, the probability of misjudgment of jumps is about

$$p_n := P\{|\sigma w_{h_n}| > Lh_n^\rho\} = 2[1 - \Phi(\sigma^{-1}Lh_n^{\rho-1/2})], \quad \rho \in [0, 1/2).$$

In the setting of Fig. 5.4, $p_{10000} = 0.031326$ and $p_{15000} = 0.031318$, therefore the misjudgment is hardly improved at all. We would need to choose the sample size n enormously large to demonstrate the consistency of f_n regardless of the selection of L in that setting, and it would exceed the ability of the usual calculator. However it must be possible to obtain the same result as in Fig. 5.6.

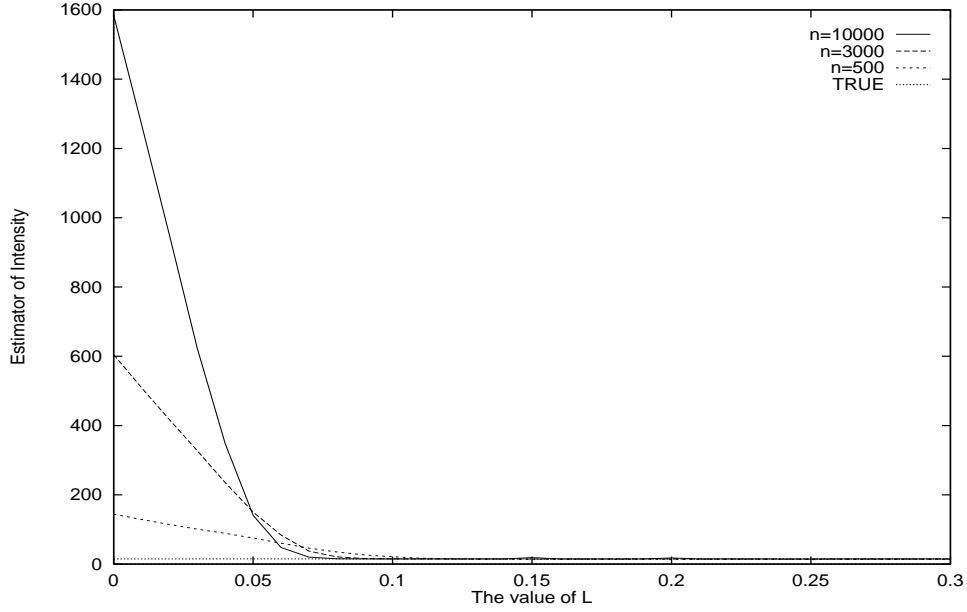


Fig. 5.6: Estimation of the intensity λ as $f(z) = 2^{-1}\lambda\beta^\alpha|x|^{\alpha-1}\Gamma^{-1}(\alpha)\exp(-\beta|x|)$ with $(\mu, \sigma, \lambda, \alpha, \beta) = (-0.03, 0.01, 15.0, 4.0, 5.0)$ for sample size 500, 3000 and 10000 respectively.

Remark 5.7 Although we discussed the selection method for L in the case where X is ergodic, we recently find that the estimator $\hat{\lambda}_n(L)$ is consistent to the true λ even if X is non-ergodic; see Shimizu [95]. Therefore the above method could be available when X is non-ergodic.

5.4.4 Data adaptive filter

In the previous section, we described the case where the diffusion coefficient is a constant. When the diffusion coefficient depends on X_t ; for example $b(x) = \sigma x$, some simulations show that the behaviors of $\hat{\lambda}_n(L)$ in L are unstable. However, if the diffusion coefficient $b(x)$ is known then we can construct the data adaptive filter which improves the performance of the estimation.

In this section, we consider the special case where the coefficient b is known. Although the method described below is just a numerical one, it would sometimes useful in some applications.

One can see by the discussion in Chapter 3 that it may be possible to replace a constant L by \mathcal{F}_t -adapted process $L_t(h_n)$ which satisfies some moment conditions and $L_{t_{i-1}^n}(h_n) = O_p(h_n^\rho)$ as $n \rightarrow \infty$, and to make the filter $\{|\Delta_i X^n| > L_{t_{i-1}^n}(h_n)\}$.

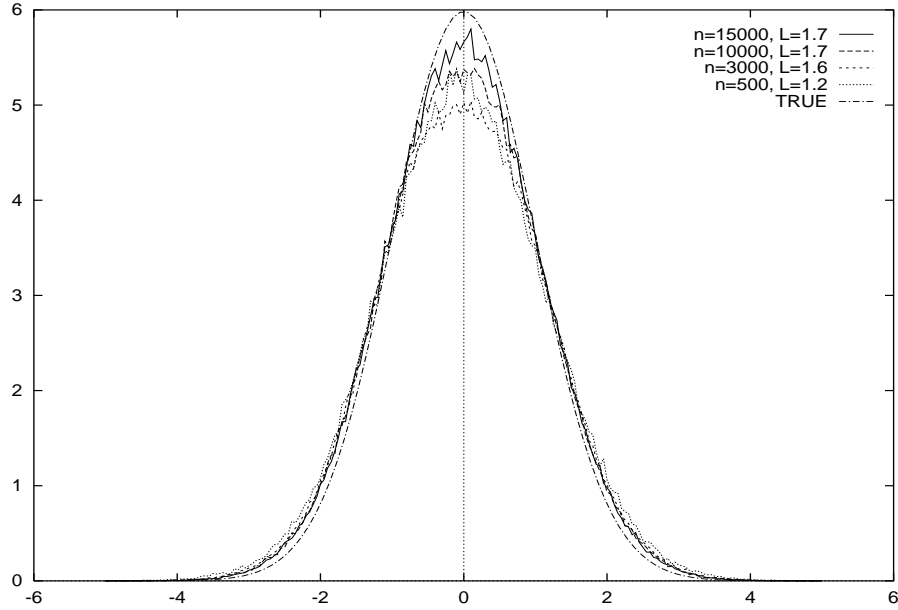


Fig. 5.7: Estimation of the Lévy density $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \sigma, \lambda, \nu) = (-0.01, 0.5, 15.0, 1.0)$, $\delta_n = n^{-0.1}$ for sample size 500, 3000 and 10000 respectively.

Suppose that $b^2(X_{t_{i-1}^n}) > 0$. Note that, in one dimensional case,

$$N_{i,n} := (\Delta_i X^n - a(X_{t_{i-1}^n})h_n)(b^2(X_{t_{i-1}^n})h_n)^{-1/2}$$

approximately follows the standard normal distribution, which satisfies for large n that

$$P \left\{ \left| \frac{\Delta_i X^n - a(X_{t_{i-1}^n})h_n}{\sqrt{b^2(X_{t_{i-1}^n})h_n}} \right| > u_{\alpha_n/2} \right\} \approx \alpha_n \approx P \left\{ \frac{|\Delta_i X^n|}{\sqrt{b^2(X_{t_{i-1}^n})h_n}} > u_{\alpha_n/2} \right\},$$

where u_z is the z -percentile of the standard normal distribution. Therefore

$$L_{t_{i-1}^n} = u_{\alpha_n/2} \sqrt{b^2(X_{t_{i-1}^n})h_n} \vee h_n^\rho \quad (5.13)$$

will be expected to improve the judgment of jumps. Here, we should choose the level α_n so that it depends on the observation number n since the extent of misjudgments depends on n .

Let us consider the following model, which is a non-ergodic case:

$$dX_t = \mu X_t dt + \sigma_0 X_t dw_t + dz_t,$$

with the Lévy density $f(z) = \frac{\lambda}{\sqrt{2\pi\nu}} \exp\left(-\frac{z^2}{2\nu}\right)$ and $(\mu, \lambda, \nu) = (-0.01, 15.0, 1.0)$. We assume that $\sigma_0 = 0.5$ is known. We chose α_n as $\alpha_n n = 15$ for $n = 500, 3000$, and 10000 .

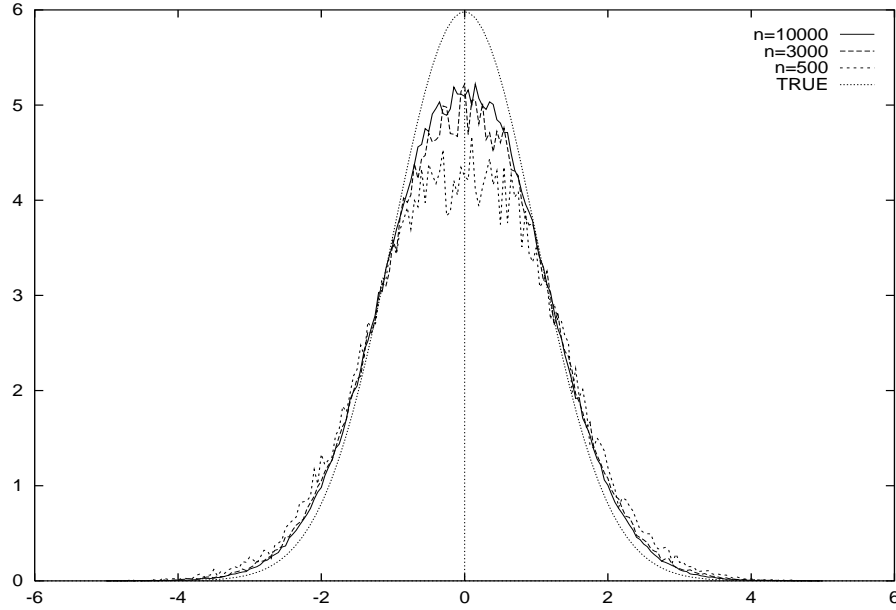


Fig. 5.8: Estimation of the Lévy density in the model $dX_t = \mu X_t dt + \sigma_0 X_t dw_t + dz_t$ ($\sigma_0 = 0.5$ is known) and $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \lambda, \nu) = (-0.01, 15.0, 1.0)$, $\rho = 0.49$, $h_n = n^{-0.8}$, $\delta_n = n^{-0.1}$ and $\alpha_n n = 15$ for sample size 500, 3000 and 10000 respectively.

This implies that respectively 3%, 0.5%, and 0.15% of the *large* shocks by diffusions are misjudged as jumps.

The estimation results are presented in Fig. 5.8. It seems that the influences by diffusions are well excluded. However, it is not necessarily the case where f_n becomes a good estimator as n increases with $\alpha_n n = 15$. The next Fig. 5.9 shows the results of the case $100\alpha_n = 0.1(\%)$ and $0.15(\%)$ as $n = 15000$. As $100\alpha_n = 0.15(\%)$, it seems that f_n is a good estimator. On the other hand, as $100\alpha_n = 0.1(\%)$ ($\alpha_n n = 15$), the filter cuts too many *small* jumps, and f_n underestimates around the origin.

Intuitively speaking, the mean of the number of misjudgments is roughly $e_n = \alpha_n(n - \lambda n h_n) \approx \alpha_n n$ for sufficiently large n and $\lambda h_n \ll 1$, but the level e_n should be chosen by an observer suitably so that those errors will not influence on estimation. However an observer could not determine the level without prior information. Therefore there remains room to consider how to choose the level e_n , and this should be done for the future.

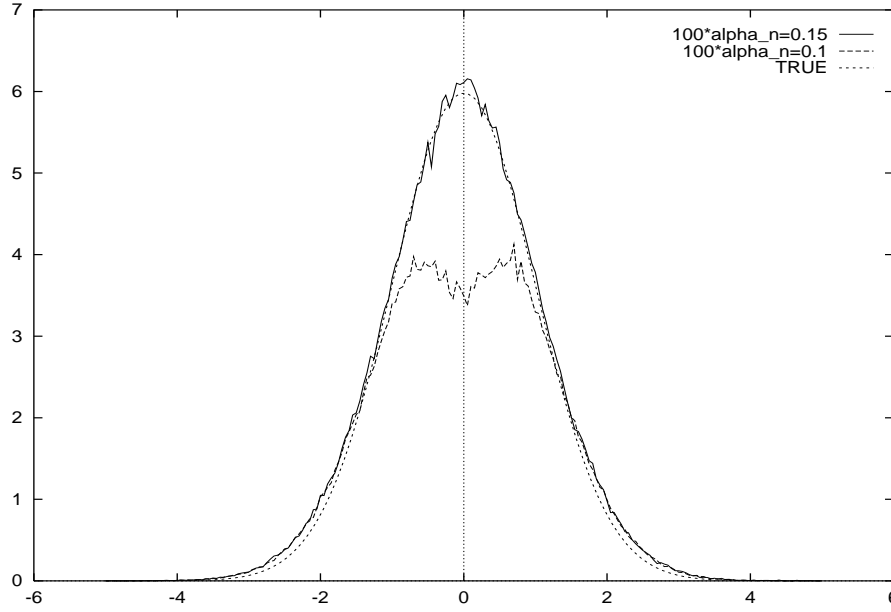


Fig. 5.9: Estimation of the Lévy density in the model $dX_t = \mu X_t dt + \sigma_0 X_t dw_t + dz_t$ ($\sigma_0 = 0.5$ is known) and $f(z) = \lambda(\sqrt{2\pi\nu})^{-1} \exp(-(2\nu)^{-1}z^2)$ with $(\mu, \lambda, \nu) = (-0.01, 15.0, 1.0)$, $\rho = 0.49$, $h_n = n^{-0.8}$, $\delta_n = n^{-0.1}$ and $n = 15000$ for $100\alpha_n = 0.15$ and 0.01 (%).

5.5 Conclusion of this chapter

This chapter tackles the identification of the Lévy density, which completely controls the behavior of jumps, from the frequently observed data. Though we presented the inference from continuous observations in Theorem 5.1, this should be a benchmark for the inference from discrete observations, and it should certainly be considered before the inference from discrete observations.

One of the advantages of this chapter is that we do not need to mind whether the continuous part is known or unknown, and whether the model is ergodic or non-ergodic which is often important in the parametric setting. Therefore the theory in this chapter can be applied to a wide class of models in these sense. However there are some annoying restriction on the order of h_n ; the experimental design and δ_n ; the bandwidth: the conditions

$$T_n h_n^\nu = O(1), \quad T_n \delta_n^{2r+d} = O(1), \quad \delta_n h_n^{\rho-1/2} = O(1), \quad T_n h_n^{1/2} \delta_n^{r-1} = o(1), \quad (5.14)$$

should be satisfied as n tends to infinity, where $T_n = nh_n$, $\nu \in (0, \frac{2r+d}{2(2r+d+1)})$, $\rho \in [0, \frac{1}{2})$, d is the dimension of the process and r is the smoothness of the Lévy density.

We used the asymptotic filter $\{|\Delta_i X^n| > Lh_n^\rho\}$ for a constant L . Though this filter asymptotically works well for arbitrary constant L in theory, their capabilities actually depend on the constant L . Therefore it is critical to choose L under the practical situation that the number of observations n is fixed.

What we can learn from the results of simulations is, although it is intuitively clear, that it is easy to identify the existence of jumps if the absolute value of the diffusion coefficient is relatively small compared with the amplitudes of jumps. Actually, in two examples presented in Section 5.4.1, the estimator $f_n(z)$ does not diverge as long as we do not choose an L which is extremely small. However, as described in the example in Section 5.4.2, it becomes difficult to choose L when increments by diffusion shocks are as large as the ones by jumps. As a consequence, it is possible that the asymptotic theory does not work well even under enormous sample size.

This chapter proposed some intuitive methods to choose the constant L from the data. The selection method of L in Section 5.4.3 focuses attention on the difference between the order of the frequency of continuous noise and that of discontinuous breaks. In consequence of many simulations, we at least find the following:

- In ergodic cases:
 - Our method can find the suitable constant L .
- In non-ergodic cases:
 - If the diffusion coefficient is uniformly bounded on the state space, then our method can find the suitable constant L .
 - If the diffusion coefficient is unbounded Markovian type as in Section 5.4.4, it is difficult to choose L because of the unstable behavior of the index (5.12).

In the last case, we can exclude the influence of diffusions by using the predictable $L_{t_{i-1}^n}$ as shown in (5.13) if the diffusion coefficient is known. However there remains a problem to determine α_n ; the rate of misjudgments.

The assumptions for Lévy densities in this paper are little restrictive, that is, $f(z)$ is bounded and continuous. However it is also important in practice to consider Lévy densities which diverge at the origin and is not integrable as in Chapter 4. In this case, the construction of new filters would become a major problem.

5.6 Some moment estimates

In this section, we make the same assumptions as in Theorem 5.2, and we use the following notations: for fixed $z \in \mathbb{R}^d$ and $q > 0$,

$$k_n(z; x) := K\left(\frac{z-x}{\delta_n}\right), \quad g_{n,q}(z) := \int_{\mathbb{R}^d} \frac{1}{\delta_n^d} K^q\left(\frac{z-u}{\delta_n}\right) f(u) du$$

$$\begin{aligned} K_{i,n,q}^{(1)}(z) &:= k_n^q(z; \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} - k_n^q(z; \Delta_i z^n) \mathbf{1}_{D_{i,1}^n}, \\ K_{i,n,q}^{(2)}(z) &:= k_n^q(z; \Delta_i z^n) \mathbf{1}_{D_{i,1}^n} - k_n^q(z; \Delta z_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n}, \\ K_{i,n,q}^{(3)}(z) &:= k_n^q(z; \Delta z_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n} - k_n^q(z; \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n = 1\}}, \end{aligned}$$

where τ_i^n is given in (5.30). Moreover we simply write $K_{i,n}^{(j)}(z)$ for $K_{i,n,1}^{(j)}(z)$.

We prepare some useful lemmas when we show Theorem 5.1 and 5.2.

Lemma 5.2 *For any $p, q \geq 1$, it follows that*

$$E_{i-1}^n[|K_{i,n,q}^{(1)}(z)|^p] = R(z, h_n^2, X_{t_{i-1}^n}), \quad (5.15)$$

$$E_{i-1}^n[|K_{i,n,q}^{(2)}(z)|^p] = R(z, h_n^{p/2+1} \delta_n^{-p}, X_{t_{i-1}^n}), \quad (5.16)$$

$$E_{i-1}^n[|K_{i,n,q}^{(3)}(z)|^p] = R(z, h_n^{\rho+1}, X_{t_{i-1}^n}). \quad (5.17)$$

Proof . On the equation (5.15), noticing that k_n is bounded, we have

$$\begin{aligned} E_{i-1}^n[|K_{i,n,q}^{(1)}(z)|^p] &\leq C_{p,q} E_{i-1}^n[|\mathbf{1}_{D_{i,0}^n} + \mathbf{1}_{D_{i,2}^n}|^p] \\ &\leq C_{p,q} (P_{i-1}^n\{D_{i,0}^n\} + P_{i-1}^n\{D_{i,2}^n\}) \\ &= R(z, h_n^2, X_{t_{i-1}^n}). \end{aligned}$$

The equation (5.17) is also similarly proved by using Lemma 5.1, that is,

$$E_{i-1}^n[|K_{i,n,q}^{(3)}(z)|^p] \leq C_{p,q} P_{i-1}^n\{C_{i,1}^n\} = R(z, h_n^{\rho+1}, X_{t_{i-1}^n}).$$

On the equation (5.16), applying the mean value theorem to k_n^q , we have

$$\begin{aligned} &E_{i-1}^n[|K_{i,n,q}^{(2)}(z)|^p] \\ &\leq C_p E_{i-1}^n \left[\left| \delta_n^{-1} \partial_x k_n^q(z; \tilde{z})^* \Delta X_{\tau_i^n} \right|^p \left(|\Delta X_i^n - \Delta X_{\tau_i^n}|^p + |X_{\tau_i^n} - X_{t_{i-1}^n}|^p \right) \mathbf{1}_{D_{i,1}^n} \right], \\ &\leq C_p \delta_n^{-1} E_{i-1}^n \left[\left| \Delta X_{\tau_i^n} \right|^p \left(|\Delta X_i^n - \Delta X_{\tau_i^n}|^p + |X_{\tau_i^n} - X_{t_{i-1}^n}|^p \right) \mathbf{1}_{\{J_i^n = 1\}} \right] \times \\ &\quad \times P\{J_i^n = 1\}, \end{aligned}$$

where \tilde{z} is a random variable which values between $\Delta_i z^n$ and $\Delta z_{\tau_i^n}$. Since

$$\begin{aligned} E_{i-1}^n \left[\left| X_{\tau_i^n} - X_{t_{i-1}^n} \right|^p \mathbf{1}_{D_{i,1}^n} \right] &\leq E_{i-1}^n \left[\left| X_{\tau_i^n} - X_{t_{i-1}^n} \right|^p \mathbf{1}_{\{J_i^n=1\}} \right] \\ &= R(z, h_n^{p/2+1}, X_{t_{i-1}^n}) \end{aligned}$$

by the same argument as for the evaluation of I_1 in the proof of Proposition 3.4 in Chapter 3, and similarly

$$E_{i-1}^n \left[\left| X_{t_i^n} - X_{\tau_i^n} \right|^p \mathbf{1}_{D_{i,1}^n} \right] = R(z, h_n^{p/2+1}, X_{t_{i-1}^n}),$$

we obtain that $E_{i-1}^n[|K_{i,n,q}^{(2)}(z)|^p] = R(z, h_n^{p/2+1} \delta_n^{-p}, X_{t_{i-1}^n})$. \square

Lemma 5.3 *For any $p > 0$, it follows that*

$$\begin{aligned} E_{i-1}^n [K_{\delta_n}^p(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}] \\ = h_n e^{-\lambda h_n} \delta_n^{-d(p-1)} g_{n,p}(z) + R(z, h_n^{3/2} \delta_n^{-(pd+1)}, X_{t_{i-1}^n}). \end{aligned}$$

Proof . Notice that

$$\begin{aligned} E_{i-1}^n [K_{\delta_n}^p(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}] \\ = \delta_n^{-pd} \left\{ E_{i-1}^n [K_{i,n,p}^{(1)}(z)] + E_{i-1}^n [K_{i,n,p}^{(2)}(z)] + E_{i-1}^n [K_{i,n,p}^{(3)}(z)] \right\} \\ + E_{i-1}^n [\delta_n^{-pd} k_n^p(z; \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}}]. \end{aligned}$$

Since $(\Delta z_{\tau_i^n}, J_i^n)$ and \mathcal{F}_{i-1}^n are independent for each i , we see that

$$\begin{aligned} E_{i-1}^n [K_{\delta_n}^p(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}] \\ = \delta_n^{-d(p-1)} E_{i-1}^n \left[E \left[\delta_n^{-d} k_n^p(s; \Delta z_{\tau_i^n}) | J_i^n = 1 \right] \mathbf{1}_{\{J_i^n=1\}} \right] \\ + R(z, h_n^{3/2} \delta_n^{-(pd+1)}, X_{t_{i-1}^n}) (\sqrt{h_n} \delta_n + 1 + \delta_n h_n^{\rho-1/2}) \\ = \delta_n^{-d(p-1)} P\{J_i^n = 1\} \int_{\mathbb{R}^d} \frac{1}{\delta_n^d} K^p \left(\frac{z-u}{\delta_n} \right) \lambda^{-1} f(u) du \\ + R(z, h_n^{3/2} \delta_n^{-(pd+1)}, X_{t_{i-1}^n}) (\sqrt{h_n} \delta_n + 1 + \delta_n h_n^{\rho-1/2}) \\ = h_n e^{-\lambda h_n} \delta_n^{-d(p-1)} g_{n,p}(z) + R(z, h_n^{3/2} \delta_n^{-(pd+1)}, X_{t_{i-1}^n}). \quad \square \end{aligned}$$

Remark 5.8 We can easily obtain the following equalities by the same argument as above.

$$E_{i-1}^n [K_{\delta_n}^p(z - \Delta_i z^n) \mathbf{1}_{D_{i,1}^n}]$$

$$= h_n e^{-\lambda h_n} \delta_n^{-d(p-1)} g_{n,p}(z) + R(z, h_n^{3/2} \delta_n^{-(pd+1)}, X_{t_{i-1}^n}), \quad (5.18)$$

$$\begin{aligned} & E_{i-1}^n [K_{\delta_n}^p(z - \Delta z_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n}] \\ &= h_n e^{-\lambda h_n} \delta_n^{-d(p-1)} g_{n,p}(z) + R(z, h_n^{1+\rho} \delta_n^{-pd}, X_{t_{i-1}^n}). \end{aligned} \quad (5.19)$$

We shall omit these proofs.

Lemma 5.4 *Let $Q(x)$ be a real valued function defined on \mathbb{R}^d which is of polynomial growth. We set $Q_j := Q(X_{t_j^n})$. Then it follows for any $p > 1$ and $i < j$ that*

$$E_{i-1}^n [Q_{j-1} K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}] = R(z, h_n^{1/p} \delta_n^{-d}, X_{i-1}^n), \quad (5.20)$$

$$E_{i-1}^n [Q_{j-1} K_{i,n}^{(1)}(z)] = R(z, h_n^{2/p}, X_{i-1}^n), \quad (5.21)$$

$$E_{i-1}^n [Q_{j-1} K_{i,n}^{(2)}(z)] = R(z, h_n^{1/2+1/p} \delta_n^{-1}, X_{i-1}^n), \quad (5.22)$$

$$E_{i-1}^n [Q_{j-1} K_{i,n}^{(3)}(z)] = R(z, h_n^{(\rho+1)/p}, X_{i-1}^n). \quad (5.23)$$

Proof . Notice that $E[Q_{j-1} | \mathcal{F}_{i-1}^n] = R(z, 1, X_{i-1}^n)$ by Lemma 3.5 (3.17). Using Lemma 5.2, 5.3 and Hölder's inequality, we can obtain that (5.20) - (5.23) by the straightforward calculation. \square

Lemma 5.5 *For $i < j$ and arbitrary $\mu \in (0, 1)$, it follows that*

$$\begin{aligned} & E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right] \\ &= h_n^2 \left\{ e^{-2\lambda h_n} g_{n,1}^2(z) + R(z, h_n^{1/2-\mu} \delta_n^{-(d+1+\mu d)}, X_{i-1}^n) \right\}. \end{aligned}$$

Proof . Since $(\Delta z_{\tau_i^n}, J_i^n)$, $(\Delta z_{\tau_j^n}, J_j^n)$ and \mathcal{F}_{i-1}^n are independent each other for $i < j$, we see that

$$\begin{aligned} M_n &:= E_{i-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}} K_{\delta_n}(z - \Delta z_{\tau_j^n}) \mathbf{1}_{\{J_j^n=1\}} \right] \\ &= \left\{ \int_{\mathbb{R}^d} \frac{1}{\delta_n^d} K\left(\frac{z-u}{\delta_n}\right) \lambda^{-1} f(u) du \cdot \lambda h_n e^{-\lambda h_n} \right\}^2 \\ &= h_n^2 e^{-2\lambda h_n} g_{n,1}^2(z). \end{aligned}$$

Let $L_n := E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right]$. Then

$$\begin{aligned} \delta^d (L_n - M_n) &= E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} E_{j-1}^n \left[K_{j,n}^{(1)}(z) \right] \right] \\ &\quad + E_{i-1}^n \left[E_{j-1}^n \left[K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{D_{j,1}^n} \right] K_{i,n}^{(1)}(z) \right] \\ &\quad + E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{D_{i,1}^n} E_{j-1}^n \left[K_{j,n}^{(2)}(z) \right] \right] \end{aligned}$$

$$\begin{aligned}
& + E_{i-1}^n \left[E_{j-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_j^n}) \mathbf{1}_{D_{j,1}^n} \right] K_{i,n}^{(2)}(z) \right] \\
& + E_{i-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n} E_{j-1}^n \left[K_{j,n}^{(3)}(z) \right] \right] \\
& + E_{i-1}^n \left[E_{j-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_j^n}) \mathbf{1}_{\{J_j^n=1\}} \right] K_{i,n}^{(3)}(z) \right].
\end{aligned}$$

From (5.18) and (5.19) with $p = 1$, we find that

$$\begin{aligned}
E_{j-1}^n \left[K_{\delta_n}(z - \Delta z_j^n) \mathbf{1}_{D_{j,1}^n} \right] &= R(z, h_n, X_{t_{j-1}^n}), \\
E_{j-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_j^n}) \mathbf{1}_{D_{j,1}^n} \right] &= R(z, h_n, X_{t_{j-1}^n}), \\
E_{j-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_j^n}) \mathbf{1}_{\{J_j^n=1\}} \right] &= R(z, h_n, X_{t_{j-1}^n})
\end{aligned}$$

since $g_{n,1}(z)$ is bounded by Bochner's lemma. Hence we obtain that

$$L_n - M_n = E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > L h_n^\rho\}} R(z, h_n^2 \delta_n^{-d}, X_{t_{j-1}^n}) \right] \quad (5.24)$$

$$+ E_{i-1}^n \left[R(z, h_n \delta_n^{-d}, X_{t_{j-1}^n}) K_{i,n}^{(1)}(z) \right] \quad (5.25)$$

$$+ E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{D_{i,1}^n} R(z, h_n^{3/2} \delta_n^{-(d+1)}, X_{t_{j-1}^n}) \right] \quad (5.26)$$

$$+ E_{i-1}^n \left[R(z, h_n \delta_n^{-d}, X_{t_{j-1}^n}) K_{i,n}^{(2)}(z) \right] \quad (5.27)$$

$$+ E_{i-1}^n \left[K_{\delta_n}(z - \Delta z_{\tau_i^n}) \mathbf{1}_{D_{i,1}^n} R(z, h_n^{\rho+1} \delta_n^{-d}, X_{t_{j-1}^n}) \right] \quad (5.28)$$

$$+ E_{i-1}^n \left[R(z, h_n \delta_n^{-d}, X_{t_{j-1}^n}) K_{i,n}^{(3)}(z) \right]. \quad (5.29)$$

On the term (5.26), using Hölder's inequality and (5.18),

$$\begin{aligned}
& E_{i-1}^n \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{D_{i,1}^n} R(z, h_n^{3/2} \delta_n^{-(d+1)}, X_{t_{j-1}^n}) \right] \\
& \leq C h_n^{3/2} \delta_n^{-(d+1)} \left\{ E_{i-1}^n \left[K_{\delta_n}^p(z - \Delta_i z^n) \mathbf{1}_{D_{i,1}^n} \right] \right\}^{1/p} \\
& = R(z, \alpha_n, X_{t_{j-1}^n})
\end{aligned}$$

for any $p > 1$, where $\alpha_n = h_n^{3/2+1/p} \delta_n^{-(d+1)-(p-1)d/p}$. Similarly, we can obtain that (5.24) and (5.28) are also $R(z, \alpha_n, X_{i-1}^n)$ since

$$\begin{aligned}
R(z, h_n^2 \delta_n^{-d}, X_{t_{j-1}^n}) &= \delta_n h_n^{1/2} R(z, h_n^{3/2} \delta_n^{-(d+1)}, X_{i-1}^n), \\
R(z, h_n^{\rho+1} \delta_n^{-d}, X_{i-1}^n) &= \delta_n h_n^{\rho-1/2} R(z, h_n^{3/2} \delta_n^{-(d+1)}, X_{i-1}^n).
\end{aligned}$$

Moreover we see that (5.27) is also $R(z, \alpha_n, X_{i-1}^n)$ by (5.22). Therefore, applying lemma 5.4 to (5.25) and (5.29), we have

$$L_n - M_n = R(z, h_n^{1+2/p} \delta_n^{-d}, X_{i-1}^n) + R(z, h_n^{1+(\rho+1)/p'} \delta_n^{-d}, X_{i-1}^n) + R(z, \alpha_n, X_{i-1}^n)$$

$$= \alpha_n \{ R(z, h_n^{1/p-1/2} \delta_n^{1+d-1/p}, X_{t_{i-1}^n}) + R(z, \beta_n, X_{t_{i-1}^n}) + R(z, 1, X_{t_{i-1}^n}) \}$$

for any $p, p' > 1$, where $\beta_n := \alpha_n^{-1} h_n^{1+(\rho+1)/p'} \delta_n^{-d}$. Let $1/p = 1 - \mu$ and $(\rho + 1)/p' = \rho + 1 - \mu'$ for $\mu, \mu' \in (0, 1)$, then

$$\beta_n = (h_n^{\rho-1/2} \delta_n) \cdot h_n^{\mu-\mu'} \delta_n^{1+\mu d}.$$

Hence, if we take $\mu \geq \mu'$ then $\beta_n \rightarrow 0$. Consequently,

$$\begin{aligned} L_n &= M_n + R(z, \alpha_n, X_{i-1}^n) \\ &= M_n + h_n^2 R(z, h_n^{1/2-\mu} \delta_n^{-(d+1+\mu d)}, X_{i-1}^n). \end{aligned}$$

We can take $\mu \in (0, 1)$ arbitrary since $\mu = 1 - 1/p$ for any $p > 1$. This completes the proof. \square

5.7 Proofs of the main theorems

5.7.1 Proof of Theorem 5.1

The proof follows the usual way of decomposing the MSE into bias and variance components and to show that they converge with optimal rate or faster. Notice the following decomposition.

$$E|f_T(z) - f(z)|^2 = b_T^2(z) + V f_T(z),$$

where $b_T(z) = E f_T(z) - f(z)$, $V f_T(z) = E f_T^2(z) - \{E f_T(z)\}^2$.

Note that $N_t - \lambda t$ is \mathcal{F}_t -martingale. It follows from Theorem 3.2 in Prakasa Rao [79] that, for $K_{\delta_T}(z - x) = \frac{1}{\delta_T^d} K\left(\frac{z - x}{\delta_T}\right)$,

$$\begin{aligned} E f_T(z) &= \frac{1}{T} E \left[\int_0^T K_{\delta_T}(z - \Delta z_t) dN_t \right] \\ &= \frac{\lambda}{T} \int_0^T E [K_{\delta_T}(z - \Delta z_t) | |\Delta z_t| > 0] dt \\ &= \frac{1}{\delta_T^d} \int_{\mathbb{R}^d} K\left(\frac{z - u}{\delta_T}\right) \lambda F(u) du \\ &= \int_{\mathbb{R}^d} K(v) f(z - \delta_T v) dv. \end{aligned}$$

As $m \neq 0$, by the same argument as in the proof of Theorem 4.1 in Bosq [10], we easily obtain that

$$|b_T(z)| \leq c_{(r,m)} \delta_T^{m+l},$$

where $c_{(r,m)} = \sum_{j_1+\dots+j_d=m} \frac{\ell}{j_1! \dots j_d!} \int_{\mathbb{R}^d} |u|^\ell |u_1|^{j_1} \dots |u_d|^{j_d} |K(u)| du$. Moreover, as $m = 0$,

$$\begin{aligned} |b_T(z)| &\leq \int_{\mathbb{R}^d} K(v) |f(z - \delta_T v) - f(z)| dv \\ &\leq \ell \int_{\mathbb{R}^d} K(v) |\delta_T v|^\ell dv \\ &\leq c_{(r,0)} \delta_T^\ell. \end{aligned}$$

Hence, for $m \geq 0$,

$$T \delta_T^d b_T^2(z) \leq T \delta_T^{2r+d} c_{(r,m)}^2 = \eta_T^{2r+d} c_{(r,m)}^2.$$

Noticing $T \delta_T^d = T^{\frac{2r}{2r+d}} \eta_T^d$, we have $T^{\frac{2r}{2r+d}} b_T^2(z) \leq \eta_T^{2r} c_{(r,m)}^2$.

The variance term $V f_T$ is dominated from above as follows:

$$\begin{aligned} V f_T(z) &= E \left| \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} K_{\delta_T}(z - \zeta) (p - q)(dt, d\zeta) \right|^2 \\ &= \frac{1}{T^2 \delta_T^d} \int_0^T \int_{\mathbb{R}^d} \frac{\lambda}{\delta_T^d} K^2 \left(\frac{z - \zeta}{\delta_T} \right) F(\zeta) d\zeta dt \\ &\leq \frac{1}{T \delta_T^d} \sup_{z \in \mathbb{R}^d} |f(z)| \int_{\mathbb{R}^d} K^2(u) du. \end{aligned}$$

This completes the proof. \square

5.7.2 Proof of Theorem 5.2

The proof is analogous to that of Theorem 5.1. We notice the following decomposition:

$$\begin{aligned} E|f_n(z) - f(z)|^2 &= (E f_n(z) - f(z))^2 + E f_n^2(z) - (E f_n(z))^2 \\ &=: b_n^2 + V f_n, \end{aligned}$$

where $b_n(z) = E f_n(z) - f(z)$ and $V f_n(z) = E f_n^2(z) - (E f_n(z))^2$. We define the stopping time τ_i^n as

$$\tau_i^n := \inf \{t \in [t_{i-1}^n, t_i^n]; |\Delta z_t| > 0\}, \quad (5.30)$$

and let $b_n = b_n^{(1)} + b_n^{(2)} + b_n^{(3)}$, where

$$\begin{aligned} b_n^{(1)}(z) &= E f_n(z) - \frac{1}{T_n} \sum_{i=1}^n E [K_{\delta_n}(z - \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}}], \\ b_n^{(2)}(z) &= \frac{1}{T_n} \sum_{i=1}^n E [K_{\delta_n}(z - \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n=1\}}] - \frac{1}{T_n} \sum_{i=1}^n \lambda^{-1} f(z) P\{J_i^n = 1\}, \end{aligned}$$

$$b_n^{(3)}(z) = \frac{1}{T_n} \sum_{i=1}^n \lambda^{-1} f(z) P\{J_i^n = 1\} - f(z).$$

On $b_n^{(1)}(z)$, applying Lemma 5.2 with $p = 1$, we have

$$\begin{aligned} \sup_{z \in \mathbb{R}^d} |b_n^{(1)}(z)| &\leq \frac{1}{T_n \delta_n^d} \sum_{i=1}^n \sup_{z \in \mathbb{R}^d} E \left[\left| k_n(z; \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > L h_n^\rho\}} - k_n(z; \Delta z_{\tau_i^n}) \mathbf{1}_{\{J_i^n = 1\}} \right| \right] \\ &\leq \frac{1}{T_n \delta_n^d} \sum_{j=1}^n \sum_{r=1}^3 \sup_{z \in \mathbb{R}^d} E \left| K_{i,n}^{(r)}(z) \right| \\ &= \sqrt{h_n} \delta_n^{-(d+1)} \{O(1) + O(\delta_n h_n^{\rho-1/2})\} = O(\sqrt{h_n} \delta_n^{-(d+1)}), \end{aligned}$$

where k_n and $K_{i,n}^{(r)}$ ($r = 1, 2, 3$) are given in Section 5.6. Therefore

$$T_n \delta_n^d \sup_{z \in \mathbb{R}^d} \{b_n^{(1)}\}^2 = O(\eta_n^{-(2r+d)} (T_n h_n^{1/2} \delta_n^{r-1})^2) \rightarrow 0.$$

On $b_n^{(2)}(z)$,

$$\begin{aligned} b_n^{(2)}(z) &= \frac{1}{T_n} \sum_{i=1}^n \int_{\mathbb{R}^d} K_{\delta_n}(z - u) \lambda^{-1} f(u) du \times P\{J_i^n = 1\} \\ &\quad - \frac{1}{T_n} \sum_{i=1}^n \lambda^{-1} f(z) P\{J_i^n = 1\} \\ &= \frac{1}{T_n} \sum_{i=1}^n h_n e^{-\lambda h_n} \left\{ \int_{\mathbb{R}^d} K_{\delta_n}(z - u) f(u) du - f(z) \right\} \\ &= e^{-\lambda h_n} \int_{\mathbb{R}^d} K(v) \{f(z - \delta_n v) - f(z)\} dv. \end{aligned}$$

As $m \neq 0$, applying Taylor's formula to f ,

$$b_n^{(2)}(z) = e^{-\lambda h_n} \delta_n^m \int_{\mathbb{R}^d} K(v) \sum_{j_1 + \dots + j_d = m} \frac{v_1^{j_1} \dots v_d^{j_d}}{j_1! \dots j_d!} \frac{\partial f^{(m)}}{\partial z_1^{j_1} \dots \partial z_d^{j_d}}(z - \theta \delta_n v) dv,$$

where $0 < \theta < 1$. Hence, by the same argument as in the proof of Theorem 4.1 in Bosq [10],

$$T_n \delta_n^d \sup_{z \in \mathbb{R}^d} \{b_n^{(2)}(z)\}^2 \leq \eta_n^{2r+d} \eta_{(r,m)}^2. \quad (5.31)$$

where $\eta_{(r,m)} = \sum_{j_1 + \dots + j_d = m} \frac{\ell}{j_1! \dots j_d!} \int_{\mathbb{R}^d} |u|^l |u_1|^{j_1} \dots |u_d|^{j_d} |K(u)| du$. Moreover, as $m = 0$, we can obtain (5.31) with $m = 0$ by the same argument as in the proof of Theorem 5.1.

On $b_n^{(3)}(z)$,

$$\begin{aligned} \sup_{z \in \mathbb{R}^d} |b_n^{(3)}(z)| &= \sup_{z \in \mathbb{R}^d} \left| \frac{1}{T_n} \sum_{i=1}^n h_n e^{-\lambda h_n} f(z) - f(z) \right| \\ &= \sup_{z \in \mathbb{R}^d} |f(z)| |e^{-\lambda h_n} - 1| \\ &= O(h_n). \end{aligned}$$

This implies that $T_n \delta_n^d \sup_{z \in \mathbb{R}^d} \{b_n^{(3)}(z)\}^2 = O((T_n h_n^\nu) h_n^{2-\nu} \delta_n^d) \rightarrow 0$. Clearly, the cross terms from $b_n^{(1)}$ to $b_n^{(3)}$ converge zero at rate $T_n \delta_n^d$ uniformly in z . Hence,

$$\limsup_{n \rightarrow \infty} T_n^{\frac{2r}{2r+d}} \sup_{z \in \mathbb{R}^d} b_n^2(z) \leq \eta^{2r} \cdot \eta_{(r)}.$$

Next, let us consider the term Vf_n . We can decompose Vf_n as $Vf_n := V_n + C_n$, where

$$\begin{aligned} V_n(z) &= \frac{1}{T_n^2} \sum_{i=1}^n \text{Var} (K_{\delta_n}(z - \Delta_i z^n)) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}, \\ C_n(z) &= \frac{1}{T_n^2} \sum_{1 \leq i \neq j \leq n} \text{Cov} \left(K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}, K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right). \end{aligned}$$

On $V_n(z)$, applying lemma 5.3, we have

$$\begin{aligned} V_n(z) &= \frac{1}{T_n^2} \sum_{i=1}^n \left\{ E \left[K_{\delta_n}^2(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} \right] \right. \\ &\quad \left. - \left(E \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} \right] \right)^2 \right\} \\ &= \frac{1}{T_n \delta_n^d} (g_{n,2}(z) + Ch_n^{1/2} \delta_n^{-(d+1)}). \end{aligned}$$

where $g_{n,p}$ ($p > 0$) is given in Section 5.6.

On $C_n(z)$, noticing that

$$\begin{aligned} &\text{Cov} \left(K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}}, K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right) \\ &= E \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right] \\ &\quad - E \left[K_{\delta_n}(z - \Delta_i z^n) \mathbf{1}_{\{|\Delta X_i^n| > Lh_n^\rho\}} \right] E \left[K_{\delta_n}(z - \Delta_j z^n) \mathbf{1}_{\{|\Delta X_j^n| > Lh_n^\rho\}} \right], \end{aligned}$$

and applying lemma 5.5 to the first term and lemma 5.3 to the second term in the right-hand side, we easily obtain $\sup_{z \in \mathbb{R}^d} |C_n(z)| = O(h_n^{1/2-\mu} \delta_n^{-(d+1+\mu d)})$ for arbitrary $\mu \in (0, 1)$. Hence,

$$T_n \delta_n^d \sup_{z \in \mathbb{R}^d} Vf_n(z)$$

$$\begin{aligned}
&= \sup_{z \in \mathbb{R}^d} g_{n,2}(z) + O(h_n^{1/2} \delta_n^{-(d+1)}) + O(T_n h_n^{1/2-\mu} \delta_n^{-1-\mu d}) \\
&= \sup_{z \in \mathbb{R}^d} g_{n,2}(z) + O(\eta_n^{-(2r+d)} \delta_n^{2r-1} T_n h_n^{1/2}) + O(T_n h_n^{1/2-\mu} T_n^{(1+\mu d)/(2r+d)}) \\
&\leq \sup_{z \in \mathbb{R}^d} |f(z)| \int K^2(u) du + O(\delta_n^{2r-1} T_n h_n^{1/2}) + O\left(T_n^{\frac{2r+d+1+\mu d}{2r+d}} h_n^{\frac{1}{2}-\mu}\right).
\end{aligned}$$

The last two terms in the last right-hand side tends to zero if we take $\mu \in (0, 1)$ arbitrary small. Actually, $\delta_n^{2r-1} T_n h_n^{1/2} = (T_n h_n^{1/2} \delta_n^{r-1}) \delta_n^r \rightarrow 0$ and

$$T_n^{\frac{2r+d+1+\mu d}{2r+d}} h_n^{\frac{1}{2}-\mu} = (T_n h_n^\nu)^{\frac{2r+d+1+\mu d}{2r+d}} h_n^{\frac{1}{2}-\frac{2r+d+1+\mu d}{2r+d}\nu-\mu}$$

whose last index $\frac{1}{2} - \frac{2r+d+1+\mu d}{2r+d}\nu - \mu$ can be positive if we take μ sufficiently small since $0 < \nu < \frac{2r+d}{2(2r+d+1)}$. As a result, we have

$$T_n \delta_n^d \sup_{z \in \mathbb{R}^d} V f_n(z) \leq \sup_{z \in \mathbb{R}^d} |f(z)| \int K^2(u) du + o(T_n h_n^\nu). \quad (5.32)$$

This completes the proof. \square

Chapter 6

Practical inference from finite samples

As we have seen until the previous chapter, when we consider the inference for jump-type processes, it is useful to take the information of jumps and the information of continuous transition separately. For that purpose, we made use of the filter such as $\{|\Delta_i X^n| \leq Lh_n^\rho\}$ for the inference for jump-diffusions from sampled data, and we showed that this filter gave a *good* judgement if jump had occurred or not asymptotically. However, as we already saw in Section 5.4.2, this filter did not work well without selecting the constants L and ρ suitably under some practical situations where the sample size n was fixed. Therefore we proposed some intuitive methods to improve the performance of the estimation. In this chapter, we discuss how to construct the filter depending on the fixed sample size from a more theoretical point of view.

6.1 Asymptotic filter and its problem

Our essential idea in Chapter 3 - 5 is to use the size of $|\Delta_i X^n|$ in order to judge the existence of a jump in the interval $[t_{i-1}^n, t_i^n]$. For that purpose, we used the jump-judgment filter of the form

$$\mathcal{H}_i^n = \{\omega \in \Omega; |\Delta_i X^n| > Lh_n^\rho\}, \quad \rho \in (0, 1/2). \quad (6.1)$$

The key result to show the validity to use such a filter was Lemma 3.2, 4.1 or 5.1.

However, as pointed out in Section 5.4.2, the accuracy of the judgment of jumps for given L and ρ depends on the sample size n and the each model, especially the diffusion coefficient and the distribution of jumps. That is, when we fix the constant

L and ρ independent of the sample size n then the filter does not work well in some models.

Although we already show some examples where the filter does not work well in Section 5.4, we shall show again an another numerical experiment. Consider an one-dimensional SDE as follows.

$$dX_t = -\mu X_t dt + \sqrt{\sigma} dw_t + z_t^{(\lambda, \theta)},$$

where $z^{(\lambda, \theta)}$ is a compound Poisson process with the Lévy density

$$f(z) = \lambda(2\pi\theta_2)^{-1/2} e^{-(z-\theta_1)^2/(2\theta_2)};$$

the normal type density with parameters $\lambda, \theta = (\theta_1, \theta_2)$. Set the true value of parameters except for σ as follows.

$$(\mu, \theta_1, \theta_2, \lambda) = (0.3, 0.5, 0.1, 3.0),$$

and consider the two models as $\sigma = 0.1$ and $\sigma = 0.3$. We estimate the parameter $(\mu, \sigma, \theta_1, \theta_2, \lambda)$ jointly from discrete observations $\mathbf{X}^n = (X_{t_0^n}, X_{t_1^n}, \dots, X_{t_n^n})$, where $t_i^n = ih_n$ for $i = 0, 1, \dots, n$, via the method of Chapter 3 with a slight extension to the case where the Lévy density is bounded around the origin. In the following simulation, we set $h_n = n^{-0.8}$ and $\rho = 0.49$.

In the case where $\sigma = 0.1$, we first set $L = 1.0$ and obtained Tab. 6.1. From this result, one might think that the filter works well. This is because the diffusion parameter σ is relatively small compared with jump sizes, and there is only a few misjudgments. However, as $\sigma = 0.3$, we obtained a result as in Tab. 6.2. The estimator $\hat{\theta}_1$ behaves strange, and $\hat{\lambda}$ is overestimated with the standard deviation increases. This is because the filter misjudges the jumps, especially it overestimates the jump's number. This simulation indicates that the asymptotic theory does not work yet, and the sample size n needs being larger. In order to improve the performance of the estimation, we change the value $L = 1.0$ to a suitable one. Here we chose $L = 1.8$, and the result is shown in Tab. 6.3. Then the result were dramatically improved.

These results indicate that we need to select the constant L (and of course ρ , too) according to the model and the sample size n .

In Chapter 5, we left a constant L in the threshold to describe some intuitive procedures for selecting the filter. However there is no theoretical reason to separate L and h_n^ρ generally as pointed out in Remark 5.3. Therefore, in this chapter, we rewrite the filter simply as

$$\mathcal{H}_i^n(r_n) = \{\omega \in \Omega; |\Delta_i X^n| > r_n\}, \quad (6.2)$$

	n	50	500	3000	TRUE
$\sigma = 0.1$	$\hat{\mu}$	0.2702	0.3009	0.3013	0.3
$L = 1.0$	s.d.	0.1474	0.0477	0.0277	
	$\hat{\sigma}$	0.1169	0.1003	0.0998	0.1
	s.d.	0.0295	0.0066	0.0025	
	$\hat{\theta}_1$	0.6450	0.5408	0.5135	0.5
	s.d.	0.1294	0.0745	0.0558	
	$\hat{\theta}_2$	0.0919	0.0924	0.0956	0.1
	s.d.	0.0886	0.0495	0.0269	
	$\hat{\lambda}$	2.0659	2.6893	2.9067	3.0
	s.d.	0.7266	0.6514	0.5076	

Tab. 6.1: The mean and the standard deviation (s.d.) of estimators over 500 times iterations.

	n	50	500	3000	TRUE
$\sigma = 0.3$	$\hat{\mu}$	0.2358	0.2528	0.2536	0.3
$L = 1.0$	s.d.	0.2366	0.0809	0.0428	
	$\hat{\sigma}$	0.2310	0.2381	0.2471	0.3
	s.d.	0.0433	0.0133	0.0054	
	$\hat{\theta}_1$	0.4505	0.2230	0.1108	0.5
	s.d.	0.0151	0.0522	0.0162	
	$\hat{\theta}_2$	0.2131	0.1445	0.0839	0.1
	s.d.	0.1178	0.0331	0.0162	
	$\hat{\lambda}$	2.7419	5.7280	11.453	3.0
	s.d.	0.8759	0.9882	1.0682	

Tab. 6.2: The mean and the standard deviation (s.d.) of estimators over 500 times iterations.

	n	50	500	3000	True
$\sigma = 0.3$	$\hat{\mu}$	0.2001	0.2977	0.3010	0.3
$L = 1.8$	s.d.	0.2366	0.0866	0.0428	
	$\hat{\sigma}$	0.3902	0.3044	0.2978	0.3
	s.d.	0.0901	0.0191	0.0074	
	$\hat{\theta}_1$	0.7517	0.5750	0.5147	0.5
	s.d.	0.1879	0.0800	0.0552	
	$\hat{\theta}_2$	0.2014	0.0965	0.1002	0.1
	s.d.	0.2054	0.0531	0.0279	
	$\hat{\lambda}$	1.5156	2.4717	2.8938	3.0
	s.d.	0.6450	0.5881	0.4682	

Tab. 6.3: The mean and the standard deviation (s.d.) of estimators over 500 times iterations.

and consider the selection problem of not L but r_n itself, which is the most important parameter in applications.

In the next section, we discuss what kind of r_n is suitable to improve the performance for fixed n .

6.2 A criterion for selecting the filter

Throughout this chapter, we consider the following d -dimensional SDE with jumps on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, P)$:

$$dX_t = a(X_t) dt + b(X_t) dw_t + dz_t, \quad X_0 = x, \quad (6.3)$$

where x is an \mathbb{R}^d -valued random variable, a and b are respectively \mathbb{R}^d and $\mathbb{R}^d \otimes \mathbb{R}^m$ -valued measurable functions defined on \mathbb{R}^d , w is an m -dimensional Wiener process, z is an \mathbb{R}^d -valued compound Poisson process of the form $z_t = \sum_{i=1}^{N_t} \varepsilon_i$ and N is a Poisson process with intensity parameter $\lambda (\geq 0)$, ε_i 's ($i \in \mathbb{N}$) are \mathbb{R}^d -valued random variables satisfying $P\{\varepsilon_i \in A\} = \int_A F(z) dz$ for any $A \subset \mathbb{R}^d$, and we put $f(z) = \lambda F(z)$. If z has a jump ε_i at the time τ_i then $|\Delta X_{\tau_i}| > 0$ a.s., where $\Delta X_t := X_t - X_{t-}$ for any $t \geq 0$. We assume that w , N and ε_i 's are independent each other, and that (6.3) has a solution.

Such a model is often used in financial literatures to model the dynamics of asset returns, index returns, exchange rates, interest rates, insurance risks, and so on. Therefore to determine the *good* threshold of the filter is important problem in practice.

Notice that, in a time interval $(s_1, s_2]$ where any jump does not occur, X is a solution to the following stochastic integral equation:

$$\alpha_t^{(s_1)} = X_{s_1} + \int_{s_1}^t a(\alpha_u^{(s_1)}) du + \int_{s_1}^t b(\alpha_u^{(s_1)}) dw_u. \quad (6.4)$$

The statistics

$$\hat{\lambda}_n(r_n) = \frac{1}{nh_n} \sum_{i=1}^n \mathbf{1}_{\mathcal{H}_i^n(r_n)} \quad (6.5)$$

is an estimator of the Poisson intensity of jumps over $[0, 1]$ time interval. It is shown in Shimizu [94] that this estimator was consist to λ unless X is ergodic, and it is easily found that (6.5) is asymptotically unbiased, that is, $E[\hat{\lambda}_n(r_n)] \rightarrow \lambda$ as n tends to infinity. Therefore it is desired that, at least, the bias of the estimator $\hat{\lambda}_n(r_n)$ for largely fixed n is as small as possible.

Let us estimate the following expectation analytically:

$$E[\hat{\lambda}_n(r_n)] = \frac{1}{nh_n} \sum_{i=1}^n P\{\mathcal{H}_i^n(r_n)\} = \frac{1}{nh_n} \sum_{i=1}^n P\{|\Delta_i X^n| > r_n\}.$$

In the sequel, we use the following notations:

$$\varepsilon(r_n) := \frac{1}{n} \sum_{i=1}^n P\{|\Delta_i \alpha^n(t_{i-1}^n)| > r_n\}, \quad (6.6)$$

$$\delta(r_n, A) := \frac{1}{n} \sum_{i=1}^n P\{|S_i^n(\tau)| > r_n | \{J_i^n = 1\} \cap A\} \quad (6.7)$$

for $A \in \mathcal{F}$, where $\Delta_i \alpha^n(t_{i-1}^n) = \alpha_{t_i^n}^{(t_{i-1}^n)} - \alpha_{t_{i-1}^n}^{(t_{i-1}^n)}$ and $S_i^n(\tau) = (\alpha_{\tau_i^n}^{(\tau_i^n)} - \alpha_{\tau_i^n}^{(\tau_i^n)}) + (\alpha_{\tau_{i-1}^n}^{(t_{i-1}^n)} - \alpha_{t_{i-1}^n}^{(t_{i-1}^n)})$, $\tau_i^n = \inf\{t \in (t_{i-1}^n, t_i^n]; |\Delta X_t| > 0\}$ and $J_i^n := \#\{t \in (t_{i-1}^n, t_i^n]; |\Delta X_t| > 0\}$.

Lemma 6.1 *For any integer n , it follows that*

$$\begin{aligned} E[\hat{\lambda}_n(r_n)] &= h_n^{-1} \varepsilon(r_n) e^{-\lambda h_n} + (1 - \delta(r_n, |\Delta z_{\tau_i^n}| > 2r_n)) e^{-\lambda h_n} \int_{|z| > 2r_n} f(z) dz \\ &\quad + T_n + U_n + e_n, \end{aligned}$$

where e_n , T_n and U_n are sequences satisfying the following inequalities, respectively:

$$\begin{aligned} 0 &\leq e_n \leq \lambda^2 h_n, \\ 0 &\leq e^{\lambda h_n} T_n \leq \int_{|z| > 2r_n} f(z) dz \cdot \delta(r_n, |\Delta z_{\tau_i^n}| > 2r_n), \\ 0 &\leq e^{\lambda h_n} U_n \leq \int_{\frac{r_n}{2} < |z| \leq 2r_n} f(z) dz + \delta\left(\frac{r_n}{2}, |\Delta z_{\tau_i^n}| \leq \frac{r_n}{2}\right) \int_{|z| \leq \frac{r_n}{2}} f(z) dz. \end{aligned}$$

Proof . Notice the following decomposition:

$$\begin{aligned} E \left[\hat{\lambda}_n(r_n) \right] &= \frac{1}{nh_n} \sum_{i=1}^n \left[P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n = 0\} \} + P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n = 1\} \} \right. \\ &\quad \left. + P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n \geq 2\} \} \right]. \end{aligned} \quad (6.8)$$

Let the last term be e_n then it follows from the Poisson property that

$$\begin{aligned} e_n &:= \frac{1}{nh_n} \sum_{i=1}^n P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n \geq 2\} \} \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n P \{ J_i^n \geq 2 \} \leq \lambda^2 h_n. \end{aligned} \quad (6.9)$$

Since $dX_t = d\alpha_t^{(t_{i-1}^n)}$ in $(t_{i-1}^n, t_i^n]$ on the set $\{J_i^n = 0\}$, we have on the first term of (6.8) that

$$P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n = 0\} \} = e^{-\lambda h_n} P \{ |\Delta_i \alpha^n(t_{i-1}^n)| > r_n \}.$$

It remains to estimate the second term of (6.8). Noticing that

$$S_i^n(\tau) = (X_{t_i^n} - X_{\tau_i^n}) + (X_{\tau_i^n} - X_{t_{i-1}^n})$$

on $\{J_i^n = 1\}$, we obtain the following decomposition.

$$\begin{aligned} &P \{ \mathcal{H}_i^n(r_n) \cap \{J_i^n = 1\} \} \\ &= P \{ |S_i^n(\tau) + \Delta z_{\tau_i^n}| > r_n, |\Delta z_{\tau_i^n}| > 2r_n, J_i^n = 1 \} \\ &\quad + P \{ |S_i^n(\tau) + \Delta z_{\tau_i^n}| > r_n, |\Delta z_{\tau_i^n}| \leq 2r_n, J_i^n = 1 \} \\ &\leq P \{ |S_i^n(\tau)| \leq r_n, |\Delta z_{\tau_i^n}| > 2r_n, J_i^n = 1 \} \end{aligned} \quad (6.10)$$

$$+ P \{ |S_i^n(\tau) + \Delta z_{\tau_i^n}| > r_n, |S_i^n(\tau)| > r_n, |\Delta z_{\tau_i^n}| > 2r_n, J_i^n = 1 \} \quad (6.11)$$

$$+ P \left\{ |S_i^n(\tau) + \Delta z_{\tau_i^n}| > r_n, \frac{r_n}{2} < |\Delta z_{\tau_i^n}| \leq 2r_n, J_i^n = 1 \right\} \quad (6.12)$$

$$+ P \left\{ |S_i^n(\tau) + \Delta z_{\tau_i^n}| > r_n, |\Delta z_{\tau_i^n}| \leq \frac{r_n}{2}, J_i^n = 1 \right\}. \quad (6.13)$$

Putting (6.10) ~ (6.13) as A_i^n, B_i^n, C_i^n and D_i^n respectively, we have

$$A_i^n = \lambda h_n e^{-\lambda h_n} \int_{|z| > 2r_n} F(z) dz \cdot P \{ |S_i^n(\tau)| \leq r_n, |\Delta z_{\tau_i^n}| > 2r_n, J_i^n = 1 \},$$

$$B_i^n \leq \lambda h_n e^{-\lambda h_n} \int_{|z| > 2r_n} F(z) dz \cdot P \{ |S_i^n(\tau)| > r_n, |\Delta z_{\tau_i^n}| > 2r_n, J_i^n = 1 \},$$

$$C_i^n \leq \lambda h_n e^{-\lambda h_n} \int_{\frac{r_n}{2} < |z| \leq 2r_n} F(z) dz,$$

$$D_i^n \leq \lambda h_n e^{-\lambda h_n} \int_{|z| \leq \frac{r_n}{2}} F(z) dz \cdot P \left\{ |S_i^n(\tau)| > \frac{r_n}{2} \mid |\Delta z_{\tau_i^n}| \leq \frac{r_n}{2}, J_i^n = 1 \right\}.$$

Then we obtain the consequence by putting as follows:

$$\begin{aligned} T_n &= \frac{1}{nh_n} \sum_{i=1}^n B_i^n, \\ U_n &= \frac{1}{nh_n} \sum_{i=1}^n (C_i^n + D_i^n). \end{aligned}$$

This completes the proof. \square

According to this lemma, we can give an intuitive explanation to the curve of $\hat{\lambda}_n(r_n)$ in Fig. 5.4 in Chapter 5. First, let us observe the behavior of $E \left[\hat{\lambda}_n(r_n) \right]$ as $r_n \uparrow \infty$ under n is fixed; which corresponds to that $L \rightarrow \infty$ in (6.1). Since

$$\begin{aligned} \varepsilon(r_n) \downarrow 0, \quad \delta(r_n, |\Delta z_{\tau_i^n}| > 2r_n) + \delta \left(\frac{r_n}{2}, |\Delta z_{\tau_i^n}| \leq \frac{r_n}{2} \right) \downarrow 0, \\ \int_{\frac{r_n}{2} \leq |z| \leq 2r_n} f(z) dz + \int_{|z| > 2r_n} f(z) dz \downarrow 0 \end{aligned}$$

as $r_n \uparrow \infty$, we find that $E \left[\hat{\lambda}_n(r_n) \right] \rightarrow 0$. Intuitively speaking, this phenomenon is natural since the filter can hardly catch jumps if r_n is too large.

Next, let us consider the case where $r_n \downarrow 0$ under n is fixed. In this case, it follows from $\varepsilon(r_n), \delta(r_n) \uparrow 1$ and $U(r_n) \downarrow 0$ that

$$E \left[\hat{\lambda}_n(r_n) \right] \rightarrow h_n^{-1} e^{-\lambda h_n} + \lim_{r_n \rightarrow 0} T(r_n) \approx h_n^{-1} + \lambda.$$

Therefore $E \left[\hat{\lambda}_n(r_n) \right]$ has the large bias by the influence of the term $h_n^{-1} \varepsilon(r_n) e^{-\lambda h_n}$ as r_n is too small. Particularly the term h_n^{-1} becomes enormously large if n is large.

6.3 A bias correction

Let $b(r_n)$ be the exact bias of $\hat{\lambda}_n(r_n)$, that is,

$$b(r_n) := E \left[\hat{\lambda}_n(r_n) \right] - \lambda. \quad (6.14)$$

Our goal is to select an r_n which minimizes the absolute bias $|b(r_n)|$. However it would be difficult to estimate the exact bias directly, and what we show here is the upper and the lower bound of the bias. We easily obtain the following theorem from the previous lemma.

Theorem 6.1 Define $\tilde{\ell}_n$ and ℓ_n as follows.

$$\begin{aligned}\tilde{\ell}_n &:= h_n^{-1} \varepsilon(r_n) - \int_{|z| < 2r_n} f(z) dz, \\ \ell_n &:= \tilde{\ell}_n + \int_{\frac{r_n}{2} < |z| \leq 2r_n} f(z) dz.\end{aligned}$$

Then it follows that

$$\tilde{\ell}_n + O(\tilde{\delta}_n \vee h_n) \leq e^{\lambda h_n} b(r_n) \leq \ell_n + O(\tilde{\delta}_n \vee h_n) \quad (6.15)$$

as $h_n \rightarrow 0$, where $\tilde{\delta}_n = \delta(r_n, |\Delta z_{\tau_i^n}| > 2r_n) \vee \delta(r_n/2, |\Delta z_{\tau_i^n}| \leq r_n/2)$.

Proof . This is the direct result from Lemma 6.1 and the fact $\lambda = \int_{\mathbb{R}^d} f(z) dz$. \square

Roughly speaking, we can regard ℓ_n and $\tilde{\ell}_n$ as the terms of the first order of the upper and the lower bound, respectively since $h_n \vee \tilde{\delta}_n \rightarrow 0$ as $n \rightarrow \infty$ under $r_n \sim h_n^\rho$ for $\rho < 1/2$, which follows from (3.11). For these bounds, it is desired that the amplitude of the bias $|\tilde{\ell}_n - \ell_n|$, which corresponds to a kind of *variance*, is as small as possible from the aspect of stability of the estimated bias. Therefore the threshold r_n should be selected so that $|\tilde{\ell}_n - \ell_n| \approx 0$. On the other hand, if we concentrate only on minimizing the distance $|\tilde{\ell}_n - \ell_n|$ then $\tilde{\ell}_n$ might be strictly positive, or ℓ_n might be strictly negative, which induce strictly biased estimators. Therefore the center of the interval $[\tilde{\ell}_n, \ell_n]$ should be nearly to zero in order to make the maximum of the absolute bias small as possible; the aspect of unbiasedness. Therefore r_n should also be selected so that $(\tilde{\ell}_n + \ell_n) \approx 0$. From these points of view, it would be natural to select r_n which minimizes the following quantity:

$$(1-u)|\tilde{\ell}_n + \ell_n| + u|\tilde{\ell}_n - \ell_n| \quad (0 \leq u < 1),$$

The weight u should not be 1 since $r_n = 0$ or ∞ is clearly selected in this case and each of them does not play a role as the filter. Therefore it is convenient to rewrite it as follows.

$$\begin{aligned}L_{n,w}(r_n) &:= |\tilde{\ell}_n + \ell_n| + w|\tilde{\ell}_n - \ell_n| \\ &= |L(r_n)| + w \int_{\frac{r_n}{2} < |z| \leq 2r_n} f(z) dz,\end{aligned}$$

for $w \geq 0$, where $L(r_n) = 2h_n^{-1}\varepsilon(r_n) - \mathcal{J}(r_n)$, $\mathcal{J}(r_n) = \int_{|z| \leq \frac{r_n}{2}} f(z) dz + \int_{|z| < 2r_n} f(z) dz$. We would like to select r_n which minimize the function $L_{n,w}(r_n)$, though it is still the unknown function.

Definition 6.1 We denote by $r_{opt}^{(n,w)}$ a minimizer of the function $L_{n,w}(r)$:

$$r_{opt}^{(n,w)} := \arg \min_{r \geq 0} L_{n,w}(r).$$

It is easy to see that $r_{opt}^{(n,w)}$ is well defined for any n and w from the form of $L_{n,w}(r)$.

A constant w is the weight on the amplitude of the bias; $|\tilde{\ell}_n - \ell_n|$. Setting w as being large implies that one puts weight on not the unbiased estimation but the stable estimation, and too large w can induce the definitely positive or negative bias. Since we do not obtain the exact bias but only less strict bounds from both sides, the aspect of unbiasedness should be weighted on rather than compulsory minimization of the bias range. Moreover, from the technical point of view, it is not necessarily that $r_{opt}^{(n,w)}$ is determined uniquely if $w > 1$. Actually, if $w \rightarrow \infty$ then the selected threshold would tend to zero or infinity as already pointed out. Therefore it would not be suitable to choose large w , and be suitable that $w = 1$ if we have no prior information about the true bias.

In application, if we restrict $w \in [0, 1]$ then $r_{opt}^{(n,w)}$ is unique and independent of w as the next lemma shows. The lemma implies that, if $0 \leq w \leq 1$, minimizing $L_{n,w}$ is equivalent to finding the unique root of $L(r) = 0$.

Lemma 6.2 If $0 \leq w \leq 1$ then $r_{opt}^{(n,w)}$ is the unique solution to the equation $L(r) = 0$, that is, $r_{opt}^{(n,w)}$ is independent of $w \in [0, 1]$.

Proof . Note that

$$L_{n,w}(r) = \begin{cases} 2h_n^{-1}\varepsilon(r) - \mathcal{J}_w(r) & \text{if } L(r) \geq 0 \\ \mathcal{J}_{-w}(r) - 2h_n^{-1}\varepsilon(r) & \text{if } L(r) \leq 0 \end{cases},$$

where $\mathcal{J}_w(r) = \mathcal{J}(r) - w \int_{\frac{r}{2} < |z| \leq 2r} f(z) dz$. We see that the function $\mathcal{J}_{\tilde{w}}(r)$ is increasing in r for each $\tilde{w} \in [-1, 1]$ since we can rewrite $\mathcal{J}_{\tilde{w}}$ as follows.

$$\mathcal{J}_{\tilde{w}}(r) = (1 + \tilde{w}) \int_{|z| \leq \frac{r}{2}} f(z) dz + (1 - \tilde{w}) \int_{|z| < 2r} f(z) dz.$$

Hence, both the function $l_1(r) := 2h_n^{-1}\varepsilon(r) - \mathcal{J}_w(r)$ and the function $l_2(r) := 2h_n^{-1}\varepsilon(r) - \mathcal{J}_{-w}(r)$ are decreasing in $r \geq 0$ for each $w \in [0, 1]$. Similarly, the function $L(r)$ is also decreasing. Moreover the equation $L(r) = 0$ has the unique root since $\varepsilon(r)$ is decreasing in $r \geq 0$ and $\varepsilon(0) = 1$, $\lim_{r \rightarrow \infty} \varepsilon(r) = 0$, and $\mathcal{J}(r)$ is increasing in $r \geq 0$ with $\mathcal{J}(0) = 0$.

From these facts, it follows that l_1 is minimized at $r_0 := \max\{r \geq 0; L(r) \geq 0\}$ on the closed set $\{r \geq 0; L(r) \geq 0\} \subset [0, \infty)$, and also

$$r_0 = \arg \min L(r) \quad \text{on} \quad \{r \geq 0; L(r) \geq 0\},$$

that is, $L(r_0) = 0$. Similarly, we can show that l_2 is maximized at r'_0 satisfying $L(r'_0) = 0$. Consequently we obtain that $r_0 = r'_0 = r_{opt}^{(n,w)}$. \square

In the sequel, we consider the case where $w \in [0, 1]$. Under this assumption we put

$$r_{opt}^{(n)} := r_{opt}^{(n,w)} \quad (6.16)$$

for simplicity. Then our interest is to find the solution $r_{opt}^{(n)}$ to the equation

$$L(r) = 2h_n^{-1}\varepsilon(r) - \mathcal{J}(r) = 0. \quad (6.17)$$

The equation (6.17) includes the unknown quantities ε and f . Therefore we have to substitute ε and f by some suitable estimators in order to make an estimator of $r_{opt}^{(n)}$. However they must be constructed by the filter which should be selected in our goal, so it goes back and forth!

In the next section, we propose a plug-in method in order to avoid this dilemma, and we show the performance of the method in some simulations.

6.4 Direct plug-in method

6.4.1 Plug-in rule

The goal of this section is to estimate the threshold $r_{opt}^{(n)}$ from finitely fixed n . For that purpose, we have to construct an estimator of $L(r)$.

As a general notation, we denote by $\hat{G}_n(x; r_n)$ an estimator of a function $G(x)$ constructed by the data $\{X_{t_i^n}\}_{i=0}^n$ and the filter $\mathcal{H}_i^n(r_n)$. Using this notation, the natural estimator of \mathcal{J} is written as

$$\hat{\mathcal{J}}_n(r; r_n) = \int_{|z| \leq \frac{r}{2}} \hat{f}_n(z; r_n) dz + \int_{|z| \leq 2r} \hat{f}_n(z; r_n) dz. \quad (6.18)$$

Let us consider an 1-dimensional case of X for simplicity. In order to calculate the above integrals easily, it is convenient to use

$$\hat{f}_n(z; r_n) := \frac{1}{nh_n^{1+\delta}} \sum_{i=1}^n \phi\left(\frac{z - \Delta_i X^n}{h_n^\delta}\right) I(\mathcal{H}_i^n(r_n)) \quad (6.19)$$

proposed in Shimizu [92], where $\delta \in (0, 1/2)$ is a constant, ϕ is the standard normal kernel, i.e. $\phi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. On the other hand, the natural approximation of the $\varepsilon_n(r_n) = \frac{1}{n} \sum_{i=1}^n P\{|\Delta_i \alpha^n(t_{i-1}^n)| > r_n\}$ for diffusion (6.4) is the local-Gaussian

approximation of the transition probability of $\alpha^{(t_{i-1}^n)}$, that is, for $a_{i-1} = a(X_{t_{i-1}^n})$ and $\beta_{i-1} = b^2(X_{t_{i-1}^n})$,

$$\begin{aligned}\varepsilon(r_n) &\approx \frac{1}{n} \sum_{i=1}^n \int_{|y| > r_n} \frac{1}{\sqrt{2\pi|\beta_{i-1}|h_n}} \exp\left(-\frac{1}{2h_n\beta_{i-1}}(y - h_n a_{i-1})^2\right) dy \\ &\approx \frac{1}{n} \sum_{i=1}^n \int_{|y| > r_n} \frac{1}{\sqrt{2\pi|\beta_{i-1}|h_n}} \exp\left(-\frac{y^2}{2h_n\beta_{i-1}}\right) dy.\end{aligned}\quad (6.20)$$

Substituting the β by the estimator $\hat{\beta}_n(X_{t_{i-1}^n}; r_n)$, we have

$$\hat{\varepsilon}_n(r; r_n) = \frac{1}{n} \sum_{i=1}^n \int_{|y| > r_n} \frac{1}{\sqrt{2\pi|\hat{\beta}_n(X_{t_{i-1}^n}; r_n)|h_n}} \exp\left(-\frac{y^2}{2h_n\hat{\beta}_n(X_{t_{i-1}^n}; r_n)}\right) dy. \quad (6.21)$$

We note that the above procedure is easily applied to the multidimensional case.

Now let us proceed the algorithm to find the approximator of $r_{opt}^{(n)}$. The following *Plug-in method* is executable:

Step 0. Choose the pilot threshold $r_n^{(0)} > 0$ arbitrarily.

Step k (≥ 1). Solve the equation

$$\hat{L}_n(r; r_n^{(k-1)}) := 2h_n^{-1}\hat{\varepsilon}_n(r; r_n^{(k-1)}) - \hat{\mathcal{J}}_n(r; r_n^{(k-1)}) = 0 \quad \dots (*)$$

and define the root as $r = r_n^{(k)}$.

Iterate Step k ($k = 1, 2, \dots$) until the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ converges.

We call the k th solution to the equation $(*)$ the *k -stage threshold*, and call the function $\hat{L}_n(r; r_n^{(k-1)})$ the *k -stage threshold selector*. By the same argument as in Lemma 6.2, we see that the equation $\hat{L}_n(r; r_n^{(k-1)}) = 0$ has the unique solution $r_n^{(k)}$ for any k and $n \in \mathbb{N}$.

We expect that $\lim_{k \rightarrow \infty} r_n^{(k)}$ exists and that the limit is *near* the $r_{opt}^{(n)}$ in some sense. Before the theoretical study, we try this algorithm by simulations.

6.4.2 Simulation results

Let us show the performance of our threshold selector \hat{L}_n using (6.19) and (6.21).

We consider an one-dimensional data generating process as follows.

$$dX_t = -\mu X_t dt + \sigma dw_t + dz_t,$$

where z is a compound Poisson process with the Lévy density $f(z) = \frac{\lambda}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$, and the true parameter is $(\mu, \sigma, \lambda) = (0.03, 0.3, 15.0)$.

For fixed n , the experiment is done as follows.

- (1) Choose the pilot threshold as $r_n^{(0)} = 1.0$.
- (2) Observe one path of X at the time $t_i^n = ih_n$, where $h_n = n^{-0.8}$.
- (3) Calculate 1 to 7-stage thresholds solving $\hat{L}_n(r; r_n^{(k-1)}) = 0$ ($k = 0, 1, \dots, 6$) and the estimator of (μ, σ, λ) in each stages as follows; see Definition 3.1 and Remark 3.2.

$$\begin{aligned}\hat{\mu}_n^{(k)} &= \frac{\sum_{i=1}^n \Delta_i X^n X_{t_{i-1}^n} \mathbf{1}_{\{|\Delta_i X^n| \leq r_n^{(k)}\}}}{\sum_{i=1}^n X_{t_{i-1}^n}^2 h_n^2 \mathbf{1}_{\{|\Delta_i X^n| \leq r_n^{(k)}\}}} \\ \hat{\sigma}_n^{(k)} &= \left\{ \frac{\sum_{i=1}^n (\Delta_i X^n - \hat{\mu}_n X_{t_{i-1}^n} h_n)^2 \mathbf{1}_{\{|\Delta_i X^n| \leq r_n^{(k)}\}}}{h_n \sum_{i=1}^n \mathbf{1}_{\{|\Delta_i X^n| \leq r_n^{(k)}\}}} \right\}^{1/2} \\ \hat{\lambda}_n^{(k)} &= \frac{1}{nh_n} \sum_{i=1}^n \mathbf{1}_{\mathcal{H}_i^n(r_n^{(k)})}\end{aligned}$$

The experiment (1)-(3) is iterated 500 times. Table 6.4-6.6 below are the sample mean and the sample standard deviation (s.d.) in each stages throughout 500 iterations. The values in the last line are $r_{opt}^{(n)}$ and the true values of each parameter.

These results show that our threshold selector can find the $r_{opt}^{(n)}$ approximately as a limit of the k -stage threshold, and as a result, the parameters are estimated well. Although we do not know yet if the k -stage threshold can theoretically converge to a positive constant, we can easily imagine that $r_n^{(k)}$ stops absolutely after several stages in applications, where only one sample path of X is available and the sample size n is fixed. Because, if $r_n^{(k)}$ goes to near the $r_{opt}^{(n)}$ then the difference $|r_n^{(k)} - r_n^{(k+1)}|$ is getting small, consequently, the estimated jump's number $I_k := \#\{i; |\Delta_i X^n| > r_n^{(k)}\}$ is not updated, that is, $I_k = I_{k+1}$ for sufficiently large k . As a result, the estimators $\hat{\varepsilon}_n$ and \hat{f}_n are not updated either. This leads that $r_n^{(k)} = r_n^{(k+1)}$. Actually the variance of the last stage threshold is sufficiently small, and it indicates that we can always select the good threshold uniquely.

In this simulation, we chose the pilot threshold as $r_n^{(0)} = 1.0$, but we can check that any other choice of pilot threshold which satisfies that $0 < r_n^{(0)} < \max_{1 \leq i \leq n} |\Delta_i X^n|$ can also leads the similar results.

$n = 1000$	$r_n^{(k)}$	$\hat{\lambda}_n^{(k)}$	$\hat{\mu}_n^{(k)}$	$\hat{\sigma}_n^{(k)}$
0-stage	1.0	4.68	-0.11718	1.74590
s.d.	0.0	1.082	0.319	0.186
1-stage	0.36071	10.52	-0.04073	0.51620
s.d.	0.0336	1.471	0.079	0.077
2-stage	0.09943	13.40	-0.03253	0.30624
s.d.	0.0136	1.663	0.050	0.009
3-stage	0.05836	14.38	-0.03089	0.29766
s.d.	0.0020	1.835	0.049	0.007
4-stage	0.05599	14.66	-0.03052	0.29619
s.d.	0.0020	1.946	0.048	0.007
5-stage	0.05551	14.72	-0.03043	0.29589
s.d.	0.0019	1.986	0.048	0.007
6-stage	0.05540	14.74	-0.03042	0.29582
s.d.	0.0019	2.004	0.048	0.007
7-stage	0.05540	14.74	-0.03037	0.29582
s.d.	0.0019	2.004	0.048	0.007
$r_{opt}^{(n)}/\text{True}$	0.05560	15.0	-0.03	0.3

Tab. 6.4: The 1st-7th thresholds as the sample size $n = 1000$. Each estimator is the mean over the 300 times iterations and s.d. is the standard deviation of them.

$n = 3000$	$r_n^{(k)}$	$\hat{\lambda}_n^{(k)}$	$\hat{\mu}_n^{(k)}$	$\hat{\sigma}_n^{(k)}$
0-stage	1.0	4.77	-0.09615	1.74004
s.d.	0.0	0.995	0.257	0.173
1-stage	0.26503	11.73	-0.03510	0.40342
s.d.	0.0238	1.436	0.048	0.038
2-stage	0.05592	14.16	-0.03331	0.30113
s.d.	0.0048	1.702	0.037	0.004
3-stage	0.04098	14.80	-0.03315	0.29901
s.d.	0.0008	1.851	0.037	0.004
4-stage	0.04036	14.90	-0.03291	0.29874
s.d.	0.0008	1.883	0.037	0.004
5-stage	0.04028	14.92	-0.03288	0.29870
s.d.	0.0008	1.887	0.037	0.004
6-stage	0.04027	14.92	-0.03288	0.29870
s.d.	0.0008	1.888	0.037	0.004
7-stage	0.04027	14.92	-0.03288	0.29870
s.d.	0.0008	1.888	0.037	0.004
$r_{opt}^{(n)}/\text{True}$	0.04016	15.0	-0.03	0.3

Tab. 6.5: The 1st-7th thresholds as the sample size $n = 3000$. Each estimator is the mean over the 300 times iterations and s.d. is the standard deviation of them.

$n = 10000$	$r_n^{(k)}$	$\hat{\lambda}_n^{(k)}$	$\hat{\mu}_n^{(k)}$	$\hat{\sigma}_n^{(k)}$
0-stage	1.0	4.75	-0.07670	1.75522
s.d.	0.0	0.877	0.179	0.152
1-stage	0.18738	12.70	-0.03042	0.34009
s.d.	0.0150	1.283	0.034	0.016
2-stage	0.03205	14.57	-0.03005	0.30008
s.d.	0.0014	1.442	0.029	0.002
3-stage	0.02777	14.91	-0.02997	0.29957
s.d.	0.0003	1.495	0.029	0.002
4-stage	0.02760	14.95	-0.02995	0.29953
s.d.	0.0003	1.503	0.029	0.002
5-stage	0.02759	14.95	-0.02995	0.29952
s.d.	0.0003	1.504	0.029	0.002
6-stage	0.02759	14.95	-0.02995	0.29952
s.d.	0.0003	1.504	0.029	0.002
7-stage	0.02759	14.95	-0.02995	0.29952
s.d.	0.0003	1.504	0.029	0.002
$r_{opt}^{(n)}/\text{True}$	0.02752	15.0	-0.03	0.3

Tab. 6.6: The 1st-7th thresholds as the sample size $n = 10000$. Each estimator is the mean over the 300 times iterations and s.d. is the standard deviation of them.

6.5 Theoretical discussion

6.5.1 What is a validity?

In this section, we investigate the asymptotic behavior of $r_n^{(k)}$ as $k \rightarrow \infty$ with fixed $n \in \mathbb{N}$, and the one as $n \rightarrow \infty$ after $k \rightarrow \infty$.

According to the numerical studies in the previous section, $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ seems to converge for fixed n to a positive constant, which is near to the $r_{opt}^{(n)}$. Consequently, the limit of $r_n^{(k)}$ as $k \rightarrow \infty$ becomes a *good* threshold, and the corresponding filter shows a high-performance. Therefore, theoretically, we expect at least that the following properties are hold: there exists a positive constant γ_n for each $n \in \mathbb{N}$ such that

$$\lim_{k \rightarrow \infty} r_n^{(k)} = \gamma_n \quad a.s., \quad (6.22)$$

$$\lim_{n \rightarrow \infty} \left(\gamma_n + \frac{\sqrt{h_n}}{\gamma_n} \right) = 0. \quad (6.23)$$

On the first property (6.22), though our greatest hope is $\gamma_n = r_{opt}^{(n)}$ for each $n \in \mathbb{N}$, it may be impossible while n is finite. Therefore it will be desirable that

$$\Delta_n := |\gamma_n - r_{opt}^{(n)}| \quad \text{is sufficiently small for large } n. \quad (6.24)$$

The condition (6.24) is described in Theorem 6.2, (6.33) more clearly.

On the second property (6.23), this is the necessary condition for the asymptotic filter. Intuitively, the $\sqrt{h_n}$ -order means the order of the expected variation of Brownian shocks. Therefore it is desired that the speed of the convergence of the threshold is slower than $\sqrt{h_n}$. In Shimizu and Yoshida [96], they used the asymptotic filter as $\{|\Delta_i X^n| > Lh_n^\rho\}$ for a constant $L > 0$ and $\rho \in (0, 1/2)$, and this threshold certainly satisfies the condition (6.23). Furthermore, Mancini [64] recently proposed the similar type of the filter, for a constant $L > 0$, $\{|\Delta_i X^n| > L\sqrt{h_n} \log h_n^{-1}\}$, which also satisfied the above condition. Similarly we demand this condition to γ_n .

6.5.2 Mathematical validity

In order to show the mathematical validity (6.22)-(6.24) for our plug-in method, we first make the following assumption:

A 1 *The coefficient $a(x)$ and $b(x)$ of the stochastic differential equation (6.3) is known.*

This assumption implies that the function $\varepsilon(r)$ is implicitly known. First we suppose that $\varepsilon(r)$ is known. Therefore we consider the case where

$$\hat{L}_n(r; s) = 2h_n^{-1}\varepsilon(r) - \hat{J}_n(r; s). \quad (6.25)$$

We make some remarks later on the case where $\varepsilon(r)$ is unknown.

Let $\hat{I}_n(r; s)$ be an estimator of the integral $\int_{|z| \leq r} f(z) dz$ constructed in some way using the filter $\{|\Delta_i X^n| > s\}$. We assume the following.

A 2 For any $r > 0$ and any $n \in \mathbb{N}$,

$$\inf_{k \in \mathbb{N}} \hat{I}_n(r; r_n^{(k-1)}) > 0 \quad a.s., \quad (6.26)$$

$$s_1 \geq s_2 \quad \Rightarrow \quad \hat{I}_n(r; s_1) \leq \hat{I}_n(r; s_2) \quad a.s. \quad (6.27)$$

Remark 6.1 The family of such estimators $\hat{I}_n(r; s)$ with above conditions is not empty. Actually, the density estimator (6.18) with (6.19) satisfies (6.26) if

$$0 < r_n^{(k-1)} < \max_{1 \leq i \leq n} |\Delta_i X^n| \quad (6.28)$$

for any $k \in \mathbb{N}$, and (6.27) clearly holds true.

Though the following results are clear from the definition of the function $L(r)$ and $\hat{L}_n(r; r_n^{(k-1)})$, we present them as a lemma since these results will be used repeatedly.

Lemma 6.3 Suppose Condition A1. Then two functions $L(r)$ and $\hat{L}_n(r; r_n^{(k-1)}) : \mathbb{R}_+ \rightarrow \mathbb{R}$ are strictly decreasing in r for any $n, k \in \mathbb{N}$, and the equations

$$L(r) = 0 \quad \text{and} \quad \hat{L}_n(r; r_n^{(k-1)}) = 0$$

have the unique roots $r_{opt}^{(n)}$ and $r_n^{(k)}$ respectively. Therefore, in particular,

$$\begin{aligned} r_{opt}^{(n)} \leq r &\Leftrightarrow L(r) \leq 0, \\ r_n^{(k)} \leq r &\text{ a.s. } \Leftrightarrow \hat{L}_n(r; r_n^{(k-1)}) \leq 0 \quad a.s. \end{aligned}$$

The following theorem shows that the monotonicity of the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$.

Lemma 6.4 Let $k \in \mathbb{N}$. Suppose Conditions A1 and A2. Then it follows that $r_n^{(k)} \geq r_n^{(k+1)}$ for all $\omega \in \left\{ \omega \in \Omega; r_n^{(k-1)} \geq r_n^{(k)} \right\}$. Moreover it follows that $r_n^{(k)} \leq r_n^{(k+1)}$ for all $\omega \in \left\{ \omega \in \Omega; r_n^{(k-1)} \leq r_n^{(k)} \right\}$.

Proof . Fix an arbitrary $\omega \in \left\{ \omega \in \Omega ; r_n^{(k-1)} \geq r_n^{(k)} \right\}$. From the definition of $r_n^{(k)}$,

$$2h_n^{-1}\varepsilon(r_n^{(k)}) = \hat{\mathcal{J}}_n(r_n^{(k)}; r_n^{(k-1)}).$$

Noticing that $\hat{\mathcal{J}}_n(r; r_n^{(k)}) = \hat{\mathcal{I}}_n(r/2; r_n^{(k)}) + \hat{\mathcal{I}}_n(2r; r_n^{(k)})$, it follows from the conditions (6.26) and (6.27) that

$$\frac{2h_n^{-1}\varepsilon(r_n^{(k)})}{\hat{\mathcal{J}}_n(r_n^{(k)}; r_n^{(k)})} = \frac{\hat{\mathcal{J}}_n(r_n^{(k)}; r_n^{(k-1)})}{\hat{\mathcal{J}}_n(r_n^{(k)}; r_n^{(k)})} \leq 1.$$

Hence

$$\hat{L}_n(r_n^{(k)}; r_n^{(k)}) \leq 0.$$

Lemma 6.3 yields that $r_n^{(k+1)} \leq r_n^{(k)}$.

The last half of the statement follows by the same argument as above. \square

A 3 The process α , which is the solution to (6.4) satisfies for some $p \geq 1$ and any $s, t \geq 0$ with $|t - s| < 1$ that

$$E[|\alpha_t - \alpha_s|^p] \leq C_p |t - s|^{p/2}, \quad (6.29)$$

where C_p is a positive constant depending on p .

This assumption holds true if, for example, the coefficients a and b are bounded, or if X satisfies the conditions presented in Chapter 3.

A 4 For any $c \in (0, 1/2)$, there exists a constant $\delta > 0$ such that

$$\sup_{n \in \mathbb{N}} h_n^\delta \mathcal{J}^{-1}(Lh_n^c) < \infty$$

for any $L > 0$. Similarly, in empirical version,

$$\sup_{n, k \in \mathbb{N}} \frac{h_n^\delta}{\hat{\mathcal{J}}_n(Lh_n^c; r_n^{(k-1)})} < \infty \quad a.s.$$

Condition A4 is not so restrictive since (6.30) holds true if, for example, F is of polynomial order in a neighborhood of the origin and that the intensity λ is strictly positive. Moreover Condition (4) is also usually satisfied for a suitable estimator $\hat{\mathcal{I}}_n$ with (6.26).

Lemma 6.5 Suppose Conditions A1-A4, and that $h_n < 1$ for any $n \in \mathbb{N}$. Then, for any $c \in (0, 1/2)$, there exists a constant $\kappa > 0$ which is independent of n such that

$$0 < r_{opt}^{(n)}, r_n^{(k)} < \kappa h_n^c \quad (6.30)$$

for each $n \in \mathbb{N}$.

Proof . Under A4, for any $\delta > 0$, $c \in (0, 1/2)$ and some $p \geq 1$ with which A3 holds, there exists a constant $\kappa_1 > 0$ which is independent of n such that

$$\kappa_1 \geq \left(\sup_n \frac{h_n^\delta}{\mathcal{J}(\kappa_1 h_n^c)} \right)^{1/p} (2C_p)^{1/p} \geq \left(\frac{2C_p h_n^\delta}{\mathcal{J}(\kappa_1 h_n^c)} \right)^{1/p}, \quad (6.31)$$

where C_p is a constant given in A3, since $\mathcal{J}(\kappa_1 h_n^c)$ is increasing in κ_1 for each n .

For the constants κ_1 and c in (6.31), it follows from Chebysev's inequality and A3 that

$$\frac{2h_n^{-1}\varepsilon(\kappa_1 h_n^c)}{\mathcal{J}(\kappa_1 h_n^c)} \leq \frac{2C_p}{h_n \mathcal{J}(\kappa_1 h_n^c)} \left(\frac{\sqrt{h_n}}{\kappa_1 h_n^c} \right)^p \leq h_n^{p(1/2-c)-1-\delta}$$

Therefore, taking δ such as $0 < \delta < p(1/2 - c) - 1$, we find that the last term is less than 1. This implies by Lemma 6.3 that $r_{opt}^{(n)} < \kappa_1 h_n^c$.

Similarly we also see that, for a constant $\kappa_2 > 0$, $\hat{L}_n(\kappa_2 h_n^c; r_n^{(k-1)}) \leq 0$. Hence the statement holds for $\kappa = \kappa_1 \vee \kappa_2$. \square

Although the statement says for any $c \in (0, 1/2)$, we can not take $c = 1/2$ since it may be that $r_{opt}^{(n)} \sim |\log h_n| \sqrt{h_n}$. Such sequence satisfies $r_{opt}^{(n)} \leq \kappa h_n^c$ for sufficiently large n and any $c \in (0, 1/2)$. However $r_{opt}^{(n)} > \kappa h_n^{1/2}$ for sufficiently large n .

Lemma 6.6 Suppose Conditions A1-A3, and fix any $c \in (0, 1/2)$ and $k, n \in \mathbb{N}$. Then it follows that

$$r_n^{(k)} < r_n^{(k-1)}$$

for all $\omega \in \left\{ \omega \in \Omega; r_n^{(k-1)} > \kappa h_n^c \right\}$, where κ is given in Lemma 6.5.

Proof . By the similar argument as in the proof of Lemma 6.5, we see that

$$\begin{aligned} & \frac{2h_n^{-1}\varepsilon(r_n^{(k-1)})}{\hat{\mathcal{J}}_n(r_n^{(k-1)}; r_n^{(k-1)})} \\ &= 2 \left(n h_n \hat{\mathcal{J}}_n(r_n^{(k-1)}; r_n^{(k-1)}) \right)^{-1} \sum_{i=1}^n P \left\{ |X_{t_i^n} - X_{t_{i-1}^n}| > r_n^{(k-1)}, J_i^n = 0 \right\} \\ &\leq 2C_p \left(h_n \hat{\mathcal{J}}_n(\kappa h_n^c; r_n^{(k-1)}) \right)^{-1} \left(\frac{\sqrt{h_n}}{\kappa h_n^c} \right)^p < 1 \end{aligned}$$

for any $p \geq 1$.

Lemma 6.4 and 6.6 indicate a way to choose the pilot threshold. If we take $r_n^{(0)} > \kappa h_n^c$ then the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ is the decreasing almost surely. Since it is bounded

from the bottom, it converges to a limit γ_n . Moreover, even if one chooses the pilot threshold as being too large, Lemma 6.6 ensures the improvement of the threshold. Indeed, Lemma 6.5 implies that $r_n^{(1)} \leq \kappa h_n^c$, which would be nearer to $r_{opt}^{(n)}$ than $r_n^{(0)}$ since $r_{opt}^{(n)} \leq \kappa h_n^c$.

The validity (6.22) and the former of the validity (6.23) is obtained by the following theorem.

Theorem 6.2 *Suppose Conditions A1-A4. For arbitrary $r_n^{(0)}$ satisfying A2 (6.26), there exists a positive constant γ_n such that*

$$\lim_{k \rightarrow \infty} r_n^{(k)} = \gamma_n \quad a.s. \quad (6.32)$$

for any fixed $n \in \mathbb{N}$. Moreover

$$\Delta_n := |\gamma_n - r_{opt}^{(n)}| \leq \kappa h_n^c \quad (6.33)$$

for any $c \in (0, 1/2)$, where κ is given in Lemma 6.5.

Proof . By Lemma 6.4 and 6.5, we can see that the sequence $\{r_n^{(k)}\}_{k \in \mathbb{N}}$ is monotone and bounded. Therefore $r_n^{(k)}$ converges to a limit $\gamma_n \geq 0$. For this γ_n , we have $\hat{L}_n(\gamma_n; \gamma_n) = 0$. If $\gamma_n = 0$ then it must be $\hat{L}_n(0; 0) = 0$. However it contradicts that $\hat{L}_n(r; 0) = 2h_n^{-1} > 0$ for any $r \geq 0$. Hence $\gamma_n > 0$.

The inequality $\gamma_n \leq \kappa h_n^c$ is clear by Lemma 6.5. Therefore we obtain (6.33). \square

The result (6.33) was also one of the validities stated in (6.24). Therefore we find that γ_n is close to $r_{opt}^{(n)}$ if the sample size n is sufficiently large, and we can check this phenomenon in simulation results displayed in Section 6.4.

Let us consider the following quantity:

$$\mathcal{D}_\alpha^{(n)}(r_1, r_2) = \frac{1}{n} \sum_{i=1}^n P\{r_1 \wedge r_2 \leq |\Delta_i \alpha^n| \leq r_1 \vee r_2\}, \quad (6.34)$$

where α is the solution process to (6.4). If the distribution function of $|\alpha_t - \alpha_s|$ for any $t, s \geq 0$ is strictly increasing, or if the support of the probability density of $|\alpha_t - \alpha_s|$ is \mathbb{R}_+ , then $\mathcal{D}_\alpha^{(n)}(r_1, r_2)$ for fixed n can be a distance between r_1 and r_2 .

Let us consider an estimator of Lévy density f , and make a natural estimator $\hat{\mathcal{I}}_n$:

$$\hat{\mathcal{I}}_n(r; s) := \int_{|z| \leq r} \hat{f}_n(z; s) dz.$$

In this case, we can estimate $\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)})$ as in the next theorem. Although this estimate says nothing about the direct estimate of the error Δ_n , this gives us indirectly

the order conditions for γ_n and $r_{opt}^{(n)}$ which should satisfy as the asymptotic thresholds; Condition (6.23).

Theorem 6.3 *Suppose Conditions A1-A3. Let $\bar{f}_n(z; r) := \hat{f}_n(z; r) - f(z)$. Then*

$$\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \leq \sqrt{\frac{33}{2}} \kappa h_n^{1+c} \sup_{z \in \mathbb{R}^d} \|\bar{f}_n(z; \gamma_n)\|_{L^2(P)}$$

for any $c \in (0, 1/2)$, where κ is given in Lemma 6.5.

Proof . First, we suppose that $r_{opt}^{(n)} \geq \gamma_n$. By the definition of γ_n and $r_{opt}^{(n)}$, we have

$$\begin{aligned} \hat{\mathcal{J}}_n(\gamma_n; \gamma_n) &= 2h_n^{-1} \varepsilon(\gamma_n) \\ \mathcal{J}(r_{opt}^{(n)}) &= 2h_n^{-1} \varepsilon(r_{opt}^{(n)}), \end{aligned}$$

hence we obtain that

$$\begin{aligned} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) &= \varepsilon(\gamma_n) - \varepsilon(r_{opt}^{(n)}) \\ &= \frac{h_n}{2} \left[\hat{\mathcal{J}}_n(\gamma_n; \gamma_n) - \mathcal{J}(r_{opt}^{(n)}) \right] \\ &\leq \frac{h_n}{2} \left[\hat{\mathcal{J}}_n(r_{opt}^{(n)}; \gamma_n) - \mathcal{J}(r_{opt}^{(n)}) \right]. \end{aligned}$$

According to Jensen's inequality, it follows that

$$\begin{aligned} \left\{ \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \right\}^2 &\leq \frac{h_n^2 (r_{opt}^{(n)})^2}{2} \left[\left(\frac{1}{r_{opt}^{(n)}} \int_{|z| \leq r_{opt}^{(n)}/2} \bar{f}_n(z; \gamma_n) dz \right)^2 \right. \\ &\quad \left. + 16 \left(\frac{1}{4r_{opt}^{(n)}} \int_{|z| \leq 2r_{opt}^{(n)}} \bar{f}_n(z; \gamma_n) dz \right)^2 \right] \\ &\leq \frac{h_n^2 r_{opt}^{(n)}}{2} \left[\int_{|z| \leq r_{opt}^{(n)}/2} \bar{f}_n^2(z; \gamma_n) dz + 4 \int_{|z| \leq 2r_{opt}^{(n)}} \bar{f}_n^2(z; \gamma_n) dz \right]. \end{aligned}$$

Integrating the both sides by the measure P , we obtain that

$$\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \leq \sqrt{\frac{33}{2}} h_n r_{opt}^{(n)} \sup_{z \in E} \|\bar{f}_n(z; \gamma_n)\|_{L^2(P)},$$

and Lemma 6.5 yields the consequence .

When $r_{opt}^{(n)} < \gamma_n$, the same argument as above is hold since

$$\begin{aligned} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) &= \varepsilon(r_{opt}^{(n)}) - \varepsilon(\gamma_n) \\ &\leq \frac{h_n}{2} \left[\mathcal{J}(r_{opt}^{(n)}) - \hat{\mathcal{J}}_n(r_{opt}^{(n)}; \gamma_n) \right]. \end{aligned}$$

This completes the proof. \square

Let us use the kernel density estimator proposed in Chapter 5 as \hat{f}_n :

$$\hat{f}_n(z) = \frac{1}{nh_n^{1+\delta}} \sum_{i=1}^n K\left(\frac{z - \Delta z_{T_i^n}}{\delta_n}\right) \mathbf{1}_{\mathcal{H}_i^n(r_n)}, \quad (6.35)$$

where δ_n satisfies $\delta_n^d = h_n^\delta$ for a constant $\delta \in (0, 1/2)$ and K is a bounded kernel with some conditions described in Section 5.3.1. If the true density f is bounded then we find by the roughly estimate that $\sup_{z \in \mathbb{R}^d} \|\bar{f}_n(z; \gamma_n)\|_{L^2(P)} = O(h_n^{-(1+\delta)})$ for any sequence γ_n . Therefore we obtain from Lemma 6.3 that

$$\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \leq O(h_n^{c-\delta}). \quad (6.36)$$

Noticing that $c \in (0, 1/2)$ can be taken arbitrarily, we have

$$\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \rightarrow 0 \quad (n \rightarrow \infty)$$

by taking $c > \delta$. We state this fact as a corollary.

Corollary 6.1 *Suppose Conditions A1-A3, and that f is bounded. Let \hat{f}_n be the kernel density estimator (6.35). Then*

$$\limsup_{n \rightarrow \infty} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) = 0. \quad (6.37)$$

It remains for us to study the validity (6.23).

First, let us consider the simplest case where α is an one-dimensional Brownian motion: $\alpha_t = \mu t + \sigma w_t$ for constants $\mu \in \mathbb{R}$, $\sigma > 0$. Then we can show that $\sqrt{h_n}(r_{opt}^{(n)})^{-1} = o_p(1)$. Actually, considering a sequence $r_n = M_n \sqrt{h_n} + o_p(\sqrt{h_n})$ with $M_n \rightarrow M \geq 0$, we have

$$\begin{aligned} L(r_n) &= 2h_n^{-1} \varepsilon(r_n) - \mathcal{J}(r_n) \\ &\geq 2h_n^{-1} \left[\int_{r_n}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du + \int_{-\infty}^{-r_n} \frac{1}{\sqrt{2\pi\sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du \right] \\ &\quad - 2\lambda_0 \\ &\sim 2h_n^{-1} \left[\int_{M_n}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du + \int_{-\infty}^{-M_n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{u^2}{2\sigma^2}} du \right] - 2\lambda_0. \end{aligned}$$

Therefore $L(r_n) > 0$ for sufficiently large n , and Lemma 6.3 yields $r_{opt}^{(n)} > r_n$ a.s. for sufficiently large n . Since we can take M_n as an arbitrary sequence while $M_n \rightarrow M \geq 0$, it follows that $r_{opt}^{(n)}/\sqrt{h_n} \rightarrow \infty$ a.s. by $M \rightarrow \infty$.

In this situation, we can show that $\sqrt{h_n}\gamma_n^{-1} = o(1)$ from Corollary 6.1. For example, assume that $\gamma_n \sim M\sqrt{h_n}$ for a constant $M \geq 0$. For sufficiently large n , we have

$$\mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)})$$

$$\begin{aligned}
&= E \left[\int_{\gamma_n}^{r_{opt}^{(n)}} \frac{1}{\sqrt{2\pi\sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du + \int_{-r_{opt}^{(n)}}^{-\gamma_n} \frac{1}{\sqrt{2\pi\sigma^2 h_n}} e^{-\frac{(u-\mu h_n)^2}{2\sigma^2 h_n}} du \right] \\
&\sim \int_M^{r_{opt}^{(n,w)}/\sqrt{h_n}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(v-\mu\sqrt{h_n})^2}{2\sigma^2}} dv + \int_{-r_{opt}^{(n)}/\sqrt{h_n}}^{-M} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(v-\mu\sqrt{h_n})^2}{2\sigma^2}} dv \\
&\geq \Delta(M),
\end{aligned}$$

where $\Delta(M)$ is a positive constant. This contradicts Corollary 6.1. These facts are proved more generally in the next theorem.

In general case where α is a diffusion process (6.4), we suppose the following condition.

A 5 For any $t \geq 0$, the process $\alpha^{(t)}$ has the transition density $p(h, x, y)$: $P\{\alpha_{t+h}^{(t)} \in A | X_t = x\} = \int_A p(h, x, y) dy$ for any $h > 0$, such that

$$p(h, x, y) \geq \frac{K}{\sqrt{h}} \exp \left(ch|x|^2 - \frac{|x-y|^2}{ch} \right) \quad (6.38)$$

for constants $c > 1$ and $K > 1$.

Although one might think that this condition is not easy to be checked, we can replace this condition with more concrete one using the coefficients of the SDE (6.4); see Gobet [35] which gives us a sufficient condition for A5 under the elliptic diffusion case by using the coefficient of the SDE, and it can be easily checked. However we need only the fact (6.38) for our purpose, so we dare to impose this condition directly.

We obtain the following theorem, which was one of the validities presented in (6.23).

Theorem 6.4 Suppose Conditions A1-A5 and that f is bounded. Then

$$\lim_{n \rightarrow \infty} \sqrt{h_n} \left(\gamma_n^{-1} + (r_{opt}^{(n)})^{-1} \right) = 0 \quad a.s. \quad (6.39)$$

Proof . The process α follows the equation

$$\alpha_t^{(t_{i-1}^n)} = X_{t_{i-1}^n} + \int_{t_{i-1}^n}^t a(\alpha_s^{(t_{i-1}^n)}) ds + \int_{t_{i-1}^n}^t b(\alpha_s^{(t_{i-1}^n)}) dw_s,$$

for $t \in (t_{i-1}^n, t_i^n]$, and suppose that $\alpha^{(t_{i-1}^n)}$ has a transition density satisfying (6.38).

It follows from the lower bound (6.38) that

$$\begin{aligned}
&P \{ r_n \leq \Delta_i \alpha^n(t_{i-1}^n) \leq R_n \} \\
&= E \left[P \left\{ r_n + X_{t_{i-1}^n} \leq \alpha_{t_i^n}^{(t_{i-1}^n)} \leq R_n + X_{t_{i-1}^n} \middle| X_{t_{i-1}^n} \right\} \right]
\end{aligned}$$

$$\begin{aligned}
&\geq E \left[\int_{r_n + X_{t_{i-1}^n}^n}^{R_n + X_{t_{i-1}^n}^n} \frac{K}{\sqrt{h_n}} \exp \left(ch_n |X_{t_{i-1}^n}^n|^2 - \frac{|y - X_{t_{i-1}^n}^n|^2}{ch_n} \right) dy \right] \\
&= E \left[\exp \left(ch_n |X_{t_{i-1}^n}^n|^2 \right) \int_{r_n/\sqrt{h_n}}^{R_n/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy \right] \\
&\geq E \left[\int_{r_n/\sqrt{h_n}}^{R_n/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy \right].
\end{aligned}$$

Putting $r_n = M_n \sqrt{h_n} + o(\sqrt{h_n})$, where M_n is an arbitrary sequence satisfying $M_n \rightarrow M$ ($0 \leq M < \infty$), and $R_n = \infty$, then we obtain that

$$L(r_n) \geq 2h_n^{-1} \int_{M_n + o(1)}^{\infty} K e^{-\frac{y^2}{c}} dy - 2\lambda_0 > 0$$

for sufficiently large n . Hence we obtain that $r_{opt}^{(n)} > r_n$ *a.s.* by Lemma 6.3. This implies that $\sqrt{h_n}(r_{opt}^{(n)})^{-1} = o_p(1)$.

Let $r_n = \gamma_n \wedge r_{opt}^{(n)}$ and $R_n = \gamma_n \vee r_{opt}^{(n)}$. Then

$$\begin{aligned}
\mathcal{D}_\alpha^{(n)}(r_n, R_n) &= P \{ r_n \leq |\Delta_i \alpha^n(t_{i-1}^n)| \leq R_n \} \\
&= P \{ r_n \leq \Delta_i \alpha^n(t_{i-1}^n) \leq R_n \} + P \{ -R_n \leq \Delta_i \alpha^n(t_{i-1}^n) \leq -r_n \} \\
&\geq E \left[\int_{r_n/\sqrt{h_n}}^{R_n/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy + \int_{-R_n/\sqrt{h_n}}^{-r_n/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy \right].
\end{aligned}$$

If we suppose that $\gamma_n \leq r_{opt}^{(n)}$ *a.s.* for sufficiently large n , we have

$$\begin{aligned}
&\liminf_{n \rightarrow \infty} \mathcal{D}_\alpha^{(n)}(\gamma_n, r_{opt}^{(n)}) \\
&\geq E \left[\liminf_{n \rightarrow \infty} \left\{ \int_{\gamma_n/\sqrt{h_n}}^{r_{opt}^{(n)}/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy + \int_{-r_{opt}^{(n)}/\sqrt{h_n}}^{-\gamma_n/\sqrt{h_n}} K e^{-\frac{y^2}{c}} dy \right\} \right] \\
&\geq 0.
\end{aligned}$$

This implies that $\sqrt{h_n}\gamma_n^{-1} + \sqrt{h_n}(r_{opt}^{(n)})^{-1} = o(1)$ since $\sqrt{h_n}(r_{opt}^{(n)})^{-1} = o(1)$, and the right hand side of the last inequality must be zero from Corollary 6.1. When $\gamma_n \geq r_{opt}^{(n)}$ for sufficiently large n , the same argument holds. This completes the proof. \square

6.6 To a practical approach

So far, we have discussed a kind of validity of the plug-in method in the case where the function $\varepsilon(r)$ is known, and we obtained the sufficient conditions for (6.22)-(6.24).

- (i) The property (6.22) and (6.24) hold under A1-A4.
- (ii) The property (6.23) holds under A1-A5 and that f is bounded.

Of course we are also interested in the case where $\varepsilon(r)$ is unknown; $\varepsilon(r)$ is replaced by $\hat{\varepsilon}_n(r; s)$. Actually, in Section 6.4, we tried some simulation in such a case, and we obtained some good results. However we can not show yet the rigorous validity when $\varepsilon(r)$ is unknown. This is the critical problem for the future.

Nevertheless, this method is often useful. Let

$$\tilde{L}_n(r; r_n^{(k-1)}) = 2h_n^{-1}\hat{\varepsilon}_n(r; r_n^{(k-1)}) - \hat{\mathcal{J}}_n(r; r_n^{(k-1)}),$$

and note that

$$\tilde{L}_n(r; r_n^{(k-1)}) = \hat{L}_n(r; r_n^{(k-1)}) + 2h_n^{-1}\Delta\varepsilon_n(r; r_n^{(k-1)}),$$

where $\Delta\varepsilon_n(r; r_n^{(k-1)}) = \hat{\varepsilon}_n(r; r_n^{(k-1)}) - \varepsilon(r)$ and \hat{L}_n is given in (6.25).

If we choose the pilot threshold satisfying $r_n^{(0)} < \sqrt{\kappa h_n}$ and (6.28) then many simulation shows that the estimator $\hat{b}_n(x; r_n^{(0)})$ becomes relatively robust; see also Section 5.4.3, and $\hat{\varepsilon}_n(r; r_n^{(k-1)})$ gives a good approximation of $\varepsilon(r)$. Consequently, $\Delta\varepsilon_n$ becomes almost zero, and this implies that finding the root of $\tilde{L}_n(r; r_n^{(0)}) = 0$ is similar to finding the root of $\hat{L}_n(r; r_n^{(0)}) = 0$ in the 1st-stage. The same argument is possible after this stage, and the discussion in the previous section can be approximately applied to the case where $\varepsilon(r)$ is unknown.

These are the empirical and intuitive explanation why the plug-in rule with unknown $\varepsilon(r)$ works well, and the more rigorous study is desired. Moreover the case where the jump part is $c(X_{t-})dz_t$, or more general case as in (3.1) should be also studied in the future.

Appendix A

Central limit theorems for arrays

The aim of this section is to present some versions of the central limit theorem for general triangular arrays, and give a result which is the most relevant from the point of view of applications, in particular, the version used in Chapter 3 and 4. The central limit theorems for arrays are often useful when we discuss the asymptotic normality for discretely observed time-continuous stochastic processes.

Before we introduce a general result, we present the central limit theorem for martingale difference arrays under certain assumptions. Then we discuss several types of sufficient conditions with which some of those assumptions can be substitute. Finally, we shall introduce the theorem for general arrays with checkable conditions.

All the facts here are well known, and one can find the details on the central limit theorem for arrays in e.g. Hall and Heyde [39], Jacod and Shiriyayev [43] and Shiriyayev [97], and so on. However we give this appendix to make this thesis self-contained.

A.1 Martingale difference arrays

Let $X^n = \{X_i^n\}_{i=1}^{k_n}$ be a family of 1-dimensional random variables defined on a probability space (Ω, \mathcal{F}, P) and let $(\mathcal{F}_i^n)_{0 \leq i \leq k_n}$ be a filtration such that X_i^n is \mathcal{F}_i^n -measurable for each $n, i \in \mathbb{N}$.

Definition A.1 *An array X^n is a martingale difference array if and only if*

$$E[X_i^n | \mathcal{F}_{i-1}^n] = 0$$

for any $1 \leq i \leq k_n$.

Let $L_n = \sum_{i=1}^n X_i^n$. Note that this sum becomes a martingale with respect to the filtration $(\mathcal{F}_i^n)_{0 \leq i \leq k_n}$ if X^n is a martingale difference array. The following basic result is found in McLeish [69].

Theorem A.1 *Suppose that X^n is a martingale difference array satisfying following conditions:*

$$\sup_{n \in \mathbb{N}} E \left[\max_{1 \leq i \leq k_n} |X_i^n|^2 \right] < \infty, \quad (\text{A.1})$$

$$\max_{1 \leq i \leq k_n} |X_i^n| \xrightarrow{p} 0, \quad (\text{A.2})$$

$$\sum_{i=1}^n (X_i^n)^2 \xrightarrow{p} 1. \quad (\text{A.3})$$

Then $L_n \xrightarrow{d} \mathcal{N}(0, 1)$.

Condition (A.2) is called the *asymptotic negligibility*, which is equivalent to the weak Lindberg condition: $\sum_{i=1}^{k_n} (X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \xrightarrow{p} 0$ as $n \rightarrow \infty$ for all $\varepsilon > 0$, since

$$P \left\{ \max_{1 \leq i \leq k_n} |X_i^n| > \varepsilon \right\} = P \left\{ \sum_{i=1}^{k_n} (X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} > \varepsilon^2 \right\}.$$

Moreover they are also equivalent to

$$\sum_{i=1}^{k_n} P \left\{ |X_i^n| > \varepsilon \mid \mathcal{F}_{i-1}^n \right\} \xrightarrow{p} 0;$$

see Shirayev [97], Theorem VII.7.2. The negligibility can be lead from the following Lindberg condition: for all $\varepsilon > 0$,

$$\sum_{i=1}^{k_n} E \left[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \right] \rightarrow 0 \quad (n \rightarrow \infty) \quad (\text{A.4})$$

by the Chebyshev's inequality. Condition (A.4) implies (A.1) since

$$E \left[\max_{1 \leq i \leq k_n} |X_i^n|^2 \right] \leq \varepsilon^2 + \sum_{i=1}^{k_n} E \left[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \right] \quad (\text{A.5})$$

for all $\varepsilon > 0$. Furthermore some authors have imposed the following conditional Lindberg condition instead of them:

$$\sum_{i=1}^{k_n} E \left[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \mid \mathcal{F}_{i-1}^n \right] \xrightarrow{p} 0 \quad (n \rightarrow \infty). \quad (\text{A.6})$$

This also yields (A.1) and, at the same time, this implies the asymptotic negligibility, too. In fact, we note the following lemma

Lemma A.1 *Let $\{\mathcal{G}_i\}_{0 \leq i \leq n}$ be a filtration and $A_i \in \mathcal{G}_i$ be an event for each $0 \leq i \leq n$. Then for any $\varepsilon > 0$,*

$$P \left\{ \bigcup_{i=0}^n A_i \middle| \mathcal{G}_0 \right\} \leq \varepsilon + P \left\{ \sum_{i=1}^n P \{A_i | \mathcal{G}_{i-1}\} > \varepsilon \middle| \mathcal{G}_0 \right\}. \quad (\text{A.7})$$

See Hall and Hyde [39], Lemma 2.5 for the proof. This lemma and the conditional version of Chebyshev's inequality yields that

$$P \left\{ \max_{1 \leq i \leq k_n} |X_i^n| > \varepsilon \right\} \leq \varepsilon + P \left\{ \sum_{i=1}^n E \left[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} \middle| \mathcal{F}_{i-1}^n \right] > \varepsilon^3 \right\}.$$

This inequality and (A.6) implies (A.2).

Clearly Condition (A.6) can be replaced by the conditional Liapnov condition: for a constant $\delta > 0$,

$$\sum_{i=1}^{k_n} E \left[|X_i^n|^{2+\delta} \middle| \mathcal{F}_{i-1}^n \right] \xrightarrow{p} 0 \quad (n \rightarrow \infty). \quad (\text{A.8})$$

Condition (A.3) corresponds to the variance estimates, and under Condition (A.6), we can replace it by the following conditional variance version:

$$\sum_{i=1}^{k_n} E \left[(X_i^n)^2 \middle| \mathcal{F}_{i-1}^n \right] \xrightarrow{p} 1 \quad (n \rightarrow \infty). \quad (\text{A.9})$$

In fact, by the same argument as in the proof of Theorem 2.23 in Hall and Hyde [39], we have, without any martingale properties, that

$$P \left\{ \left| \sum_{i=1}^{k_n} \{ (X_i^n)^2 - E \left[(X_i^n)^2 \middle| \mathcal{F}_{i-1}^n \right] \} \right| > \eta \right\} \leq \frac{\varepsilon}{\eta}$$

for arbitrary $\varepsilon > 0$. Therefore such replacements of conditions are possible for any array X^n .

For the usefulness to obtain our desired general version later, we show the revised version of Theorem A.1.

Theorem A.2 *Suppose that X^n is a martingale difference array satisfying Condition (A.9). Moreover assume Condition (A.6) or (A.8). Then $L_n \xrightarrow{d} \mathcal{N}(0, 1)$.*

A.2 General triangular arrays

In this section, we do not suppose that X^n is the martingale difference, that is, the sequence X^n is a general triangular array. In order to use the preceding results, it is useful to decompose L_n as follows:

$$L_n = \sum_{i=1}^{k_n} Y_i^n + \sum_{i=1}^{k_n} E[X_i^n | \mathcal{F}_{i-1}^n],$$

where $Y_i^n = X_i^n - E[X_i^n | \mathcal{F}_{i-1}^n]$. Then Y^n is a martingale difference.

Suppose that X^n satisfies Conditions (A.1) - (A.3), and $\sum_{i=1}^{k_n} \{E[X_i^n | \mathcal{F}_{i-1}^n]\}^j \xrightarrow{p} 0$ for $j = 1, 2$. Then Y^n satisfies (A.6) and (A.9). In fact,

$$\begin{aligned} & \sum_{i=1}^{k_n} E[(Y_i^n)^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] \\ & \leq 2 \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] + 2 \sum_{i=1}^{k_n} \{E[X_i^n | \mathcal{F}_{i-1}^n]\}^2. \end{aligned}$$

Here we have to show $\sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] \xrightarrow{p} 0$.

$$\begin{aligned} & \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] \\ & = \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{2(X_i^n)^2 + 2\{E[X_i^n | \mathcal{F}_{i-1}^n]\}^2 > \varepsilon^2\}} | \mathcal{F}_{i-1}^n] \\ & \leq \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon/2\}} | \mathcal{F}_{i-1}^n] + \sum_{i=1}^{k_n} E[(X_i^n)^2 | \mathcal{F}_{i-1}^n] \mathbf{1}_{\{|E[X_i^n | \mathcal{F}_{i-1}^n]| > \varepsilon/2\}} \\ & \leq \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon/2\}} | \mathcal{F}_{i-1}^n] + \sum_{i=1}^{k_n} E[(X_i^n)^2 | \mathcal{F}_{i-1}^n] \sum_{i=1}^{k_n} \mathbf{1}_{\{|E[X_i^n | \mathcal{F}_{i-1}^n]| > \varepsilon/2\}} \\ & \leq \sum_{i=1}^{k_n} E[(X_i^n)^2 \mathbf{1}_{\{|X_i^n| > \varepsilon/2\}} | \mathcal{F}_{i-1}^n] + \sum_{i=1}^{k_n} E[(X_i^n)^2 | \mathcal{F}_{i-1}^n] \sum_{i=1}^{k_n} 4\{E[X_i^n | \mathcal{F}_{i-1}^n]\}^2 \varepsilon^{-2} \\ & \xrightarrow{p} 0. \end{aligned}$$

Therefore we obtain that $\sum_{i=1}^{k_n} E[(Y_i^n)^2 \mathbf{1}_{\{|Y_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] \xrightarrow{p} 0$. Furthermore

$$\sum_{i=1}^{k_n} E[(Y_i^n)^2 | \mathcal{F}_{i-1}^n] = \sum_{i=1}^{k_n} E[(X_i^n)^2 | \mathcal{F}_{i-1}^n] - \sum_{i=1}^{k_n} \{E[X_i^n | \mathcal{F}_{i-1}^n]\}^2 \xrightarrow{p} 1.$$

These considerations lead the fact that $L_n \xrightarrow{d} \mathcal{N}(0, 1)$.

Generally, we have the next theorem.

Theorem A.3 *Let C be a $d \times d$ deterministic matrix, and $\varepsilon > 0$. Suppose a d -dimensional triangular array X^n satisfies the following conditions:*

$$\sum_{i=1}^{k_n} E [X_i^n | \mathcal{F}_{i-1}^n] \xrightarrow{p} 0, \quad (\text{A.10})$$

$$\sum_{i=1}^{k_n} |E [X_i^n | \mathcal{F}_{i-1}^n]|^2 \xrightarrow{p} 0, \quad (\text{A.11})$$

$$\sum_{i=1}^n E [X_i^n (X_i^n)^* | \mathcal{F}_{i-1}^n] \xrightarrow{p} C, \quad (\text{A.12})$$

$$\sum_{i=1}^{k_n} E [|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] \xrightarrow{p} 0. \quad (\text{A.13})$$

Then $L_n \xrightarrow{d} \mathcal{N}_d(0, C)$.

The similar version of this theorem is also found in Shirayev [97], Theorem VII.8.1 where the sufficient conditions are essentially the same as in the above ones. Although a more general version; the case where C is a random variable, is found in Hall and Hyde [39], Theorem 3.4 without the detailed proof, the above version is enough for our purpose in this thesis.

Remark A.1 Instead of (A.13), it is sometimes convenient to assume (A.8), which is more restrictive than (A.13) but is relatively tractable in applications. Conditions (A.10) and (A.11) in Theorem A.3 can be replaced by the following:

$$\sum_{i=1}^{k_n} |E [X_i^n | \mathcal{F}_{i-1}^n]| \xrightarrow{p} 0. \quad (\text{A.14})$$

Indeed, it derives (A.10) obviously, and

$$\begin{aligned} & \sum_{i=1}^{k_n} |E [X_i^n | \mathcal{F}_{i-1}^n]|^2 \\ &= \sum_{i=1}^{k_n} |E [X_i^n (\mathbf{1}_{\{|X_i^n| > \varepsilon\}} + \mathbf{1}_{\{|X_i^n| \leq \varepsilon\}}) | \mathcal{F}_{i-1}^n]|^2 \\ &\leq 2 \sum_{i=1}^{k_n} E [|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n] + 2 \sum_{i=1}^{k_n} |E [X_i^n \mathbf{1}_{\{0 < |X_i^n| \leq \varepsilon\}} | \mathcal{F}_{i-1}^n]|^2 \end{aligned}$$

$$\begin{aligned}
&\leq 2 \sum_{i=1}^{k_n} E \left[|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n \right] + 2 \sum_{i=1}^{k_n} \left| E \left[X_i^n \mathbf{1}_{\{0 < |E[X_i^n | \mathcal{F}_{i-1}^n]| \leq \varepsilon\}} | \mathcal{F}_{i-1}^n \right] \right|^2 \\
&= 2 \sum_{i=1}^{k_n} E \left[|X_i^n|^2 \mathbf{1}_{\{|X_i^n| > \varepsilon\}} | \mathcal{F}_{i-1}^n \right] + 2\varepsilon \sum_{i=1}^{k_n} |E[X_i^n | \mathcal{F}_{i-1}^n]| \\
&\xrightarrow{p} 0
\end{aligned}$$

under Conditions (A.13) and (A.14).

Appendix B

Weak Convergence in C -space

This appendix is also to make this thesis self-contained as well as Appendix B. This chapter will help us to understand the sufficient condition for the uniform convergence of random functions on the compact parameter space, which was appeared in Chapter 3 and 4; see Remark B.1. We referred to Billingsley [10], Ibragimov and Has'minskii [41] and Kallenberg [47] for writing this appendix.

B.1 Weak Convergence and tightness

The aim of this section is to introduce the notion of the *tightness*, which is one of the *good* properties for a probability measure. When we consider the weak convergence for probability measures in a general metric space, the concept of the tightness is essential. In this section, we particularly consider the C -space; the space which consists of every continuous functions on some compact metric space and endowed with the supremum norm, and describe the relation between the tightness and the weak convergence. In the last section, we introduce some tractable tightness criteria.

First of all, we give several types of definitions of the tightness.

Definition B.1 *A probability measure P on a measurable space (E, \mathcal{E}) is tight if and only if, for every $\varepsilon > 0$, there exists a compact set $K \in \mathcal{E}$ such that*

$$P(K) > 1 - \varepsilon. \tag{B.1}$$

Moreover, a family of distributions Π on E is tight if and only if, for every $\varepsilon > 0$, there exists a compact set $K \in \mathcal{E}$ such that the inequality (B.1) is hold for every $P \in \Pi$. Furthermore E -valued random elements $\{X_\lambda\}_{\lambda \in \Lambda}$ is tight if and only if the family of distributions $\{P_\lambda\}_{\lambda \in \Lambda}$ is tight.

Theorem B.1 *If the space E is a Polish space, that is, separable and complete, then every probability measure P on (E, \mathcal{E}) is tight .*

The tightness of a sequence of random variables $\{X^n\}_{n \in \mathbb{N}}$ is equivalent to the following:

$$\sup_{K: \text{compact}} \liminf_{n \rightarrow \infty} P\{X^n \in K\} = 1. \quad (\text{B.2})$$

Indeed if we assume (B.2) then, for every $\varepsilon > 0$, there exists a compact K' such that $\liminf_{n \rightarrow \infty} P_n(K') > 1 - \varepsilon$, where $P_n = P \circ (X^n)^{-1}$. Therefore, for sufficiently large m , $\inf_{n > m} P_n(K') > 1 - \varepsilon$. On the other hand, for every $n \leq m$, there exists a compact set K'' that is independent of n such that $\inf_{1 \leq n \leq m} P_n(K'') > 1 - \varepsilon$. Consequently, putting $K = K' \cup K''$, we obtain that $P_n(K) > 1 - \varepsilon$ for every n . The necessity will be clear .

One of the simplest cases is when $E = \mathbb{R}$. In this case, the tightness of $\{X^n\}_{n \in \mathbb{N}}$ is obviously equivalent to

$$\lim_{r \rightarrow \infty} \sup_n P\{X^n > r\} = 0. \quad (\text{B.3})$$

This is called the *uniformly tightness*.

In the sequel, let (K, d) be a metric space with a metric d , and suppose K is compact. Moreover let (S, ρ) be a metric space with metric ρ , and suppose S is complete and separable . We denote by $C(K, S)$ the set of all continuous functions from K to S endowed with the uniform metric $\hat{\rho}(x, y) = \sup_{t \in K} \rho(x_t, y_t)$. We sometimes write $C(K, S)$ as simply C . Furthermore, denote by π_t the mapping $\pi_t : C(K, S) \ni x \mapsto x_t \in S$. Then the following theorem holds.

Theorem B.2

$$\mathcal{B}(C(K, S)) = \sigma\{\pi_t; t \in K\}, \quad (\text{B.4})$$

where $\mathcal{B}(X)$ means the set of all Borel subsets of X .

The σ -algebra in the right-hand side is called Kolmogorov's σ -algebra. This theorem holds when K is not compact if we choose a metric ρ suitably; see Ito [42], Theorem 5.2.

By Theorem B.2, that $x \in C(K, S)$ is measurable is equivalent to that x_t is a random variable. Indeed, if we suppose x is a measurable mapping from a probability space (Ω, \mathcal{F}, P) to C ; $x^{-1}\mathcal{B}(C) \subset \mathcal{F}$, then $x_t^{-1}\mathcal{S} = x^{-1}(\pi_t^{-1}\mathcal{S}) = x^{-1}(\mathcal{B}(C)) \subset \mathcal{F}$. This implies that x_t is \mathcal{F} -measurable. The mapping x is called an S -valued *random*

function in the sense that an S -valued function x_t is determined for given ω , and this is so-called an S -valued *stochastic process*. We also call it an S -valued *random element* as the generalization of the term a “random variable”.

Lemma B.1 *Let (S, \mathcal{S}) be a measurable space and T be a compact subspace of \mathbb{R}_+ . Let $X = \{X_t\}_{t \in T}$ and $Y = \{Y_t\}_{t \in T}$ be S -valued stochastic processes, that is, X and Y are $C(T, S)$ -valued random elements. Then $X \stackrel{d}{=} Y$ is equivalent to $(X_{t_1}, \dots, X_{t_n}) \stackrel{d}{=} (Y_{t_1}, \dots, Y_{t_n})$ for every $t_1, \dots, t_n \in T$ and $n \in \mathbb{N}$.*

By this lemma, we find that a distribution of a kind of stochastic processes can be completely characterized by its finite dimensional distribution. Generally, the necessary and sufficient condition that a sequence of distributions on an infinite dimensional space as $(\mathbb{R}^\infty, \mathcal{B}^\infty)$ converges to a distribution weakly is that any finite dimensional distribution of its distribution family converges to the finite dimensional distribution of the corresponding limit; see Billingsley [10] for details .

Similarly, one might think that Lemma B.1 seems to imply that a sequence of distributions of any stochastic process converges weakly to a limit if and only if a sequence of their finite dimensional distributions converges weakly to the finite dimensional distribution of the corresponding limit. However the intuition is not correct; see a counterexample in Billingsley [10], page 20. Since a *dimension* of the path space of such a stochastic processes is *higher* than the one of \mathbb{R}^∞ , we need a further tight condition. In fact, the *relatively compactness* described below is important in the weak convergence in stochastic processes.

Definition B.2 *A sequence of random elements $\{X^n\}_{n \in \mathbb{N}}$ is relatively compact if and only if, for an arbitrary subsequence, there exists a sub-subsequence that converges weakly.*

We write $X^n \xrightarrow{fd} X$ if the finite dimensional distributions of a sequence of random elements X^n 's converges weakly to that of X .

Theorem B.3 *Let X and $\{X^n\}_{n \in \mathbb{N}}$ be $C(K, S)$ -valued random elements. Then $X^n \xrightarrow{d} X$ if and only if $\{X^n\}_{n \in \mathbb{N}}$ is relatively compact and $X^n \xrightarrow{fd} X$.*

Proof . First, we show the necessity. Consider the following canonical projection π :

$$\pi_{t_1, \dots, t_k} : X \in C(K, S) \mapsto (X(t_1), \dots, X(t_k)) \in \mathbb{R}^k.$$

Since this is clearly continuous, $\pi_{t_1, \dots, t_k}(X^n) \xrightarrow{d} \pi_{t_1, \dots, t_k}(X)$. Therefore $X^n \xrightarrow{fd} X$. The relatively compactness is obvious.

Next, we show the sufficiency. Assume that $X^n \not\stackrel{d}{\rightarrow} X$. Then, for every $\varepsilon > 0$, there exist $N' \subset \mathbb{N}$ and a continuous bounded function $f : C(K, S) \rightarrow \mathbb{R}$ such that $|Ef(X^n) - Ef(X)| > \varepsilon$ for any $n \in N'$. On the other hand, by the relatively compactness, there exists a sub-subsequence $N'' \subset N'$ such that $\{X^n\}_{n \in N''}$ converges to a process Y weakly. By the assumption that $X^n \xrightarrow{fd} X$, ($n \in N''$), we see that $Y \stackrel{d}{=} X$ by Theorem B.2. This implies that $X^n \xrightarrow{d} X$, $n \in N''$. This contradicts the assumption. \square

There are Prohorov's theorem and Ascoli-Arzerà's theorem to judge the relatively compactness.

Theorem B.4 (Prohorov) *If an arbitrary family Π of probability measures on a measurable space (E, \mathcal{E}) is tight then Π is relatively compact. Moreover a space E is Polish then a family Π that is relatively compact is tight.*

Theorem B.5 (Ascoli-Arzerà) *Let*

$$w(x, h) = \sup\{\rho(x_t, x_s); d(s, t) \leq h\} \quad (\text{B.5})$$

for any $x \in C(K, S)$ and $h > 0$. Suppose that there exists a dense subset D of K . Then the following (i) and (ii) are equivalent:

- (i) *$A \subset C(K, S)$ is relatively compact, that is, \bar{A} is compact.*
- (ii) *For any $t \in D$, $\pi_t A$ is a relatively compact subset of S and $\limsup_{h \rightarrow 0} \limsup_{x \in A} w(x, h) = 0$.
In particular, $\bigcup_{t \in K} \pi_t A$ is a relatively compact subset of S .*

B.2 Tightness criteria

In this section, we remain to suppose that (K, d) be a metric space with a metric d , and that K is compact. Moreover suppose (S, ρ) be a metric space with metric ρ , and that S is complete and separable. We denote by $C(K, S)$ the set of all continuous functions from K to S endowed with the uniform metric $\hat{\rho}(x, y) = \sup_{t \in K} \rho(x_t, y_t)$.

There is the following criterion for the tightness; see Kallenberg [47] Theorem 14.5.

Theorem B.6 (Tightness Criteria) *$C(K, S)$ -valued random elements $\{X^n\}_{n \in \mathbb{N}}$ are tight if and only if S -valued random variables $\{X^n(t)\}_{n \in \mathbb{N}}$ are tight for any $t \in K$ and*

$$\lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} E[w(X^n, h) \wedge 1] = 0, \quad (\text{B.6})$$

where w is given in (B.5).

Proof . Assume that $\{X^n\}_{n \in \mathbb{N}}$ is tight. Then, for arbitrary $\varepsilon > 0$, there exists a compact set $B \subset C(K, S)$ such that

$$\limsup_{n \rightarrow \infty} P\{X^n \in B^c\} < \varepsilon.$$

By the Ascoli-Arzerà's theorem, for the above ε , there exists a constant $h > 0$ such that $w(x, h) \leq \varepsilon$ for any $x \in B$. This implies that $\{w(X^n, h) > \varepsilon\} \subset \{X^n \in B^c\}$, hence $\limsup_{n \rightarrow \infty} P\{w(X^n, h) > \varepsilon\} < \varepsilon$. The tightness of each family of marginal distributions is clear from Theorem B.3 and B.4.

Next we assume that the equality (B.6). Then $w(X^n, h) \rightarrow 0$ a.s. as $h \rightarrow 0$ since the path of X^n is continuous almost surely. By (B.6), for every $\varepsilon > 0$, there exist some h_0 and n_0 such that $\sup_{k \leq n_0} E[w(X_k, h) \wedge 1] < \varepsilon$ for any $h \geq h_0$. Hence (B.6) is equivalent to

$$\limsup_{h \rightarrow 0} \sup_{n \in \mathbb{N}} E[w(X^n, h) \wedge 1] = 0.$$

Therefore, for every $\varepsilon > 0$, there exist $h_1, h_2, \dots > 0$ such that

$$\sup_{n \in \mathbb{N}} P\{w(X^n, h_k) > 2^{-k}\} \leq 2^{-k-1}\varepsilon, \quad k \in \mathbb{N}. \quad (\text{B.7})$$

Moreover, since the family of distributions of $X^n(t)$ for each $t \geq 0$ is tight, there exist compact sets $C_1, C_2, \dots \subset S$ such that

$$\sup_n P\{X^n(t_k) \in C_k^c\} \leq 2^{-k-1}\varepsilon, \quad k \in \mathbb{N}, \quad (\text{B.8})$$

where $\{t_1, t_2, \dots\}$ is a dense subset in K . Here, putting

$$B = \bigcap_{k \in \mathbb{N}} \{x \in C(K, S); x(t_k) \in C_k, w(x, h_k) \leq 2^{-k}\}$$

we see from Ascoli-Arzerà's theorem again that \bar{B} is compact. The equality (B.7) and (B.8) yield that $\sup_{n \in \mathbb{N}} P\{X^n \in B^c\} \leq \varepsilon$, which means that $\{X^n\}_{n \in \mathbb{N}}$ are tight. \square

The following lemma, which is easily deduced from the above theorem, is useful in applications.

Corollary B.1 *Let $\{X^n\}_{n \in \mathbb{N}}$ be $C(K, S)$ -valued random elements, and suppose that the path of $X^n(t)$ is differentiable with respect to $t \in K$. Then Condition (B.6) holds true if*

$$\sup_{n \in \mathbb{N}} E \left[\sup_{t \in K} |\partial_t X^n(t)| \right] < \infty. \quad (\text{B.9})$$

Proof .

$$\begin{aligned}
\lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} E[w(X^n, h) \wedge 1] &= \lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} E \left[\sup_{|t-t'| \leq h} |X^n(t) - X^n(t')| \wedge 1 \right] \\
&= \lim_{h \rightarrow 0} \limsup_{n \rightarrow \infty} E \left[\sup_{|t-t'| \leq h} |\partial_t X^n(\tilde{t})| |t - t'| \wedge 1 \right] \\
&\leq \lim_{h \rightarrow 0} h \sup_{n \in \mathbb{N}} E \left[\sup_{t \in K} |\partial_t X^n(t)| \right] \quad \square
\end{aligned}$$

Remark B.1 If $X^n(\cdot) \xrightarrow{d} x(\cdot)$ in $C(K, S)$ and $x \in C(K, S)$ is deterministic then $X^n(\cdot) \xrightarrow{P} x(\cdot)$, which implies that, for any $\varepsilon > 0$, $P\{\hat{\rho}(X^n, x) > \varepsilon_n\} \rightarrow 0$ as $n \rightarrow \infty$, that is,

$$\sup_{t \in K} \rho(X^n(t), x(t)) \xrightarrow{P} 0.$$

as $n \rightarrow \infty$.

There are other types of tightness criteria. In the sequel, we suppose that $C(\mathbb{R}^d, S)$ is endowed with the topology of the uniform convergence on compacts.

Theorem B.7 *Let $\{X^n\}_{n \in \mathbb{N}}$ be $C(\mathbb{R}^d, S)$ -valued random elements. Then $\{X^n\}_{n \in \mathbb{N}}$ is tight in $C(\mathbb{R}^d, S)$ if the following conditions are hold:*

(i) $\{X^n(0)\}_{n \in \mathbb{N}}$ is tight in S .

(ii) For constants $a, b, c > 0$, $\sup_{n \in \mathbb{N}} E|\rho(X^n(s), X^n(t))|^a \leq c|s - t|^{d+b}$, $s, t \in \mathbb{R}^d$.

The following criterion is seen in Ibragimov and Has'minskii [41] Appendix I, Theorem 20.

Theorem B.8 *Let $\{X^n\}_{n \in \mathbb{N}}$ be $C(\mathbb{R}^d, S)$ -valued random elements. Then $\{X^n\}_{n \in \mathbb{N}}$ is tight in $C(\mathbb{R}^d, S)$ if there exist constants $m \geq r > d$, $H > 0$ such that the following conditions are hold:*

(i) $E[|\rho(X^n(t), 0)|^m] \leq H$.

(ii) $\sup_{n \in \mathbb{N}} E|\rho(X^n(t+h), X^n(t))|^m \leq H|h|^r$.

This is sometimes more tractable than Theorem B.7 in applications; it would be easier to check (i) of Theorem B.8 than to check (i) of Theorem B.7.

Bibliography

- [1] Aase, K. K. and Guttorp, P. (1987). Estimation in Models for Security Prices, *Bernoulli*, **4**, 211–224.
- [2] Akritas, M. G. and Johnson, R. A. (1981). Asymptotic inference in Levy processes of the discontinuous type, *Ann. Statist.*, **9**, no. 3, 604–614.
- [3] Athreya, K. B. and Keiding, N. (1977). Estimation theory for continuous-time branching processes. Asymptotic inference in Levy processes of the discontinuous type, *Sankhya Ser. A*, **39**, no. 2, 101–123.
- [4] Asmussen, S. and Rosinski, J. (2001). Approximations of small jumps of Lévy processes with a view towards simulation, *J. Appl. Probab.*, **38**, 482–493.
- [5] Bandi, F. M. and Nguyen, T. H. (2003). On the functional estimation of jump-diffusion models, *J. Econometrics*, **116**, 211–229.
- [6] Basawa, I. V. and Brockwell, P. J. (1978). Inference for gamma and stable processes, *Biometrika*, **65**, 129–133.
- [7] Basawa, I. V. and Prakasa Rao, B. L. S. (1980). *Statistical inference for stochastic processes*, Probability and Mathematical Statistics, Academic Press, Inc., London-New York.
- [8] Basawa, I. V. and Scott, D. J. (1983). *Asymptotic Optimal Inference for Non-ergodic Models*, Springer-Verlag, New York.
- [9] Billingsley, P. (1961). *Statistical Inference for Markov processes*, University of Chicago Press.
- [10] Billingsley, P. (1968). *Convergence of Probability Measures*, Willey, New York.

- [11] Barndorff-Nielsen, O. E. and Sørensen, M. (1994). A review of some aspects of asymptotic likelihood theory for stochastic processes, *Int. Stoch. Review*, **62**, 133–165.
- [12] Bibby, B. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes, *Bernoulli*, **1**, 17–39.
- [13] Bichteler, K. and Jacod, J. (1986). Calcul de Malliavin pour les diffusions avec sauts: existence d’une densité dans le cas unidimensionnel, *Lecture Notes in Math*, Springer-Verlag, Berlin, 132–157.
- [14] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities, *J. Polit. Economy*, **81**, 637–654.
- [15] Blackwell, D. (1951). Comparison of experiments, *Proc. 2nd Berkeley Symp. Math. Statist.*, Springer-Verlag, Berlin, 132–157.
- [16] Bodo, B. A. and Thomason, M. E. (1987). A review on stochastic differential equations for applications in hydrology, *Stochastic Hydrol. Hydraul.*, **1**, 81–100.
- [17] Brillinger, D. R. (1975). Statistical inference for stationary point processes, *Stochastic processes and related topics*, **1**, Academic Press, New York, 55–99.
- [18] Brown, M. (1972). Statistical analysis of non-homogeneous Poisson processes, *Stochastic point processes: statistical analysis, theory, and applications*, Wiley-Interscience, New York, 67–89.
- [19] Cramér, H. (1946). Contribution to the theory of statistical estimation, *Skand. Aktuariidskr.*, **29**, 85–94.
- [20] Cramér, H. (1949). *Mathematical method of statistics*, Princeton University Press, Princeton.
- [21] Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficient of a diffusion from discrete observations, *Stochastics*, **19**, 263–284.
- [22] 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.
- [23] Embrechts, P. and Schmidli, H. (1994). Ruin estimation for a general insurance risk model, *Adv. Appl. Prob.*, **26**, 404–422.

- [24] Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence*, Wiley.
- [25] Feigin, P. D. (1977). A note on maximum likelihood estimation for simple branching processes, *Austral. J. Statist.*, **19**, no. 2, 152–154.
- [26] Feuerverger, A. (1990). An efficiency result for the empirical characteristic function in stationary time-series models, *Canad. J. Statist.*, **18**, no. 2, 155–161.
- [27] Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics, *Phil. Trans. Roy. Soc. Ser. A*, **222**, 309–368.
- [28] Fisher, R. A. (1956). Statistical methods and scientific induction, *J. Roy. Statist. Soc. Ser. B*, **17**, 69–78.
- [29] Florens-Zmirou, D. (1989). Approximate discrete schemes for statistics of diffusion processes, *Statistics*, **20**, 547–557.
- [30] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**, 99–116.
- [31] Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion process, *Ann. Inst. Henri Poincaré Prob. Statist.*, **29**, 119–151.
- [32] Gerber, H. U. (1979). *An introduction to mathematical risk theory*, S.S. Heubner Foundation Monograph Series, 8. University of Pennsylvania, Wharton School.
- [33] Gihman, I. I. and Skorohod, A. V. (1972). *Stochastic Differential Equation*, Springer-Verlag, Berlin.
- [34] Gnedenko, B. V. and Kolmogorov, A. N. (1954). *Limit Distributions for Sums of Independent Random Variables*, Addison-Wesley, Reading.
- [35] Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations, *Ann. Inst. H. Poincaré Probab. Statist.*, **38**, no. 5, 711–737.
- [36] Grenander, U. (1950). Stochastic processes and statistical inference, *Ark. Mat.*, **1**, 195–277.
- [37] Gukhal, C. R. (2001). Analytical valuation of American options on jump-diffusion processes, *Math. Finance*, **1**, 220–242.

- [38] Guttrop, P. and Kulperger, R. (1984). Statistical inference for some Volterra population processes in a random environment, *Can. J. Statist.*, **12**, 289–302.
- [39] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and its Applications*, Academic Press, New York.
- [40] Hanson, F. B. and Tuckwell, H. C. (1981). Logistic growth with random density independent disasters, *Theor. Pop. Biol.*, **19**, 1–18.
- [41] Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*, Springer-Verlag, Berlin.
- [42] Ito, K. (1990). *Probability theory*, Iwanami sugaku sensho (in Japanese) .
- [43] Jacod, J. and Shiriyayev, A. N. (1987). *Limit Theorems for Stochastic Processes*, Springer-Verlag, Berlin.
- [44] Jiang, J. and Knight, L. (2002). Estimation of continuous-time processes via the empirical characteristic function, *J. Bus. Econom. Statist.*, **20**, no. 2, 198–212.
- [45] Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the loglikelihood ratio is mixed normal, *Sankhya Ser. A*, **44**, 173–212.
- [46] Jeganathan, P. (1995). Some aspects of asymptotic theory with applications to time series models, *Econometric Theory*, **11**, no. 5, 818–887.
- [47] Kallenberg, O. (1997). *Foundations of Modern Probability*, Springer-Verlag, New York.
- [48] Kasonga, R. A. (1988). The consistency of a non-linear least squares estimator from diffusion processes, *Stoch. Proc. Appl.*, **24**, 211–229.
- [49] Kessler, M. (1995). Estimation des paramètres d'une diffusion par des contrastes corrigés, *C. R. Acad. Sci. Paris Ser. I*, **320**, 359–362.
- [50] Kessler, M. (1997). Estimation of Diffusion Processes From Discrete Observations, *Scand. J. Statist.*, **24**, 211–229.
- [51] Kreiss, J. P. (1987). On adaptive estimation in stationary ARMA processes, *Ann. Statist.*, **15**, no. 1, 112–133.
- [52] Küchler, U. and Sørensen, M. (1997). *Exponential families of stochastic processes*, Springer-Verlag, New York.

- [53] Kunitomo, N. and Takahashi, A. (2004). Applications of the asymptotic expansion approach based on Malliavin-Watanabe calculus in financial problems, *Stochastic processes and applications to mathematical finance*, World Sci. Publ., 195–232.
- [54] Kutoyants, Y. (1984). *Parameter estimation for stochastic processes*, Heldermann Verlag, Berlin.
- [55] Kwon, Y. (1997). Stability in distribution for a class of diffusion with jump, *Bull. Korean Math. Soc.*, **34**, no. 2, 287–293.
- [56] Kwon, Y. and Lee, C. (1999). Strong Feller Property and Irreducibility of diffusions with jumps, *Stoch. Stoch. Rep.*, **67**, 147–157.
- [57] Le Cam, L. (1953). On asymptotic properties of maximum likelihood estimates and related Bayes estimates, *Univ. of Calif. Publ. Statist.*, **1**, 277–330.
- [58] Le Cam, L. (1960). Locally asymptotically normal families of distributions, *Univ. of Calif. Publ. Statist.*, **3**, 27–98.
- [59] Le Cam, L. (1986). *Asymptotic Methods on Statistical Decision Theory*, Springer-Verlag, New York.
- [60] Liptser, R. S. and Shiriyayev, A. N. (2000). *Statistics of random processes I. 2nd ed.*, Springer-Verlag.
- [61] Liptser, R. S. and Shiriyayev, A. N. (2000). *Statistics of random processes II. 2nd ed.*, Springer-Verlag.
- [62] Luschgy, H. (1992). Local asymptotic mixed normality for semimartingale experiments, *Probab. Theory Related Fields.*, no. 2, 151–176.
- [63] Luschgy, H. (1994). Asymptotic inference for semimartingale models with singular parameter points, *J. Statist. Plann. Infer.*, no. 2, 155–186.
- [64] Mancini, C. (2004). Estimation of the Characteristics of the jumps of a General Poisson-Diffusion model, *Scand. Actuarial J.*, no. 1, 42–52.
- [65] Masry, E. (1983). Probability Density Estimation from Sampled Data, *IEEE Trans. Inform. Theory* **29**, 696–709.

- [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, *Stoch. Proc. Appl.*, **117**, 35–56.
- [68] Masuda, H. and Yoshida, N. (2005). Asymptotic expansion for Barndorff-Nielsen and Shephard's stochastic volatility model, *Stochastic Proc. Appl.*, **115**, no. 7, 1167–1186.
- [69] McLeish, D. L. (1974). Dependent central limit theorems and invariance principles, *Ann. Prob.*, **3**, 829–839.
- [70] Merton, R. (1973). Theory of Rational Option Pricing, *Bell Journal of Econom. Manage. Sci.*, **4**, 141–183.
- [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. and Tweedie, R. L. (1993). Stability of Markov processes II: Foster-Lyapunov criteria for continuous-time processes, *Adv. in Appl. Probab.*, **25**, 487–517.
- [73] Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markov processes III: Foster-Lyapunov criteria for continuous-time processes, *Adv. in Appl. Probab.*, **25**, 518–548.
- [74] Mulinacci, S. (1996). An approximation of American option prices in a jump-diffusion model, *Stoch. Proc. Appl.*, **62**, 1–17.
- [75] Mtundu, N. D. and Koch, R. W. (1987). A stochastic differential equation approach to soil moisture, *Stochastic Hydrol. Hydraul.*, **1**, 101–116.
- [76] Prakasa-Rao, B. L. S. (1972). Maximum likelihood estimation for Markov processes, *Ann. Inst. Statist. Math.*, **24**, 333–345.
- [77] Prakasa-Rao, B. L. S. (1983). Asymptotic theory for non-linear least square estimator for diffusion processes, *Math. Operationsforsch. Statist. Ser. Statist.*, **14**, 195–209.

- [78] Prakasa Rao, B.L.S (1988). Statistical inference from sampled data for stochastic processes, *Statistical inference from stochastic processes (Ithaca, NY, 1987) Contemp. Math., 80, Amer. Math. Soc., Providence, RI.* 249–284.
- [79] Prakasa Rao, B.L.S (1990). Nonparametric density estimation for stochastic processes from sampled data, *Publ. Inst. Statist. Univ. Paris*, **35**, no. 3, 51–83.
- [80] Prakasa Rao, B.L.S. (1999). *Semimartingales and their statistical inference*, Monographs on Statistics and Applied Probability, 83, Chapman & Hall/CRC.
- [81] Prakasa Rao, B.L.S. (2000). *Statistical Inference for Diffusion Type Processes*, Kendall's Library of Statistics 8, ARNOLD.
- [82] Protter, P. E. (2003). *Stochastic Integration and Differential Equations*, 2nd. Springer-Verlag.
- [83] Roussas, G. G. (1965). Extension to Markov processes of a result by A. Wald about the consistency of the maximum likelihood estimate, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **4**, 69–73.
- [84] Roussas, G. G. (1968). Asymptotic normality of the maximum likelihood estimate in Markov processes, *Metrika*, **14**, 62–70.
- [85] Roussas, G. G. (1969). Nonparametric estimation in Markov processes, *Ann. Inst. Statist. Math.*, **21**, 73–87.
- [86] Roussas, G. G. (1972). *Contiguity of probability measures*, Cambridge University Press, London-New York.
- [87] Sakamoto, Y. and Yoshida, N (1996). Expansion of perturbed random variables based on generalized Wiener functionals, *J. Multivariate Anal.*, **59**, no. 1, 34–59.
- [88] Sakamoto, Y. and Yoshida, N (1998). Asymptotic expansion of M -estimator over Wiener space., *Stat. Infer. Stoch. Proc.*, **1**, no. 1, 85–103.
- [89] 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.
- [90] Shimizu, Y. (2004). Nonparametric estimation of diffusion processes with jumps by method of nearest neighbors, Preprint, *Statistical analysis by nonparametric/semiparametric methods and their applications*, University of Kyoto (in Japanese).

- [91] Shimizu, Y. (2005). Model selection for Lévy characteristics in discretely observed jump-diffusion models from the aspect of information criteria, Preprint, *Symposium for Asymptotic Statistics*, University of Tokyo, *submitted*.
- [92] Shimizu, Y. (2006). M -estimation for discretely observed ergodic diffusion processes with infinitely many jumps, *Statist. Infer. Stochastic Proc.*, **9**, no. 2, 179–225.
- [93] Shimizu, Y. (2006). Density estimation of Lévy measure for discretely observed diffusion processes with jumps, *J. Japan Statist. Soc.*, **36**, no. 1, 37–62.
- [94] Shimizu, Y. (2007). The practical inference for discretely observed jump-diffusions from finite samples, Preprint, *Asymptotic Methods in Probability and Statistics*, University of Tokyo, *to appear in J. Japan Statist. Soc.*
- [95] Shimizu, Y. (2007). Functional estimation for Levy measures of semimartingales with poissonian jumps, Research Report Series 07-013, Department of Mathematical Science, Osaka University, *submitted*.
- [96] Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations, *Statist. Infer. Stochastic Proc.*, **9**, no. 3, 227–277.
- [97] Shiriyayev, A. N. (1984). *Probability*, Springer-Verlag.
- [98] Sørensen, M. (1989). A note on the existence of a consistent maximum likelihood estimator for diffusions with jumps, *Markov Process and Control Theory*, Akademie-Verlag, Berlin, 229–234.
- [99] Sørensen, M. (1991). Likelihood methods for diffusions with jumps, *Statist. Infer. Stochastic Proc.*, Marcel Dekker, New York, 67–105.
- [100] Sørensen, M. and Uchida, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations, *Bernoulli*, **9**, no. 6, 1051–1069.
- [101] Taniguchi, M. and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series*, Springer, New York.
- [102] Taraskin, A. F. (1984). Behavior of the likelihood ratio of semimartingales (Russian), *Teor. Veroyatnost. i Primenen.*, **29**, no. 3, 440–451.

- [103] Takahashi, A. and Yoshida, N. (2004). An asymptotic expansion scheme for optimal investment problems, *Stat. Infer. Stoch. Proc.*, **2**, no. 1, 153–188.
- [104] Tjøstheim, D. (1986). Estimation in nonlinear time series models, *Statist. Infer. Stochastic Proc.*, Marcel Dekker, New York, 67–105.
- [105] Uchida, M. (2003). Estimation for dynamical systems with small noise from discrete observations., *J. Japan Statist. Soc.*, **33**, no. 2, 157–167.
- [106] Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions, *Scand. J. Statist.*, **31**, no. 4, 553–566.
- [107] Uchida, M. and Yoshida, N. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe, *Stat. Infer. Stoch. Proc.*, **7**, no. 1, 35–67.
- [108] Wald, A. (1939). Contributions to the theory of statistical estimation and testing hypothesis, *Ann. Math. Statist.*, **10**, 299–326.
- [109] Wald, A. (1948). Asymptotic properties of the maximum likelihood estimate of an unknown parameter of a discrete stochastic processes, *Ann. Math. Statist.*, **19**, 40–46.
- [110] Wald, A. (1949). Note on the consistency of the maximum likelihood estimate, *Ann. Math. Statist.*, **20**, 595–601.
- [111] Wald, A. (1950). *Statistical decision function*, Wiley, New York.
- [112] Whittle, P. (1962). Gaussian estimation in stationary time series, *Bull. Inst. Internat. Statist.*, **39**, 105–129.
- [113] Woerner, J. (2001). Statistical Analysis for Discretely Observed Lévy Processes, *Ph. D. dissertation*, Albert Ludwigs University of Freiburg .
- [114] van der Vaart, A. W. (1998). *Asymptotic statistics*, Cambridge Series in Statistical and Probabilistic Mathematics, 3. Cambridge University Press.
- [115] Yoshida, N. (1988). On asymptotic mixed normality of the maximum likelihood estimator in a multidimensional diffusion process, *Statistical theory and data analysis, II* (Tokyo, 1986), North-Holland, Amsterdam, 559–566.

- [116] Yoshida, N. (1990). Asymptotic behavior of M -estimator and related random field for diffusion process, *Ann. Inst. Statist. Math.*, **42**, no. 2, 221–251.
- [117] Yoshida, N. (1992). Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe, *Probab. Theory Related Fields*, **92**, no. 3, 275–311.
- [118] Yoshida, N. (1992). Estimation for diffusion processes from discrete observations, *J. Multivar. Anal.*, **41**, 220–242.
- [119] Yoshida, N. (1992). Asymptotic expansion for statistics related to small diffusions, *J. Japan Statist. Soc.*, **22**, no. 2, 139–159.
- [120] Yoshida, N. (1993). Malliavin calculus and higher order statistical inference, *Statistical sciences and data analysis* (Tokyo, 1991), **22**, 175–180.