When is it no longer possible to estimate a compound Poisson process?

Céline Duval

When is it no longer possible to estimate a compound Poisson process?

Céline Duval*

Abstract

We consider centered compound Poisson processes with finite variance, discretely observed over $[0, T]$ and let the sampling rate $\Delta = \Delta_T \to \infty$ as $T \to \infty$. From the central limit theorem, the law of each increment converges to a Gaussian variable. Then, it should not be possible to estimate more than one parameter at the limit. First, from the study of a parametric example we identify two regimes for Δ_T and we observe how the Fisher information degenerates. Then, we generalize these results to the class of compound Poisson processes. We establish a lower bound showing that consistent estimation is impossible when Δ_T grows faster than $\sqrt{T}$. We also prove an asymptotic equivalence result, from which we identify, for instance, regimes where the increments cannot be distinguished from Gaussian variables.

AMS subject classifications: 62B15, 62K99, 62M99.
Keywords: Discretely observed random process, Compound Poisson process, Information loss.

1 Introduction

1.1 Motivation and statistical setting

Continuous diffusive models are often used for phenomena observed at large sampling rate, even though they present discontinuities or jumps at lower frequencies. For example in finance, asset prices or volumes change at discrete random times (see for instance Gerber and Shiu [7], Russell and Engle [18] or Guilbaud and Pham [8]), however it is common to use continuous diffusive processes to model them when the sampling rate is large (see *e.g.* Masoliver *et al.* [13], Önalan [16] or Hong and Satchell [9]). This opposition in the observations' behavior between small frequencies and large sampling rate is evoked in Cont and de Larrard [4]: "over time scales much larger than the interval between individual order book events, prices are observed to have diffusive dynamics and modeled as such." In physics the opposition between large scale diffusive behavior and point process at small scale is also popular (see *e.g.* Metzler and Klafter [14] or Uchaikin and Zolotarev [23]). The usual justification for using diffusive approximations

*MAP5, UMR CNRS 8145, Université Paris Descartes, Sorbonne Paris Cité

is as follows. Suppose we have discrete observations of a centered pure jump process X observed at a sampling rate $\Delta > 0$, *e.g.* a centered compound Poisson process with finite variance, namely we observe

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

To lighten notation we set $n_T := \lfloor T\Delta^{-1}\rfloor$, the number of observations. We make explicit the dependence in T since it is the asymptotic of the paper. If Δ is large, between two observations of X many jumps occurred, the central limit theorem gives for every increments the approximation

$$X_{i\Delta} - X_{(i-1)\Delta} \approx \sigma\big(W_{i\Delta} - W_{(i-1)\Delta}\big)$$

where W is a standard Wiener process and σ is positive. Hence, only the variance parameter σ^2 should be identifiable from (1). If X depends on more parameters their identifiability should be lost. Yet the use of diffusive approximations conceals the jump's dynamic observed at lower frequencies. The following questions naturally come across.

i) Is it possible to estimate the parameters characterizing X from (1)?

ii) Is the experiment generated by (1) asymptotically equivalent to a Gaussian experiment when $\Delta = \Delta_T \to \infty$ as $T \to \infty$?

The asymptotic equivalence of a Poisson experiment with variable intensity has been studied in Brown *et al.* [3]. Shevtsova [19] looks at the accuracy of Gaussian approximations for Poisson random sums.

Definition 1. *Observations (1) are said to be on a macroscopic regime if $\Delta = \Delta_T \to \infty$ and $T/\Delta_T \to \infty$ as $T \to \infty$.*

The condition $T/\Delta_T \to \infty$ ensures there are asymptotically infinitely many observations ($\lfloor T\Delta^{-1}\rfloor \to \infty$ as $T \to \infty$). A typical example of macroscopic regime is a sampling rate Δ_T of the order of T^α as $T \to \infty$ for α in $(0,1)$ as $T \to \infty$. In this paper we restrain our study to homogeneous compound Poisson processes. A compound Poisson process X is defined as

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

where R is a Poisson process of intensity λ and (ξ_i) are independent and identically distributed random variables independent of R. The process X is characterized by the the pair $r = (\lambda, f)$, where f is the probability law of ξ_1. We denote by $\mathcal{P}$ the class of compound Poisson processes.

1.2 Main results

Investigating questions **i)** and **ii)** directly is difficult. Hence in Section 2 we first build and study a toy model: a compound Poisson process plus a drift that depends on a 2-dimensional parameter. This process does not belong to $\mathcal{P}$. From this toy model, we identify two distinct macroscopic regimes,

- A regime where Δ goes to infinity faster than $\sqrt{T}$, where the parameters cannot be consistently estimated from (1), providing a negative answer to **i)** (see Theorem 1 hereafter).
- A regime where Δ goes to infinity slower than $\sqrt{T}$, where the parameters can be estimated answering positively to **i)**. However, optimal rates are much slower than usual parametric ones (see Proposition 1 hereafter).

From the study of the toy model, we derive a lower bound in Theorem 2. It identifies regimes in which consistent estimation of the law generating a process in $\mathcal{P}$ is impossible, leading to a negative answer to **i)**. Theorem 3 gives an asymptotic equivalence result; according to the behavior of Δ_T with regard to T, the following occurs.

- The experiment generated by the observation of a process in $\mathcal{P}$ is asymptotically equivalent to a Gaussian experiment, answering positively to **ii)**.
- Compound Poisson processes depending on a large number of parameters are not identifiable, providing a negative answer to **i)**. The limit number of parameters beyond which consistent estimation is not possible is made explicit.

This paper is organized as follows, in Section 2 we construct and study our toy model. In Section 3 we establish the main Theorems 2 and 3. A discussion is proposed in Section 4. Finally, Section 5 is devoted to the proofs.

2 Information loss: A parametric example

2.1 Building up a parametric model

Consider the Lévy process Y defined by

$$Y_t = X_t - \frac{\lambda t}{\beta} = \sum_{i=1}^{R_t} \xi_i - \frac{\lambda t}{\beta}, \qquad t \geq 0, \tag{2}$$

where R is Poisson process of intensity $\lambda \in (0, \infty)$ independent of $(\xi_i)_{i\geq 0}$ which are independent and exponentially distributed random variables with parameter $\beta \in (0, \infty)$. Due to the drift part, Y does not belong to $\mathcal{P}$ (unlike X). This model, known as the Cramér-Lundberg model, is used by insurance companies to model big claims of subscribers (see *e.g.* Embrechts *et al.* [6] or Miksoch [15]). Without the drift part, it is also used in Alexandersson [1] to model rainfall.

Suppose we observe $\lfloor T\Delta^{-1}\rfloor$ increments of Y, conditional on the event $\{R_{i\Delta} - R_{(i-1)\Delta} \neq 0\}$. Namely we observe Y over $[0, S(T)\Delta]$ at a sampling rate $\Delta > 0$, where $S(T)$ is random and such that

$$\sum_{i=1}^{S(T)} \mathbf{1}_{\{R_{i\Delta}-R_{(i-1)\Delta}\neq 0\}} = \lfloor T\Delta^{-1}\rfloor.$$

Remark 1. *The following results on $S(T)$ can be easily checked. Since the probability of occurrence of a zero increment of X is $e^{-\lambda\Delta}$, we have*

$$\mathbb{P}(S(T) \neq \lfloor T\Delta^{-1}\rfloor) = 1 - (1-e^{-\lambda\Delta})^{\lfloor T\Delta^{-1}\rfloor} \sim \lfloor T\Delta^{-1}\rfloor e^{-\lambda\Delta} \to 0 \quad \text{as } T \to \infty$$

if Δ goes to infinity as a power of T. Moreover, $S(T)$ is negative binomial with parameters $(1-e^{-\lambda\Delta}, \lfloor T\Delta^{-1}\rfloor)$.

Define $J = \{i \in \{1,\dots,S(T)\}, R_{i\Delta} - R_{(i-1)\Delta} \neq 0\}$, by construction $|J| = \lfloor T\Delta^{-1}\rfloor$. Consider the $\lfloor T\Delta^{-1}\rfloor$ independent and identically distributed observations

$$\widetilde{\boldsymbol{Y}} = \big(\widetilde{Y}_{i\Delta_T} - \widetilde{Y}_{(i-1)\Delta_T} = Y_{i_\Delta} - Y_{(i-1)\Delta} | R_{i\Delta} - R_{(i-1)\Delta} \neq 0,\ i \in J\big). \tag{3}$$

We introduce the family of experiments indexed by Δ generated by the conditional observations (3)

$$\widetilde{\mathcal{Y}}^\Delta := \big(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}, \mathcal{P}(\mathbb{R}^{\lfloor T\Delta^{-1}\rfloor}), \{\widetilde{\mathbb{P}}_\theta^{T,\Delta}, \theta \in \Theta\}\big),$$

where θ denotes the unknown parameter $\theta = (\lambda,\beta) \in \Theta = (0,\infty)\times(0,\infty)$ and $\widetilde{\mathbb{P}}_\theta^{T,\Delta}$ the law of $\widetilde{\boldsymbol{Y}}$.

Remark 2. *The natural experiment to work with is the experiment $\mathcal{Y}^\Delta$ generated by the observations of $\lfloor T\Delta^{-1}\rfloor$ increments of Y*

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

But the law of $\boldsymbol{Y}$ is not dominated and the Fisher information in $\mathcal{Y}^\Delta$ does not exist. Indeed the distribution of Y_Δ can be decomposed in

$$\mathbb{P}(R_\Delta = 0)\delta_{\{-\frac{\lambda\Delta}{\beta}\}}(\cdot) + \mathbb{P}(R_\Delta > 0)\widetilde{p}_{\Delta,\theta}(\cdot),$$

where $\widetilde{p}_{\Delta,\theta}$ is dominated by the Lebesgue measure but $\delta\{-\frac{\lambda\Delta}{\beta}\}$, the mass concentrated at $-\frac{\lambda\Delta}{\beta}$, cannot be dominated over Θ. Removing null increments of the Poisson part gives the experiment $\widetilde{\mathcal{Y}}^\Delta$, dominated by the Lebesgue measure, where the Fisher information exists. Since the probability of a null increments of X is $e^{-\lambda\Delta}$, which is negligible as $\Delta \to \infty$, we show in Section 2.4 that the results established for $\widetilde{\boldsymbol{Y}}$ hold also for $\boldsymbol{Y}$: the experiments $\widetilde{\mathcal{Y}}^\Delta$ and $\mathcal{Y}^\Delta$ are asymptotically equivalent.

The intuition of the problem is the following, as ξ_1 has finite variance, the central limit theorem applies for each increments and gives for i in J

$$\frac{\widetilde{Y}_{i\Delta_T} - \widetilde{Y}_{(i-1)\Delta_T}}{\sqrt{\Delta_T}} \xrightarrow{d} \mathcal{N}\Big(0, \frac{2\lambda}{\beta^2}\Big), \qquad \text{as } T \to \infty.$$

Thus each observation converges in law to a Gaussian random variable depending on one parameter: the parameter θ should no longer be identifiable when Δ gets large, only the ratio λ/β^2 should be.

2.2 Study of the Fisher information

The increments of Y are independent and identically distributed, it follows that the Fisher information of $\widetilde{\mathcal{Y}}^\Delta$ satisfies

$$I_{\lfloor T\Delta^{-1}\rfloor,\Delta}(\theta) = \lfloor T\Delta^{-1}\rfloor I_{1,\Delta}(\theta)$$

where $I_{1,\Delta}(\theta)$ is the Fisher information corresponding to one increment. It has no closed form expression but the following Proposition gives its asymptotic behavior.

Proposition 1. *Let $\Delta = \Delta_T$ such that $\Delta_T \to \infty$ and $T/\Delta_T \to \infty$ as $T \to \infty$. Then*

$$\lim_{T\to\infty} I_{1,\Delta_T}(\theta) = I(\theta) := \begin{pmatrix} \frac{1}{2\lambda^2} & -\frac{1}{\lambda\beta} \\ -\frac{1}{\lambda\beta} & \frac{2}{\beta^2} \end{pmatrix}$$

and the eigenvalues of $I_{\lfloor T\Delta_T^{-1}\rfloor,\Delta_T}(\theta)$, denoted $e_{1,\Delta_T}(\theta)$ and $e_{2,\Delta_T}(\theta)$, satisfy

$$\begin{aligned} e_{1,\Delta_T}(\theta) &= \big(\tfrac{2}{\beta^2} + \tfrac{1}{2\lambda^2}\big)\lfloor T\Delta_T^{-1}\rfloor + \tfrac{3(7\beta^4+40\beta^2\lambda^2+56\lambda^4)}{8\beta^2\lambda^3(\beta^2+4\lambda^2)} \frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta} + O\Big(\tfrac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T^{3/2}}\Big) \\ e_{2,\Delta_T}(\theta) &= \tfrac{3}{4\beta^2\lambda+16\lambda^3} \frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta} + O\Big(\tfrac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T^{3/2}}\Big). \end{aligned}$$

Remark 3. *The matrix $I(\theta)$ is the Fisher information of an experiment consisting in one variable of distribution $\mathcal{N}(0, 2\lambda/\beta^2)$.*

From Proposition 1, whenever Δ_T goes to infinity faster than $\sqrt{T}$ the Fisher information degenerates to a rank 1 matrix: the second eigenvalue e_{2,Δ_T} goes to 0 as $\lfloor T\Delta_T^{-1}\rfloor/\Delta_T \sim T/\Delta_T^2$. Theorem 1 below shows that it is indeed not possible to build a consistent estimator of θ in those scales. Conversely, when Δ_T is slower than $\sqrt{T}$, both eigenvalues of the Fisher information go to infinity. Since the experiment $\widetilde{\mathcal{Y}}^\Delta$ is regular we deduce that the parameter θ remains identifiable and that consistent estimators of θ do exist. This is surprising, even if each observation is close to a Gaussian variable depending on one parameter, the whole sample still permits to estimate consistently all unknown parameters. However the optimal rate of convergence, determined by the slowest eigenvalue $e_{2,\Delta_T}(\theta)$, is in $\big(\lfloor T\Delta_T^{-1}\rfloor/\Delta_T\big)^{1/2}$. It is much slower than usual parametric rates in $\lfloor T\Delta_T^{-1}\rfloor^{1/2}$, the square root of the sample size.

2.3 A lower bound

In what follows $\|\cdot\|$ denotes a norm on $\mathbb{R}^2$. Define the diameter of a set A as

$$diam(A) = \sup_{a_1,a_2\in A} \|a_2 - a_1\|.$$

Theorem 1. *Let Δ_T be such that $\Delta_T \to \infty$ and $T\Delta_T^{-2} \to \mathfrak{l} \in [0,\infty)$ as $T \to \infty$. Then, for all $\theta_0 \in \Theta$ and $\delta > 0$ there exists $\mathcal{V}_\delta(\theta_0) \subset \Theta$ a neighborhood of θ_0 such that $diam(\mathcal{V}_\delta(\theta_0)) \leq \delta$ and*

$$\varliminf_{T\to\infty} \inf_{\widehat{\theta}} \sup_{\theta\in\mathcal{V}_\delta(\theta_0)} \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta} - \theta\|\big] > 0$$

where the infimum is taken over all estimators.

From Theorem 1, there is no consistent estimator of θ when Δ_T grows rapidly to infinity, faster than $\sqrt{T}$. This was expected as the Fisher information degenerates to a rank 1 matrix in those regimes (see Proposition 1). Notice that if Theorem 1 holds for every $\delta > 0$, possibly small, it is not uniform in δ. It is not possible to apply it along a vanishing sequence of δ.

2.4 Generalization to the unconditional experiment

The asymptotic equivalence of $\widetilde{\mathcal{Y}}^\Delta$ and $\mathcal{Y}^\Delta$ (defined in Remark 2) is an immediate consequence of the following Lemma.

Lemma 1. *Define the probability measures,*

$$\begin{aligned} p_n(\theta,x) &= f_n(\theta,x)dx \\ q_n(\theta,x) &= a_n(\theta)w_{n,\theta}(dx) + (1-a_n(\theta))f_n(\theta,x)dx, \end{aligned}$$

where $\theta \in \Sigma$, where Σ is a compact subset of $\mathbb{R}^d$, $d \geq 1$, $a_n(\theta) \in (0,1)$, $f_n(\theta,\cdot)$ is a density absolutely continuous with respect to the Lebesgue measure and $w_{n,\theta}$ is a probability measure. Consider the statistical experiments $\mathcal{E}^n$ and $\mathcal{G}^n$ generated by the independent and identically distributed observation of n random variables of density $p_n(\theta,\cdot)$ and $q_n(\theta,\cdot)$ respectively. If $\sup_{\theta\in\Sigma} a_n(\theta) = o(\frac{1}{n})$, then $\mathcal{E}^n$ and $\mathcal{G}^n$ are asymptotically equivalent.

Proof of Lemma 1 is given in Section 5. Since Y is a Lévy process, observations $\boldsymbol{Y}$, and $\widetilde{\boldsymbol{Y}}$, are independent and identically distributed. The distribution of Y_Δ is

$$p_{\Delta,\theta}(x) = e^{-\lambda\Delta}\delta_{\{-\frac{\lambda\Delta}{\beta}\}}(dx) + (1-e^{-\lambda\Delta})\widetilde{p}_{\Delta,\theta}(x)dx, \qquad x \in \mathbb{R} \tag{4}$$

where $\delta_{\{-\lambda\Delta/\beta\}}$ is the measure concentrated at $-\frac{\lambda\Delta}{\beta}$ and $\widetilde{p}_{\Delta,\theta}$ is the density of $\widetilde{Y}_\Delta$ absolutely continuous with respect to the Lebesgue. We consider macroscopic regimes such that $\Delta = 0(T^\alpha)$ for some $\alpha \in (0,1)$, it follows that $e^{-\lambda\Delta_T} = o(\lfloor T\Delta_T^{-1}\rfloor^{-1}) = o(\lfloor T\Delta_T^{-1}\rfloor^{-1})$ and Lemma 1 applies with $a_{\lfloor T\Delta_T^{-1}\rfloor}(\theta) = e^{-\lambda\Delta_T}$ and $w_{T,\theta}(dx) = \delta_{\{-\lambda\Delta_T/\beta\}}(dx)$. The experiments $\mathcal{Y}^\Delta$ and $\widetilde{\mathcal{Y}}^\Delta$ are asymptotically equivalent and the results established for $\widetilde{\mathcal{Y}}^\Delta$ hold for $\mathcal{Y}^\Delta$.

3 Identifiability loss for compound Poisson processes

3.1 A lower bound

In Section 2 we exhibit on a parametric example a regime where estimation is impossible. We generalize here Theorem 1 to the class of compound Poisson processes $\mathcal{P}$ whose norm

$$\|r\|_{2,\mathcal{P}} = \|(\lambda, f)\|_{2,\mathcal{P}} := \|\lambda f\|_2,$$

is finite, $\|\cdot\|_2$ stands for the usual L_2 norm.

Theorem 2. *Let $\Delta_T \to \infty$ be such that $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ and $T/\Delta_T \to \infty$ as $T \to \infty$. Then, for all $r_0 \in \mathcal{P}$, $\|r_0\|_{2,\mathcal{P}}$, and $\delta > 0$, there exists $\mathcal{V}_\delta(r_0)$, a neighborhood of r_0 such that $diam(\mathcal{V}_\delta(r_0)) \leq \delta$ and*

$$\varliminf_{T\to\infty} \inf_{\widehat{r}} \sup_{r\in\mathcal{V}_\delta(r_0)} \mathbb{E}_r^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r\|_{2,\mathcal{P}}\big] > 0$$

where the infimum is taken over all estimators.

It follows that if Δ_T is of the order of T^α for $\alpha \in (1/2, 1)$ it is not possible to build a consistent estimator of (λ, f) from (1) when f is unknown.

Remark 4. *A compound Poisson process is a renewal reward process and a Lévy process. Thus, we immediately derive from Theorem 2 that if Δ_T is such that $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ as $T \to \infty$, it is not possible to build consistent estimators of the law generating a renewal reward process or a Lévy process with jumps from (1).*

Remark 5. *The rate restriction $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ is technical and might be weakened in $T/\Delta_T^2 = O(1)$. Indeed we derive Theorem 2 from Theorem 1, which holds under the restriction $T/\Delta_T^2 = O(1)$. To apply Theorem 1 in the present setting, we show that the experiment $\mathcal{Y}^{\Delta_T}$ introduced in Section 2 is asymptotically equivalent to an experiment generated by increments of a compound Poisson process. This asymptotic equivalence result imposes the constraint $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$.*

3.2 An asymptotic equivalence result

3.2.1 Building up asymptotically equivalent experiments

A parameter transformation function. Consider some density f_θ with respect to the Lebesgue measure, centered with finite K first moments, $K \in \mathbb{N}$. Let

$$\int x f_\theta(x)dx = 0 \quad \text{and} \quad \int x^k f_\theta(x)dx = m_k, \qquad k = 2, \dots, K.$$

Define the parameter $\theta = (\lambda, m_2, \dots, m_K) \in \Sigma_K$, where Σ_K is a compact subset of

$$\begin{array}{ll} \mathbb{R}_+ \times [\mathbb{R}_+ \times \mathbb{R}] \times \cdots \times [\mathbb{R}_+ \times \mathbb{R}] \times \mathbb{R}_+ & \text{if } K \text{ is even} \\ \mathbb{R}_+ \times [\mathbb{R}_+ \times \mathbb{R}] \times \cdots \times [\mathbb{R}_+ \times \mathbb{R}] & \text{if } K \text{ is odd,} \end{array}$$

where $\mathbb{R}_+$ denotes $(0,\infty)$. Let $\gamma>0$, consider the parameter transformation function

$$h_\gamma : \theta \in \Sigma_K \to h_\gamma(\theta) = \big(\gamma\lambda, \tfrac{m_2}{\gamma}, \dots, \tfrac{m_K}{\gamma}\big).$$

Fix $\gamma \in (0,\infty)\setminus\{0,1\}$ such that $h_\gamma(\theta)\in\Sigma_K$. In the sequel we consider X and Z two compound Poisson processes, X has intensity λ and compound density f_θ and Z has intensity $\gamma\lambda$ and compound density $f_{h_\gamma(\theta)}$. Namely $f_{h_\gamma(\theta)}$ is a density with respect to the Lebesgue measure such that

$$\int x f_{h_\gamma(\theta)}(x)dx = 0 \quad \text{and} \quad \int x^k f_{h_\gamma(\theta)}(x)dx = \frac{m_k}{\gamma}, \qquad k=2,\dots,K. \tag{5}$$

To establish Theorem 3 below, we do not need to know $f_{h_\gamma(\theta)}$, only to make sure of its existence and that it is in the Sobolev space $W^{1,1}$ (*i.e.* $f_{h_\gamma(\theta)} \in C^1$ with $f'_{h_\gamma(\theta)} \in L^1$, see the class of densities (9) below). The existence of $f_{h_\gamma(\theta)}$ is an immediate consequence of the truncated Hamburger moment problem. A necessary and sufficient condition for (5) to have a solution is that the associated Hankel matrices are positive definite (see *e.g.* Athanassoulis and Gavriliadis [2] or Tagliani [21]). Since $(0, m_2, \dots, m_K)$ are the first moments of the density f_θ, the associated Hankel matrices are positive definite. Then, the Hankel matrices of $(0, \frac{m_2}{\gamma}, \dots, \frac{m_K}{\gamma})$, for any $\gamma > 0$, are positive definite as well and existence of a density $f_{h_\gamma(\theta)}$, absolutely continuous with respect to the Lebesgue measure, is thus ensured. Note that the number of solutions is infinite. A lot of papers study methods to build explicit solutions of (5); for instance maximum entropy approaches (see *e.g.* Tagliani [21] or Sobezyk and Trębicki [20]), polynomial solutions (see *e.g.* Rodriguez and Seatzu [17]) or solutions based on kernel density functions (see *e.g.* Athanassoulis and Gavriliadis [2]). All these solutions are C^1 with integrable derivative, which ensures that $f_{h_\gamma(\theta)}$ can be chosen in $W^{1,1}$. In practice if K is large, building $f_{h_\gamma(\theta)}$ is difficult and the shape of the solution highly depends on the construction method considered. Above references provide explicit examples.

Definition of the experiments. Consider also a Gaussian process W with quadratic variation λm_2. We associate the parameter $\phi = (\lambda, m_2)$ in Σ_2. Suppose X, Z and W are discretely observed at a sampling rate $\Delta > 0$ over $[0,T]$, namely

$$\big(X_{i\Delta} - X_{(i-1)\Delta},\ i = 1,\dots,\lfloor T\Delta^{-1}\rfloor\big), \tag{6}$$
$$\big(Z_{i\Delta} - Z_{(i-1)\Delta},\ \ i = 1,\dots,\lfloor T\Delta^{-1}\rfloor\big), \tag{7}$$
$$\big(W_{i\Delta} - W_{(i-1)\Delta}, i = 1,\dots,\lfloor T\Delta^{-1}\rfloor\big). \tag{8}$$

Define the families of statistical experiments indexed by Δ

$$\mathcal{X}^\Delta := \{\mathbb{P}_\theta^{T,\Delta}, \theta\in\Sigma_K\},\ \mathcal{Z}^\Delta := \{\mathbb{Q}_\theta^{T,\Delta}, \theta\in\Sigma_K\} \text{ and } \mathcal{W}^\Delta := \{\mathbb{D}_\phi^{T,\Delta}, \phi\in\Sigma_2\},$$

where $\mathbb{P}_\theta^{T,\Delta}$ denotes the law of (6), $\mathbb{Q}_\theta^{T,\Delta}$ the law of (7) and $\mathbb{D}_\phi^{T,\Delta}$ the law of (8).

3.2.2 Statement of the result

Define the subclass of densities

$$\mathcal{F} = \Big\{ f \in \mathcal{F}(\mathbb{R}),\ |\xi f^*(\xi)| \underset{|\xi|\to\infty}{=} O(1) \Big\}, \tag{9}$$

where $\mathcal{F}(\mathbb{R})$ is the class of densities with respect to the Lebesgue measure and f^* denotes the Fourier transform of f. The class $\mathcal{F}$ contains any density sufficiently regular. For instance all densities in the Sobolev space $W^{1,1}$, *i.e.* densities with integrable derivatives.

Theorem 3. *Let f_θ be in $\mathcal{F}$ and suppose $\sup_{\theta\in\Sigma_K} |\int x^3 f_\theta(x)dx| < \infty$,let $\Delta_T \to \infty$ be such that $T/\Delta_T \to \infty$ as $T \to \infty$.*

1. *Let $K \geq 2$, if $T\Delta_T^{-(K+1)/2} = o((\log(T/\Delta_T))^{-1/4})$, the experiments $\mathcal{X}^{\Delta_T}$ and $\mathcal{Z}^{\Delta_T}$ are asymptotically equivalent.*

2. *Moreover, if either one of the following holds*

   i. $T\Delta_T^{-3/2} = o((\log(T/\Delta_T))^{-1/4})$

   ii. $T\Delta_T^{-2} = o((\log(T/\Delta_T))^{-1/4})$ *and* $m_3 = 0$

   the experiment $\mathcal{X}^{\Delta_T}$ is asymptotically equivalent to the Gaussian experiment $\mathcal{W}^{\Delta_T}$.

The assumptions of Theorem 3 focus on f_θ only as it is possible to select $f_{h_\gamma(\theta)}$ solution of (5) satisfying the same assumptions (see Section 3.2.1).

3.2.3 Interpretation

Part 2 of Theorem 3 can be easily interpreted. It states that when Δ_T goes rapidly to infinity, the Gaussian approximation is valid. The increments of a compound Poisson process cannot be distinguished from the increments of a Brownian motion. Using a diffusive model even though the phenomena is *per se* discontinuous is justified in those regime (see *e.g.* Cont and de Larrard [4]).

Part 1 of Theorem 3 is more general since it holds regardless of the rate Δ_T. It should be interpreted as follows: for a given rate Δ_T, with regard to T, how many parameters are not identifiable? The response given by the theorem is if Δ_T is of the order of T^α as $T \to \infty$, $\alpha \in (0,1)$, it is not possible to identify more than $K_\alpha = \lceil \frac{2}{\alpha} - 1 \rceil$ moments of the compound law and the intensity. Indeed, it is possible to exhibit two different compound Poisson processes that cannot be distinguished from their discrete observation. Thus, compound laws characterized by their $M \geq K_\alpha$ first moments cannot be estimated consistently from observations (1).

The case $K = 1$, where parameter θ reduces to $\theta = \lambda$, is not covered by Theorem 3. Indeed θ appears in the limit variance and is always identifiable. This case is studied for a particular discrete compound law in Duval and Hoffmann [5], where an efficient estimator of θ is given and the asymptotic equivalence with a Gaussian experiment is established for Δ going rapidly to infinity, namely $T/\Delta_T^{1+1/4} = o((\log(T/\Delta_T))^{-1/4})$. This constraint is more restrictive than the one of Theorem 3 due to the discreteness of the compound law, a regularizing kernel is needed to prove the equivalence and imposes the condition. In the case $K = 2$, the parameter becomes $\theta = (\lambda, m_2)$, a particular example is studied in Section 2. Corroborating Theorem 3, Theorem 2 shows that it is not possible to estimate θ whenever $T\Delta_T^{-2} \to 0$ as $T \to \infty$ since two parameters have to be estimated.

4 Discussion

Consequences and extensions. An immediate consequence of the results of the paper is that nonparametric estimation for compound Poisson processes is impossible when Δ goes to infinity as a power of T, since it requires to estimate an infinite number of parameters (see Theorem 3). In this paper we did not investigate the existence and properties of consistent estimation procedures when they exist. From the example of Section 2, we may expect that such procedures exist but have optimal rates of convergence that deteriorate as the number of parameters increases.

A natural generalization of Theorem 3 would be to relax the constraint on the third moment of the compound law in $|\int_{\mathbb{R}} x^\eta f_\theta(x)dx| < \infty$ for some $\eta > 0$, and more specifically for $\eta \in (0, 2)$. This allows to exhibit at the limit a convergence to any stable process and not only to a Brownian motion (see for instance Kotulski [10] or Levy and Taqqu [12]). The stable limit law is parametric, then if the initial process depends on too many parameters questions **i)** and **ii)** (modifying the limit experiment accordingly) of Section 1.1 may also be extended. However, the methodology used in this paper highly rely on the hypothesis $|\int_{\mathbb{R}} x^3 f_\theta(x)dx| < \infty$ (see the proof of Theorem 3 and the use of Edgeworth expansions). Another generalization might be to add a long range dependence structure between the jump times or the jumps themselves that remains at the macroscopic limit. But our methodology uses heavily the Lévy structure of the process.

On the difficulty of giving identifiability results. Section 3 contains mostly negative results (see Theorems 2 and 3). Establishing positive results in the general case such as "parametric estimation is possible for K parameters if Δ goes to infinity slower than $T^{h(K)}$", for some function h, is much more involved, even without specifying a rate of convergence. When Δ goes to infinity, the law of each observation is asymptotically Gaussian and depends on one parameter, the asymptotic variance. If this variance is insufficient to recover the initial parameter, one has to study the limit experiment to derive identifiability and not just the law of one observation. The successful study of the example of Section 2 entirely relies on the fact that modified Bessel functions of the first kind appears in the density of the increments. Asymptotic expansions of such functions are known rendering possible the study of the limit Fisher information. Thus, the methodology adopted in Section 2 cannot be generalized to other cases.

The multivariate case. The results of the paper should apply also to multivariate compound Poisson models. In that case the limit distribution of the increments is a multivariate Gaussian variable, additional information might be extracted from the covariance structure between coordinates.

A particular model is worth mentioning. Consider the bidimentional compound Poisson model where we observe for each increment the number of events $R_{i\Delta} - R_{(i-1)\Delta}$ and the value of the increment $X_{i\Delta} - X_{(i-1)\Delta}$. One may think this additional information may improve identifiability, it is not the case. Observing R enables to estimate the intensity of the Poisson process with high accuracy; indeed a sufficient statistics is the terminal value $R_{n_T\Delta} = R_{\lfloor T\Delta_T^{-1}\rfloor \Delta_T}$ and the maximum likelihood estimator $R_{n_T\Delta}/T$ converges at the rate $\sqrt{T}$ which is much faster than $\sqrt{n_T}$! Consequently, it does improve identifiability when two parameters are to be estimated: the intensity and one jump parameter that appears in the variance. In that case, the example studied in Section 2 remains identifiable in all macroscopic regimes.

Still, identifiability of the jump probability law remains to be studied. Intuitively, if f depends on parameters that cannot be fully identified from the asymptotic variance there might be a loss of identifiability: the variance of f is identifiable but maybe not its shape. Moreover, if the construction that permits to derive Theorem 3 part 1 becomes obsolete, the results of Theorem 3 part 2 still hold.

5 Proof

5.1 Proof of Proposition 1

Preparation

The increments of Y are independent and identically distributed. Conditional on the presence of jumps, the density of $\widetilde{Y}_\Delta + \lambda\Delta/\beta$ is

$$\mathbf{P}_\Delta[f_\beta](x) = \sum_{m=1}^{\infty} \mathbb{P}(R_\Delta = m | R_\Delta \neq 0) f_\beta^{\star m}(x) = \frac{e^{-\lambda\Delta}}{1-e^{-\lambda\Delta}} \sum_{m=1}^{\infty} \frac{(\lambda\Delta)^m}{m!} f_\beta^{\star m}(x)$$

where f_β is the density of an exponentially distributed random variable with parameter β and $\star$ denotes the convolution product. It follows that $f_\beta^{\star m}$ is the density of a gamma distribution. Then, for $x \geq 0$

$$\mathbf{P}_\Delta[f_\beta](x) = \frac{e^{-\lambda\Delta}}{1-e^{-\lambda\Delta}} e^{-\beta x} \lambda\Delta\beta \sum_{m=0}^{\infty} \frac{(\lambda\Delta\beta x)^m}{m!(m+1)!}.$$

Let $k \in \mathbb{N}$ and introduce the function

$$g_k(x) = \sum_{m=0}^{\infty} \frac{x^m}{m!(m+k)!}, \qquad x \in [0,\infty). \tag{10}$$

It is related to the modified Bessel function of the first kind $\mathcal{I}_k$ as follows

$$g_k(x) = \frac{1}{x^{k/2}}\mathcal{I}_k(2\sqrt{x}), \qquad x > 0, \tag{11}$$

where
$$\mathcal{I}_k(x) = \sum_{m=0}^{\infty}\left(\frac{x}{2}\right)^{2m+k}\frac{1}{m!\Gamma(m+k+1)}.$$

Rewriting $\mathbf{P}_\Delta[f_\beta]$ and adding the drift part we get the density $\widetilde{p}_{\Delta,\theta}$ of $\widetilde{Y}_\Delta$, for $x \geq -\lambda\Delta/\beta$

$$\widetilde{p}_{\Delta,\theta}(x) = \frac{e^{-2\lambda\Delta-\beta x}}{1-e^{-\lambda\Delta}}\lambda\Delta\beta g_1\big(\lambda\Delta\beta x + \lambda^2\Delta^2\big).$$

Technical Lemmas

Lemma 2. *Let $k \in \mathbb{N}$, the modified Bessel function of the first kind $\mathcal{I}_k(x)$ satisfies for all $M \in \mathbb{N}$*

$$e^{-x}\mathcal{I}_k(x) = \frac{1}{(2\pi x)^{1/2}}\sum_{m=0}^{M}\frac{(-1)^m}{(2x)^m}\frac{\Gamma(k+m+\frac{1}{2})}{m!\Gamma(k-m+\frac{1}{2})} + O\Big(\frac{1}{x^{M+3/2}}\Big),$$

where the remainder depends on k and M.

Proof. See Watson [24]. $\square$

We need to control the moments of $\widetilde{Y}_\Delta$ and compute the first ones. For that we use relation (4) and the moments of Y_Δ derived from the Lévy-Kintchine formula

$$\phi_{Y_\Delta}(w) = \mathbb{E}\big[e^{iwY_\Delta}\big] = \exp\big(\lambda\Delta((1-iw/\beta)^{-1} - 1 - iw/\beta)\big)$$

by the relation

$$\mathbb{E}\big[Y_\Delta^m\big] = \frac{1}{i^m}\frac{\partial^m\phi_{Y_\Delta}(w)}{\partial w^m}\Big|_{w=0}, \qquad m \in \mathbb{N}.$$

The control of the moments of Y_Δ is given in Lemma 4 hereafter, which is a consequence of the following Lemma, whose proof can be found in the Appendix.

Lemma 3. *Let $K \in \mathbb{N}$, suppose X is a compound Poisson process whose compound law is centered and has moment up to order K. Then for Δ large enough and $m \leq K$ we have $|\mathbb{E}[X_\Delta^m]| \leq \mathfrak{C}\Delta^{\lfloor m/2\rfloor}$, where $\mathfrak{C}$ continuously depends on λ and the K first moments.*

Remark 6. *Lemma 3 and Cauchy-Schwarz inequality imply $\mathbb{E}[|X_\Delta^{2m+1}|] \leq \mathfrak{C}\Delta^{m+1/2}$.*

Lemma 4. *Let $K \geq 2$, then $|\mathbb{E}[\widetilde{Y}_\Delta^m]| \leq \mathfrak{C}\Delta^{\lfloor m/2\rfloor}$, where $\mathfrak{C}$ continuously depends on θ.*

Proof of Lemma 4. The process Y is not in $\mathcal{P}$, nevertheless a convex inequality leads to

$$\mathbb{E}\big[Y_\Delta^m\big] \leq 2^m\Big(\mathbb{E}\Big[\Big(\sum_{i=1}^{R_\Delta}\big(\xi_i - \tfrac{1}{\beta}\big)\Big)^m\Big] + \frac{1}{\beta^m}\mathbb{E}\big[(R_\Delta - \lambda\Delta)^m\big]\Big).$$

We apply Lemma 3 to the first term of the right hand part of the inequality. We control the second term using Faà di Bruno's formula; we compute the nth derivative of the Laplace transform of $R_\Delta - \lambda\Delta$ at 0 as follows

$$\frac{d^n}{dt^n}\mathbb{E}\big[e^{t(R_\Delta - \lambda\Delta)}\big] = \frac{d^n}{dt^n}e^{\lambda\Delta(e^t - t - 1)} = \frac{d^n}{dt^n}F(G(t))$$

where $F(t) = e^{\lambda\Delta(t-1)}$ and $G(t) = e^t - t$, which satisfy $F^{(n)}(t)|_{t=0} = (\lambda\Delta)^n$ and $G^{(n)}(t)|_{t=0} = 1_{n\neq 1}$ for all $n \geq 1$. Applying Faà di Bruno's formula we get

$$\frac{d^n}{dt^n}F(G(t)) = \sum_{\substack{m_1, m_2, \ldots, m_n \\ m_1 + 2m_2 + \cdots + nm_n = n}} \tfrac{n!}{m_1! m_2! 2!^{m_2} \ldots m_n! n!^{m_n}} F^{(m_1 + \cdots + m_n)}(G(t)) \prod_{j=1}^{n}\Big(\tfrac{G^{(j)}(t)}{j!}\Big)^{m_j}.$$

Let $t = 0$. All the terms corresponding to $m_1 \neq 0$ are null, we obtain

$$\mathbb{E}\big[(R_\Delta - \lambda\Delta)^n\big] = \sum_{\substack{m_2, \ldots, m_n \\ 2m_2 + \cdots + nm_n = n}} \tfrac{n!}{m_1! m_2! 2!^{m_2} \ldots m_n! n!^{m_n}} (\lambda\Delta)^{m_2 + \cdots + m_n} \leq \mathfrak{C}\Delta^{\lfloor n/2 \rfloor},$$

for large enough Δ. The last inequality follows from the fact that, due to the constraint $2m_2 + 3m_3 + \cdots + nm_n = n$, for large enough Δ, the exponent $m_2 + \cdots + m_n$ is maximized for $m_3 = \cdots = m_n = 0$. The constant $\mathfrak{C}$ depends on λ. To conclude we control the moments of $\widetilde{Y}_\Delta$ using (4)

$$\mathbb{E}\big[\widetilde{Y}_\Delta^m\big] = \frac{1}{(1 - e^{-\lambda\Delta})}\Big(\mathbb{E}\big[Y_\Delta^m\big] - e^{-\lambda\Delta}\tfrac{\lambda\Delta}{\beta}\Big) \leq \mathfrak{C}\Delta^{\lfloor m/2 \rfloor},$$

for Δ large enough and where $\mathfrak{C}$ continuously depends on θ. □

Completion of the proof of Proposition 1

Since observations (3) are independent and identically distributed, the Fisher information satisfies $I_{\lfloor T\Delta_T^{-1}\rfloor, \Delta_T}(\theta) = \lfloor T\Delta_T^{-1}\rfloor I_{1,\Delta_T}(\theta)$

$$I_{1,\Delta_T}(\theta) = \begin{pmatrix} I_{\Delta_T}(\lambda, \lambda) & I_{\Delta_T}(\lambda, \beta) \\ I_{\Delta_T}(\beta, \lambda) & I_{\Delta_T}(\beta, \beta) \end{pmatrix}$$

where

$$I_{\Delta_T}(\lambda, \lambda) = -\mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[\frac{\partial^2}{\partial\lambda^2}\log \widetilde{p}_{\Delta,\theta}(\widetilde{Y}_{\Delta_T}, \lambda, \beta)\Big],$$

$$I_{\Delta_T}(\beta, \beta) = -\mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[\frac{\partial^2}{\partial\beta^2}\log \widetilde{p}_{\Delta,\theta}(\widetilde{Y}_{\Delta_T}, \lambda, \beta)\Big],$$

$$I_{\Delta_T}(\lambda, \beta) = I_{\Delta_T}(\beta, \lambda) = -\mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[\frac{\partial^2}{\partial\lambda\partial\beta}\log \widetilde{p}_{\Delta,\theta}(\widetilde{Y}_{\Delta_T}, \lambda, \beta)\Big]$$

From (10) we derive $g_k'(x) = g_{k+1}(x)$. Straightforward computations lead to

$$\begin{aligned}
I_{\Delta_T}(\lambda,\lambda) &= \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[\frac{\Delta_T^2 e^{-2\lambda\Delta_T}}{(1-e^{-\lambda\Delta_T})^2} + \frac{\Delta_T^2 e^{-\lambda\Delta_T}}{1-e^{-\lambda\Delta_T}} - 2\Delta_T^2 \frac{g_2(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)} \\
&\quad + \frac{1}{\lambda^2} - (\beta\Delta_T\widetilde{Y}_{\Delta_T} + 2\lambda\Delta_T^2)^2 \\
&\quad \times \Big(\frac{g_3(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)} - \Big(\frac{g_2(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}\Big)^2\Big)\Big] \\
I_{\Delta_T}(\lambda,\beta) &= \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[-\Delta_T\widetilde{Y}_{\Delta_T}\frac{g_2(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)} - \lambda\Delta_T\widetilde{Y}_{\Delta_T}(\beta\Delta_T\widetilde{Y}_{\Delta_T} + \\
&\quad 2\lambda\Delta_T^2) \times \Big(\frac{g_3(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)} - \Big(\frac{g_2(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}\Big)^2\Big)\Big] \\
I_{\Delta_T}(\beta,\beta) &= \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}\Big[\frac{1}{\beta^2} - (\lambda\Delta_T\widetilde{Y}_{\Delta_T})^2\Big(\frac{g_3(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)} \\
&\quad - \Big(\frac{g_2(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}{g_1(\lambda\Delta_T\beta\widetilde{Y}_{\Delta_T}+\lambda^2\Delta_T^2)}\Big)^2\Big)\Big].
\end{aligned}$$

Finally equation (11), Lemma 2 applied with $M = 8$, Lemma 4 (with Remark 6) and the Taylor expansions around 0 of $z \to 1/(1+z)$ up to order 4 in Δ lead to Proposition 1. Computations are made with Mathematica.

5.2 Proof of Theorem 1

Preliminary

Lemma 5. *Let Δ_T be such that $\Delta_T \to \infty$ and $T\Delta_T^{-2} \to \mathfrak{l} \in [0,\infty)$ as $T \to \infty$, then for $\gamma > 0$ and $\gamma \neq 1$*

$$\mathbb{E}_{\theta_0}\Big[\log\Big(\frac{g_1(\lambda_0\beta_0\Delta_T\widetilde{Y}_{\Delta_T}+\lambda_0^2\Delta_T^2)}{g_1(\gamma^3\lambda_0\beta_0\Delta_T\widetilde{Y}_{\Delta_T}+\gamma^4\lambda_0^2\Delta_T^2)}\Big)\Big] = 2\lambda_0\Delta_T(1-\gamma^2) + 3\log(\gamma) - \frac{9(\gamma^2-1)}{16\gamma^2\lambda_0\Delta_T} + O\Big(\frac{1}{\Delta_T^{3/2}}\Big).$$

Proof. It is a consequence of (11), Lemma 2 applied with $M = 8$ and Lemma 4 (with Remark 6). Computations are made with Mathematica. □

Completion of the proof of Theorem 1

The following inequality holds for all $\theta_0 \in \Theta$ and $\delta > 0$

$$\sup_{\theta\in\mathcal{V}_\delta(\theta_0)} \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big] \geq \int_{\mathcal{V}_\delta(\theta_0)} \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big]\mu(d\theta)$$

where $\mathcal{V}_\delta(\theta_0)$ is a neighborhood of θ_0 such that $diam(\mathcal{V}_\delta(\theta_0)) < \delta$ and μ is the following measure on $\mathcal{V}_\delta(\theta_0)$

$$\mu(dx) = \frac{1}{2}\big(\delta_{\theta_0}(dx) + \delta_{h_\gamma(\theta_0)}(dx)\big)$$

where $h_\gamma(\theta_0) \in \mathcal{V}(\theta_0)$ is a perturbation of θ_0 and δ_θ denotes the Dirac distribution in θ. For the reader convenience, wherever there is no ambiguity, we drop the super and sub scripts as follows,

$$\mathbb{E}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}[\cdot] := \mathbb{E}_{\widetilde{\mathbb{P}}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}[\cdot], \quad \mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} := \widetilde{\mathbb{P}}_{\theta_0}^{\otimes\lfloor T\Delta_T^{-1}\rfloor} \quad \text{and} \quad \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor} := \widetilde{\mathbb{P}}_{h_\gamma(\theta_0)}^{\otimes\lfloor T\Delta_T^{-1}\rfloor},$$

with the convention $\mathbb{P}^{\otimes 1} = \mathbb{P}$. It follows that

$$\begin{aligned}\int_{\mathcal{V}_\delta(\theta_0)} \mathbb{E}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big]\mu(d\theta) &= \frac{1}{2}\Big(\mathbb{E}_{\theta_0}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta_0\|\big] + \mathbb{E}_{\theta_0}^{\lfloor T\Delta_T^{-1}\rfloor}\Big[\|\widehat{\theta}-h_\gamma(\theta_0)\|\frac{d\mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor}}\Big]\Big)\\ &\geq \mathbb{E}_{\theta_0}^{\lfloor T\Delta_T^{-1}\rfloor}\Big[\frac{e^{-s}}{2}(\|\widehat{\theta}-\theta_0\| + \|\widehat{\theta}-h_\gamma(\theta_0)\|)\mathbf{1}\Big\{\frac{d\mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor}} > e^{-s}\Big\}\Big] \qquad (12)\end{aligned}$$

for any $s > 0$. The triangle inequality applied to (12) gives

$$\int_{\mathcal{V}_\delta(\theta_0)} \mathbb{E}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big]\mu(d\theta) \geq \frac{e^{-s}}{2}\|\theta_0 - h_\gamma(\theta_0)\|\widetilde{\mathbb{P}}_{\theta_0}\Big(\frac{d\mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}}{d\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor}} > e^{-s}\Big).$$

Noticing that for any $s > 0$ and $\mathbb{P}$ and $\mathbb{Q}$ some probabilities

$$\mathbb{P}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}} > e^{-s}\Big) = 1 - \mathbb{P}\Big(1 - \frac{d\mathbb{Q}}{d\mathbb{P}} > 1 - e^{-s}\Big) \geq 1 - \mathbb{P}\Big(\Big|1 - \frac{d\mathbb{Q}}{d\mathbb{P}}\Big| > 1 - e^{-s}\Big),$$

Markov's inequality and $\|\mathbb{P}-\mathbb{Q}\|_{TV} = \int |d\mathbb{P} - d\mathbb{Q}|$, lead to

$$\mathbb{P}\Big(\frac{d\mathbb{Q}}{d\mathbb{P}} > e^{-s}\Big) \geq 1 - \frac{1}{1-e^{-s}}\|\mathbb{P}-\mathbb{Q}\|_{TV}.$$

Then, for all $s > 0$

$$\int_{\mathcal{V}_\delta(\theta_0)} \mathbb{E}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big]\mu(d\theta) \geq \|\theta_0 - h_\gamma(\theta_0)\|\frac{e^{-s}}{2}\Big(1 - \frac{1}{1-e^{-s}}\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV}\Big).$$

Hence,

$$\int_{\mathcal{V}_\delta(\theta_0)} \mathbb{E}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{\theta}-\theta\|\big]\mu(d\theta) \geq \|\theta_0 - h_\gamma(\theta_0)\|\Phi\big(\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV}\big) \qquad (13)$$

where

$$\Phi(x) = \sup_{s\in(0,\infty)} \frac{e^{-s}}{2}\Big(1 - \frac{1}{1-e^{-s}}x\Big) = \frac{(1-\sqrt{x})^2}{2}, \qquad x \in [0,1].$$

If x is bounded away from 1, Φ is strictly positive. In the remaining of the proof we choose h_γ such that

- $\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV} \leq \mathfrak{C}_1 < 1$ for some constant $\mathfrak{C}_1$,
- $\|\theta_0 - h_\gamma(\theta_0)\| \geq \mathfrak{C}_2 > 0$ for some constant $\mathfrak{C}_2$ possibly depending on θ_0.

Define the function $h_\gamma : \theta \to h_\gamma(\theta) = (\gamma^2\lambda, \gamma\beta)$ where $\gamma \neq 1$ is positive. First, Pinsker's inequality gives

$$\big\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \sqrt{\tfrac{\lfloor T\Delta_T^{-1}\rfloor}{2} K\big(\mathbb{P}, \mathbb{P}_\gamma\big)} := \sqrt{\tfrac{\lfloor T\Delta_T^{-1}\rfloor}{2} K\big(\widetilde{\mathbb{P}}_{\theta_0}, \widetilde{\mathbb{P}}_{h_\gamma(\theta_0)}\big)}, \tag{14}$$

where K is the Kullback divergence and

$$\begin{aligned} K\big(\widetilde{\mathbb{P}}_{\theta_0},\widetilde{\mathbb{P}}_{h_\gamma(\theta_0)}\big) &= \int_{-\infty}^{\infty} \big(\log(\widetilde{p}_{\theta_0}(x)) - \log(\widetilde{p}_{h_\gamma(\theta_0)}(x))\big)\widetilde{p}_{\theta_0}(x)dx \\ &= \mathbb{E}_{\widetilde{\mathbb{P}}_{\theta_0}}\big[\log\big(\tfrac{g_1(\lambda_0\beta_0\Delta_T X_{\Delta_T} + \lambda_0^2\Delta_T^2)}{g_1(\gamma^3\lambda_0\beta_0\Delta_T X_{\Delta_T} + \gamma^4\lambda_0^2\Delta_T^2)}\big)\big] - 2\lambda_0\Delta(1-\gamma^2) - 3\log(\gamma). \end{aligned}$$

In view of Lemma 5,

$$K\big(\widetilde{\mathbb{P}}_{\theta_0}, \widetilde{\mathbb{P}}_{h_\gamma(\theta_0)}\big) = \frac{9(1-\gamma^2)}{16\gamma^2\lambda_0\Delta_T} + O\Big(\tfrac{1}{\Delta_T^{3/2}}\Big),$$

and

$$\big\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \sqrt{\frac{9(1-\gamma^2)}{32\gamma^2\lambda_0}\frac{\lfloor T\Delta_T^{-1}\rfloor}{\Delta_T} + O\Big(\tfrac{T}{\Delta_T^{5/2}}\Big)}. \tag{15}$$

Then, if $T/\Delta_T^2 \to 0$ as $T \to \infty$, for large enough T there exits $\mathfrak{C}_1 < 1$ such that

$$\big\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\big\|_{TV} \leq \mathfrak{C}_1 < 1. \tag{16}$$

The inequality holds for any γ. If $T/\Delta_T^2 \to \mathfrak{l} > 0$ as $T \to \infty$, take $\gamma \neq 1$ such that

$$0 < \big(\tfrac{16\lambda_0}{9\mathfrak{l}} + 1\big)^{-1} < \gamma^2. \tag{17}$$

Then, (15) ensures that there exists $\mathfrak{C}_1 < 1$ such that

$$\|\mathbb{P}^{\lfloor T\Delta_T^{-1}\rfloor} - \mathbb{P}_\gamma^{\lfloor T\Delta_T^{-1}\rfloor}\|_{TV} \leq \mathfrak{C}_1 < 1. \tag{18}$$

Second, we bound from below $\|\theta_0 - h_\gamma(\theta_0)\|$, here $\|\cdot\|$ denotes the L_2 norm. Since

$$\|\theta_0 - h_\gamma(\theta_0)\| = \sqrt{(1-\gamma^2)^2\lambda_0^2 + (1-\gamma)^2\beta_0^2} = |1-\gamma|\sqrt{(1+\gamma)^2\lambda_0^2 + \beta_0^2},$$

we choose $\gamma \neq 1$ such that (17) is satisfied and $h_\gamma(\theta_0) \in \mathcal{V}_\delta(\theta_0)$. That latest condition can always be fulfilled since we can have either $\gamma > 1$ or $\gamma < 1$, avoiding boundary issues. Finally, there exists $\mathfrak{C}_2 > 0$, depending on γ and θ_0, such that

$$\|\theta_0 - h_\gamma(\theta_0)\| \geq \mathfrak{C}_2 > 0. \tag{19}$$

We complete the proof plugging (16), (18) and (19) into (13) and taking limits.

Remark 7. *To bound the total variation norm in (14) we prefer the Kullback divergence over the Hellinger distance since the logarithm makes easier the manipulation of the density $\widetilde{p}_{\theta,\Delta}$ (see Lemma 5).*

5.3 Proof of Lemma 1

Both experiments $\mathcal{E}^n$ and $\mathcal{G}^n$ are dominated by $\nu_{n,\theta}(dx) = \mu_{n,\theta}(dx) + dx$, where $\mu_{n,\theta}$ is a dominating measure for $w_{n,\Delta}$, therefore to establish the asymptotic equivalence it is sufficient to show (see Le Cam and Yang [11])

$$\sup_{\theta\in\Sigma}\big\|\mathbb{P}_\theta^{\otimes n} - \mathbb{Q}_\theta^{\otimes n}\big\|_{TV} \to 0 \ \text{ as } \ n \to \infty,$$

where $\|\cdot\|_{TV}$ denotes the total variation norm. Since each experiment is the n fold product of independent and identically distributed random variables[1] the result

$$\big\|\mathbb{P}_\theta^{\otimes n} - \mathbb{Q}_\theta^{\otimes n}\big\|_{TV} \to 0 \ \text{ as } \ n \to \infty.$$

is implied by

$$\big\|\mathcal{L}(X) - \mathcal{L}(Z)\big\|_{TV} = o\big(n^{-1}\big),$$

if X has density $p_n(\theta,\cdot)$ and Z has density $q_n(\theta,\cdot)$. The connection between the total variation norm and the L_1 norm leads to

$$\begin{aligned}\big\|\mathcal{L}(X) - \mathcal{L}(Z)\big\|_{TV} &= \frac{1}{2}\int_{\mathbb{R}}\big|p_n(\theta,x) - q_n(\theta,x)\big|\nu_{n,\theta}(dx)\\ &= \frac{a_n(\theta)}{2}\int_{\mathbb{R}}\big|f_n(\theta,x) - w_{n,\theta}(x)\big|\nu_{n,\theta}(dx) \le a_n(\theta).\end{aligned}$$

The condition $\sup_{\theta\in\Sigma} a_n(\theta) = o(\frac{1}{n})$ completes the proof of Lemma 1.

Remark 8. *The last inequality is an equality when $\mu_{n,\theta}$ and the Lebesgue measure are orthogonal.*

5.4 Proof of Theorem 2

Preliminary

The process Y defined in Section 2 is not in $\mathcal{P}$, we build a compound Poisson process V close to Y in total variation norm. Keeping up with notation of Section 2, $\theta = (\lambda,\beta) \in \Theta$, where Θ is compact subset of $(0,\infty)\times(0,\infty)$, consider the process V

$$V_s = \sum_{i=1}^{N_s}\epsilon_i, \quad s \ge 0 \tag{20}$$

[1]For instance, by using the bound (see Tsybakov [22] pp. 83–90)

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

where N is a Poisson process of intensity $\frac{8}{9}\lambda$ and independent of (ϵ_i) which are independent and identically distributed centered exponential variables with parameter $\frac{2}{3}\beta$. Their common density is

$$f_\theta(x) = \tfrac{2}{3}\beta e^{-\frac{2}{3}\beta(x+1/(\frac{2}{3}\beta))}, \quad x \geq -1/\tfrac{2\beta}{3}. \tag{21}$$

Remark 9. *The multiplicative constants $\frac{8}{9}$ and $\frac{2}{3}$ in front of λ and β ensure that Y_Δ defined by (2) and V_Δ have same moments of order 2 and 3.*

Consider the observations $(V_{i\Delta} - V_{(i-1)\Delta},\ i = 1, \dots, \lfloor T\Delta^{-1}\rfloor)$ and denote by $\mathbb{Q}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}$ its law. We have the following Lemma.

Lemma 6. *Let $\Delta_T \to \infty$ such that $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ and $T/\Delta_T \to \infty$ as $T \to \infty$. Then, for any compact set $\Theta \subset (0,\infty) \times (0,\infty)$*

$$\sup_{\theta\in\Theta} \big\| \mathbb{P}_\theta^{\otimes \lfloor T\Delta_T^{-1}\rfloor} - \mathbb{Q}_\theta^{\otimes \lfloor T\Delta_T^{-1}\rfloor} \big\|_{TV} \to 0,$$

where $\mathbb{P}_\theta^{\lfloor T\Delta_T^{-1}\rfloor}$ denotes the law of $(Y_{i\Delta} - Y_{(i-1)\Delta},\ i = 1, \dots, \lfloor T\Delta_T^{-1}\rfloor)$.

Proof of Lemma 6 can be found in the Appendix. The steps of the proof follows the lines of the proof of Theorem 3 hereafter.

Completion of the proof of Theorem 2

Let $r_\theta = (\lambda, f_\beta)$ defined from (20) and (21), for all $r_0 \in \mathcal{F}$ and $\delta > 0$

$$\sup_{r\in\mathcal{V}_\delta(r_0)} \mathbb{E}_{\mathbb{P}_r}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r\|_{2,\mathcal{P}}\big] \geq \sup_{r_\theta\in\mathcal{V}_\delta(r_{\theta_0})} \mathbb{E}_{\mathbb{Q}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r_\theta\|_{2,\mathcal{P}}\big]$$

where the neighborhood $\mathcal{V}_\delta(r_0)$ (resp. $\mathcal{V}_\delta(r_{\theta_0})$) of r_0 (resp. r_{θ_0}) is such that $diam(\mathcal{V}_\delta(r_0)) < \delta$, $diam(\mathcal{V}_\delta(r_{\theta_0})) < \delta$ and $\mathcal{V}_\delta(r_{\theta_0}) \subset \mathcal{V}_\delta(r_0)$. Notice that

$$\inf_{\widehat{r}} \sup_{r_\theta\in\mathcal{V}_\delta(r_{\theta_0})} \mathbb{E}_{\mathbb{Q}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r_\theta\|_{2,\mathcal{P}}\big] = \inf_{\widehat{r}\in\mathcal{V}_\delta(r_{\theta_0})} \sup_{r_\theta\in\mathcal{V}_\delta(r_{\theta_0})} \mathbb{E}_{\mathbb{Q}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r} - r_\theta\|_{2,\mathcal{P}}\big].$$

Otherwise if $\hat{r} \notin \mathcal{V}_\delta(r_{\theta_0})$, define $\Pi_{\mathcal{V}_\delta(r_{\theta_0})}$ the projection operator onto $\mathcal{V}_\delta(r_{\theta_0})$, we immediately get for all $r_\theta \in \mathcal{V}_\delta(r_{\theta_0})$

$$\|\hat{r} - r_\theta\|_{2,\mathcal{P}} \geq \|\Pi_{\mathcal{V}_\delta(r_{\theta_0})}[\hat{r}] - r_\theta\|_{2,\mathcal{P}}.$$

It follows that for all $\widehat{r},\ r_\theta$ in $\mathcal{V}_\delta(r_{\theta_0})$ we have

$$\|\widehat{r} - r_\theta\|_{2,\mathcal{P}} \leq 2(\delta + \|r_{\theta_0}\|_{2,\mathcal{P}}). \tag{22}$$

The remainder of the proof is a consequence of Scheffé's theorem. Let F be a bounded function then for every measures $\mathbb{P}$ and $\mathbb{Q}$

$$\big|\mathbb{E}_{\mathbb{P}}[F(X)] - \mathbb{E}_{\mathbb{Q}}[F(X)]\big| \leq \|F\|_\infty \int |d\mathbb{P} - d\mathbb{Q}| = 2\|F\|_\infty \|\mathbb{P}-\mathbb{Q}\|_{TV}. \tag{23}$$

It follows from (22) and (23)

$$\mathbb{E}_{\mathbb{Q}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\theta\|_{2,\mathcal{P}}\big] \geq \mathbb{E}_{\mathbb{P}_\theta}^{\lfloor T\Delta_T^{-1}\rfloor}\big[\|\widehat{r}-r_\theta\|_{2,\mathcal{P}}\big] - 2\big(2(\delta+\|r_{\theta_0}\|)\|\mathbb{P}_\theta - \mathbb{Q}_\theta\|_{TV}\big).$$

We conclude the proof with Lemma 6, Theorem 1 and taking limits.

5.5 Proof of Theorem 3

Preliminary

Lemma 7. *Let f be a density with respect to the Lebesgue measure, centered and with finite variance. Then, for every $\rho>0$ there exists $a<1$ such that $|f^*(\xi)|\leq a,\ \forall|\xi|>\rho$.*

Proof of Lemma 7. First, we establish that $|f^*(\xi)|=1$ if and only if $\xi=0$. Only the direct implication needs justification. Denote $\mathcal{D}=\{x, f(x)>0\}$ and suppose $|f^*(\xi)|=1$, taking square leads to

$$\begin{aligned}0 = |f^*(\xi)|^2 - 1 &= \mathcal{R}e\Big(\int_{\mathcal{D}^2} e^{i\xi(x-y)} f(x)f(y)dxdy\Big) - \int_{\mathcal{D}^2} f(x)f(y)dxdy\\ &= \int_{\mathcal{D}^2}\big(\cos(\xi(x-y))-1\big)f(x)f(y)dxdy,\end{aligned}$$

where $\mathcal{R}e(z)$ designates the real part of z. Since $(\cos(\xi(x-y))-1)f(x)f(y)\leq 0$ for all $x,\ y\in\mathcal{D}$ we derive that $\cos(\xi(x-y))=1$ for all $x,\ y\in\mathcal{D}$. Since $\mathcal{D}$ is an interval of $\mathbb{R}$ we have $\xi=0$.

Second, by Riemann-Lebesgue Lemma there exists $A>0$ such that $\forall|\xi|>A$, $|f^*(\xi)|<1/2$. Since f has finite expectation $\xi\to|f^*(\xi)|$ is continuous. It is continuous over the compact $[\rho, A]$ and reaches its supremum, denoted S, which is by the first part of the proof strictly lower than 1. Finally, since f has finite variance $f^{*\prime\prime}(0)>0$ and for all $\rho>0$ we have $|f^*(\rho)|<1$. Set $a=S\vee\frac{1}{2}\vee|f^*(\rho)|$. Proof is now complete. ☐

To establish Theorem 3, we show that the total variation norm between the experiments vanishes, using the Lévy structure of the processes X, Z and W. The experiments $\mathcal{X}^{\Delta_T}$, $\mathcal{Z}^{\Delta_T}$ and $\mathcal{W}^{\Delta_T}$ are dominated by the measure $\delta_0(dx)+dx$. Introduce

$$p_{\Delta_T,\theta}(x) = e^{-\lambda\Delta_T}\delta_0(x) + (1-e^{-\lambda\Delta_T})\widetilde{p}_{\Delta_T,\theta}(x) \tag{24}$$
$$q_{\Delta_T,h_\gamma(\theta)}(x) = e^{-\gamma\lambda\Delta_T}\delta_0(x) + (1-e^{-\gamma\lambda\Delta_T})\widetilde{q}_{\Delta_T,h_\gamma(\theta)}(x) \tag{25}$$

where $p_{\Delta_T,\theta}$ and $q_{\Delta_T,h_\gamma(\theta)}$ are the distributions of X_{Δ_T} and Z_{Δ_T} and $\widetilde{p}_{\Delta_T,\theta}$ and $\widetilde{q}_{\Delta_T,h_\gamma(\theta)}$ are absolutely continuous with respect to the Lebesgue measure. For the reader convenience, in absence of ambiguity we drop the subscripts and set $\widetilde{p}:=\widetilde{p}_{\Delta_T,\theta}$ and $\widetilde{q}:=\widetilde{q}_{\Delta_T,h_\gamma(\theta)}$.

Proof of Theorem 3.1

We prove that for $K \geq 2$, Δ_T satisfying the rate restriction

$$T/\Delta_T^{(K+1)/2} = o((\log(T/\Delta_T))^{-1/4}) \qquad \text{as } T \to \infty \tag{26}$$

and the condition $\sup_{\theta \in \Sigma_K} |\int x^3 f_\theta(x)dx| < \infty$ the experiments $\mathcal{X}^{\Delta_T}$ and $\mathcal{Z}^{\Delta_T}$ are asymptotically equivalent. They live on the same state space and are the $\lfloor T\Delta_T^{-1} \rfloor$ fold product of independent and identically distributed random variables, therefore it is sufficient to show (see Section 5.3 the proof of Lemma 1)

$$\big\|\mathcal{L}(X_{\Delta_T}) - \mathcal{L}(Z_{\Delta_T})\big\|_{TV} = o\big(\lfloor T\Delta_T^{-1} \rfloor^{-1}\big). \tag{27}$$

We have

$$\begin{aligned}\|\mathcal{L}(X_{\Delta_T}) - \mathcal{L}(Z_{\Delta_T})\|_{TV} = \frac{1}{2}\int_{\mathbb{R}} &|(1-e^{-\lambda\Delta_T})\widetilde{p}(x) \\ &- (1-e^{-\gamma\lambda\Delta_T})\widetilde{q}(x)|dx + \frac{1}{2}|e^{-\lambda\Delta_T} - e^{-\gamma\lambda\Delta_T}|.\end{aligned}$$

Where $\widetilde{p}$ and $\widetilde{q}$ are defined in (24) and (25), and $|e^{-\lambda\Delta_T} - e^{-\gamma\lambda\Delta_T}| = o(\lfloor T\Delta_T^{-1}\rfloor^{-1})$, as Δ is of the order of T^α for some $\alpha > 0$. Applying successively the triangle inequality and Cauchy-Schwarz we obtain

$$\int_{\mathbb{R}} |(1-e^{-\lambda\Delta_T})\widetilde{p}(x) - (1-e^{-\gamma\lambda\Delta_T})\widetilde{q}(x)|dx \leq I + II + III$$

where for any $\eta > 0$,

$$\begin{aligned} I &= \sqrt{2\eta}\Big(\int_{\mathbb{R}} \big((1-e^{-\lambda\Delta_T})\widetilde{p}(x) - (1-e^{-\gamma\lambda\Delta_T})\widetilde{q}(x)\big)^2 dx\Big)^{1/2}, \\ II &= (1-e^{-\lambda\Delta_T})\int_{|x|>\eta} \widetilde{p}(x)dx \ \text{ and } \ III = (1-e^{-\gamma\lambda\Delta_T})\int_{|x|>\eta} \widetilde{q}(x)dx. \end{aligned}$$

Set $\eta = \eta_T = \kappa\sqrt{\Delta_T \log(T/\Delta_T)}$, we claim that for $\kappa^2 > 2\lambda m$, the terms I, II and III are $o((T/\Delta_T^{-1}))$ hence (27) and the result.

Bounding terms II and III. For II we use that

$$(1-e^{-\lambda\Delta_T})\int_{|x|>\eta_T} \widetilde{p}(x)dx = \int_{|x|>\eta_T} p_{\Delta_T,\theta}(x)dx = \mathbb{P}_\theta(|X_{\Delta_T}| > \eta_T)$$

and that X_{Δ_T} is a centered compound Poisson process whose compound law has finite variance, it follows that

$$\frac{X_{\Delta_T}}{\sqrt{\Delta_T}} \to \mathcal{N}(0, \lambda m_2) \qquad \text{as } T \to \infty.$$

Let $D \sim \mathcal{N}(0, \lambda m_2)$, the triangle inequality gives

$$\mathbb{P}_\theta\big(|X_{\Delta_T}| \geq \eta_T\big) \leq \mathbb{P}\big(|D| \geq \kappa\sqrt{\log(T/\Delta_T)}\big) \\ + \big|\mathbb{P}\big(|D| \geq \kappa\sqrt{\log(T/\Delta_T)}\big) - \mathbb{P}\big(\big|\tfrac{X_{\Delta_T}}{\sqrt{\Delta_T}}\big| \geq \kappa\sqrt{\log(T/\Delta_T)}\big)\big|.$$

We readily obtain

$$\mathbb{P}\big(|D| \geq \kappa\sqrt{\log(T/\Delta_T)}\big) \leq 2(T/\Delta_T)^{-\kappa^2/(2\lambda m_2)} = o\big((T/\Delta_T)^{-1}\big).$$

We bound the second term using Edgeworth series, even if it means conditioning on the value of the Poisson process associated to X. By assumption, the compound law has finite moment of order 3, denoted m_3, uniformly bounded over Σ_K, we derive

$$\begin{aligned}\big|\mathbb{P}\big(|D| \geq \kappa\sqrt{\log(T/\Delta_T)}\big) - \mathbb{P}_\theta\big(\big|\tfrac{X_{\Delta_T}}{\sqrt{\Delta_T}}\big| \geq \kappa\sqrt{\log(T/\Delta_T)}\big)\big| \\ &\leq \bigg|\frac{\mathfrak{C}}{\sqrt{\Delta_T}}\frac{\partial^3}{\partial x^3}\int_x^\infty e^{-\frac{s^2}{2\lambda m_2}}ds\Big|_{x=\kappa\sqrt{\log(T/\Delta_T)}}\bigg| \\ &= \frac{\mathfrak{C}}{\lambda m_2\sqrt{\Delta_T}}\bigg|1 - \frac{\kappa^2\log(T/\Delta_T)}{\lambda m_2}\bigg|e^{-\frac{\kappa^2\log(T/\Delta_T)}{2\lambda m_2}} \\ &\leq \frac{\mathfrak{C}\log(T/\Delta_T)}{\sqrt{\Delta_T}}(T/\Delta_T)^{-\kappa^2/(2\lambda m_2)} = o\big((T/\Delta_T)^{-1}\big)\end{aligned}$$

where $\mathfrak{C}$ continuously depends on λ, m_2 and m_3 and which is $o((T/\Delta_T)^{-1})$ for $\kappa^2 \geq 2\lambda m_2$. The term III is treated similarly as II, the parameter γ simplifies. We do not reproduce computations. Thus II and III have the right order, the choice of κ and the bounds on II and III is made independent of θ taking the supremum over the compact set Σ_K.

Bounding term I. Plancherel theorem gives

$$\begin{aligned}A = &\int_{\mathbb{R}}\big((1-e^{-\lambda\Delta_T})\widetilde{p}(x) - (1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}(x)\big)^2 dx \\ &= \frac{1}{2\pi}\int_{\mathbb{R}}|(1 - e^{-\lambda\Delta_T})\widetilde{p}^*(\xi) - (1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}^*(\xi)|^2 d\xi,\end{aligned}$$

where f^* denotes the Fourier transform of f. The Fourier transforms are computed using (24) and (25) and the Lévy-Kintchine formula

$$\begin{aligned}(1 - e^{-\lambda\Delta_T})\widetilde{p}^*(\xi) &= \exp\big(\lambda\Delta_T(f_\theta^*(\xi) - 1)\big) - e^{-\lambda\Delta_T} \\ (1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}^*(\xi) &= \exp\big(\gamma\lambda\Delta_T(f_{h_\gamma(\theta)}^*(\xi) - 1)\big) - e^{-\gamma\lambda\Delta_T}.\end{aligned}$$

Then, $2\pi A$ can be upper bounded as follows

$$\int_{\mathbb{R}}\big|(1 - e^{-\lambda\Delta_T})\widetilde{p}^*(\tfrac{\xi}{\sqrt{\Delta_T}}) - (1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}^*(\tfrac{\xi}{\sqrt{\Delta_T}})\big|^2\frac{d\xi}{\sqrt{\Delta_T}} \leq IV + V + VI,$$

where for $\rho \geq 0$

$$IV = \frac{1}{2\pi} \int_{|\xi| \leq \rho\sqrt{\Delta_T}} \big|(1 - e^{-\lambda\Delta_T})\widetilde{p}^*(\tfrac{\xi}{\sqrt{\Delta_T}}) - (1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}^*(\tfrac{\xi}{\sqrt{\Delta_T}})\big|^2 \frac{d\xi}{\sqrt{\Delta_T}},$$

$$V = \tfrac{1}{2\pi} \int_{|\xi| > \rho} \big|(1 - e^{-\lambda\Delta_T})\widetilde{p}^*(\xi)\big|^2 d\xi \ \text{ and } \ VI = \tfrac{1}{2\pi} \int_{|\xi| > \rho} \big|(1 - e^{-\gamma\lambda\Delta_T})\widetilde{q}^*(\xi)\big|^2 d\xi.$$

Bounding term IV. Since f_θ and $f_{h_\gamma(\theta)}$ have their K first moments finite, we get the following expansion for any bounded ξ

$$f_\theta^*(\xi) - \Big(1 - \frac{m_2\xi^2}{2} + \cdots + \frac{i^K m_K \xi^K}{K!}\Big) = \xi^{K+1}\alpha_1(\xi)$$

and

$$f_{h_\gamma(\theta)}^*(\xi) - \Big(1 - \frac{m_2\xi^2}{2\gamma} + \cdots + \frac{i^K m_K \xi^K}{K!\gamma}\Big) = \xi^{K+1}\alpha_2(\xi)$$

for some bounded functions $\xi \rightsquigarrow \alpha_1(\xi)$ and $\xi \rightsquigarrow \alpha_2(\xi)$. It follows that IV is less than

$$\int_{|\xi| \leq \rho\sqrt{\Delta_T}} \Big| e^{-\lambda m_2 \frac{\xi^2}{2} + \cdots + i^K \lambda m_K \frac{\xi^K}{\sqrt{\Delta_T}^{K-2} K!}} \Big|^2 \frac{\xi^{2K+2}}{\Delta_T^{K-1}} \alpha^2(\tfrac{\xi}{\sqrt{\Delta_T}})$$
$$\times \exp\Big(2\frac{\xi^{K+1}}{\sqrt{\Delta_T}^{K-1}}\alpha\big(\tfrac{\xi}{\sqrt{\Delta_T}}\big)\Big)\frac{d\xi}{\sqrt{\Delta_T}} + 2\rho\sqrt{\Delta_T}(e^{-\lambda\Delta_T} + e^{-\gamma\lambda\Delta_T})$$

for some bounded function $\xi \rightsquigarrow \alpha(\xi)$. Set $\overline{\alpha} = \sup_x |\alpha(x)|$, then IV is bounded by

$$\overline{\alpha}^2 \int_{|\xi| \leq \rho\sqrt{\Delta_T}} \frac{\xi^{2K+2}}{\Delta_T^{K-1}} \exp\Big(- \big(\lambda(m_2 - \sum_{k=2}^{\lfloor K/2 \rfloor} \tfrac{(-1)^k m_{2k}\rho^{2k-2}}{2(2k)!}) + 2\rho^{K-1}\overline{\alpha}\big)\xi^2\Big)\frac{d\xi}{\sqrt{\Delta_T}}$$
$$+ 2\rho\sqrt{\Delta_T}(e^{-\lambda\Delta_T} + e^{-\gamma\lambda\Delta_T}).$$

We pick ρ such that $\rho > 0$

$$\lambda\Big(m_2 - \sum_{k=2}^{\lfloor K/2 \rfloor} \frac{(-1)^k m_{2k}\rho^{2k-2}}{2(2k)!}\Big) + 2\rho^{K-1}\overline{\alpha} > 0. \tag{28}$$

Even if it means taking ρ small, condition (28) can always be satisfied. Using that the Gaussian density has finite moment of order $2K+2$, term IV is of order $\Delta_T^{-(2K-1)/2}$.

Bounding terms V and VI. For any $A > \rho$,

$$\begin{aligned} V &= \frac{(1 - e^{-\lambda\Delta_T})}{2\pi} \int_{|\xi| > \rho} \big|\widetilde{p}^*(\xi)\big|^2 d\xi = \frac{e^{-\lambda\Delta_T}}{2\pi} \int_{|\xi| > \rho} \big|e^{\lambda\Delta_T f_\theta^*(\xi)} - 1\big|^2 d\xi \\ &= \frac{e^{-\lambda\Delta_T}}{2\pi} \int_{A > |\xi| > \rho} \big|e^{\lambda\Delta_T f_\theta^*(\xi)} - 1\big|^2 d\xi + \frac{e^{-\lambda\Delta_T}}{2\pi} \int_{|\xi| > A} \big|e^{\lambda\Delta_T f_\theta^*(\xi)} - 1\big|^2 d\xi \\ &= VII + VIII. \end{aligned}$$

First, by Lemma 7 VII is bounded by constant times $Ae^{-(1-a)\lambda\Delta_T} = o((T/\Delta_T)^{-1})$ as $a<1$. Second, since f_θ belongs to $\mathcal{F}$, there exist $C>0$ such that for all $\xi \geq A$, $|f^*(\xi)/\xi| \leq C$.

$$\begin{aligned}
VIII &\leq \frac{2e^{-\lambda\Delta_T}}{2\pi}\int_A^\infty \Big|\sum_{l=1}^\infty \frac{(\lambda\Delta_T C)^l}{l!}\frac{1}{\xi^l}\Big|^2 d\xi \\
&\leq \frac{e^{-\lambda\Delta_T}}{\pi}\sum_{l_1=1}^\infty\sum_{l_2=1}^\infty \frac{(\lambda\Delta_T C)^{l_1+l_2}}{l_1!l_2}\int_A^\infty \frac{1}{\xi^{l_1+l_2}}d\xi \\
&= \frac{e^{-\lambda\Delta_T}}{\pi}\sum_{l_1=1}^\infty\sum_{l_2=1}^\infty \frac{(\lambda\Delta_T C)^{l_1+l_2}}{l_1!l_2!}\frac{1}{(l_1+l_2-1)A^{l_1+l_2-1}} \leq \frac{\rho}{\pi}e^{-(1-2\frac{C}{A})\lambda\Delta_T}.
\end{aligned}$$

Fix $A>2C$, then, $(\sqrt{\eta_T}V)^{1/2}$ is of order $(\sqrt{\Delta_T}e^{-(1-2\frac{C}{A})\lambda\Delta_T})^{1/2} = o((T/\Delta_T)^{-1})$. The term VI is treated similarly and is of the same order.

Completion of the proof of Theorem 3.1. The leading quantity is IV, we deduce that I is of order $\eta_T^{1/2}\Delta_T^{-(2K-1)/4}$. The choice $\eta_T = \kappa\sqrt{\Delta_T\log(T/\Delta_T)}$ and restriction (26) imply $I = o((T/\Delta_T)^{-1})$. The proof of Theorem 3.1 is completed taking the supremum in θ over the compact set Σ_K.

Proof of Theorem 3.2

Proof of part 2 of Theorem 3 is deduced from above computations replacing Z with W and applying modifications *i.* or *ii.*

Appendix

Proof of Lemma 3

We prove the result by induction on m. The Lévy-Kintchine formula gives an explicit formula of the Fourier transform of X_Δ

$$\phi_{X_\Delta}(w) = \mathbb{E}\big[e^{iwX_\Delta}\big] = \exp\big(\lambda\Delta(f^*(w)-1)\big)$$

where $f^*(w) = \mathbb{E}[e^{iw\xi}]$ denotes the Fourier transform of the compound law and λ is the intensity of the Poisson process. The moments of X_Δ are obtained with

$$\mathbb{E}\big[X_\Delta^m\big] = \frac{1}{i^m}\frac{\partial^m\phi_{X_\Delta}(w)}{\partial w^m}\Big|_{w=0}, \qquad m\in\mathbb{N}. \tag{29}$$

We prove by induction the following property, for all $m \leq \lfloor\frac{K-1}{2}\rfloor$

$$\begin{aligned}
\frac{\partial^{2m}\phi_{X_\Delta}(w)}{\partial w^{2m}} &= \big(P_{2m}(w,\Delta)+Q_{2m}(w,\Delta)\big)\exp\big(\lambda\Delta(f^*(w)-1)\big) \\
\frac{\partial^{2m+1}\phi_{X_\Delta}(w)}{\partial w^{2m+1}} &= \big(P_{2m+1}(w,\Delta)+Q_{2m+1}(w,\Delta)\big)\exp\big(\lambda\Delta(f^*(w)-1)\big)
\end{aligned}$$

where the functions $\Delta \to P_{2m}(w,\Delta)$, $\Delta \to Q_{2m}(w,\Delta)$, $\Delta \to P_{2m+1}(w,\Delta)$ and $\Delta \to Q_{2m+1}(w,\Delta)$ are polynomials in Δ, the degree of Q_{2m} and Q_{2m+1} is smaller than m and there exist C^1 functions $(c_{2m,j}(\cdot), c_{2m+1,j}(\cdot), j = 1,\dots,m)$, continuously depending on λ, such that

$$
\begin{aligned}
P_{2m}(w,\Delta) &= \sum_{j=1}^{m} c_{2m,j}(w) f^{*\prime}(w)^{2j} \Delta^{m+j} \quad P_{2m+1}(w,\Delta) \\
&= \sum_{j=1}^{m} c_{2m+1,j}(w) f^{*\prime}(w)^{2j-1} \Delta^{m+j}.
\end{aligned}
$$

Straightforward computations lead to the result for $m = 1$

$$
\begin{aligned}
\frac{\partial^2 \phi_{X_\Delta}(w)}{\partial w^2} &= \big(\lambda\Delta f^{*(2)}(w) + (\lambda\Delta f^{*\prime}(w))^2\big) \exp\big(\lambda\Delta(f^*(w)-1)\big) \\
\frac{\partial^3 \phi_{X_\Delta}(w)}{\partial w^3} &= \big(\lambda\Delta f^{*(3)}(w) + 2\lambda^2\Delta^2 f^{*\prime}(w) f^{*(2)}(w) + \lambda\Delta f^{*\prime}(w)(\lambda\Delta f^{*(2)}(w) \\
&\qquad + (\lambda\Delta f^{*\prime}(w))^2)\big) \times \exp\big(\lambda\Delta(f^*(w)-1)\big).
\end{aligned}
$$

Assume that the property holds at rank $m-1$, we have

$$
\begin{aligned}
\frac{\partial^{2m} \phi_{X_\Delta}(w)}{\partial w^{2m}} &= \frac{\partial}{\partial w} \frac{\partial^{2m-1} \phi_{X_\Delta}(w)}{\partial w^{2m-1}} \\
&= \big(\partial_w P_{2m-1}(w,\Delta) + \partial_w Q_{2m-1}(w,\Delta) + \lambda\Delta f^{*\prime}(x)(P_{2m-1}(w,\Delta) \\
&\qquad + Q_{2m-1}(w,\Delta))\big) \times \exp\big(\lambda\Delta(f^*(w)-1)\big)
\end{aligned}
$$

$$
\begin{aligned}
\text{where} \quad \partial_w P_{2m-1}(w,\Delta) &= c_{2m-1,1}(w)' f^{*\prime}(w)\Delta^m + c_{2m-1,1}(w) f^{*\prime\prime}(w) f^{*\prime}(w)\Delta^m \\
&\quad + \sum_{j=1}^{m-2} \big(c_{2m-1,j+1}(w)' f^{*\prime}(w)^{2j+1}\Delta^{m+j} \\
&\quad + c_{2m-1,j+1}(w)(2j+1) f^{*\prime\prime}(w) f^{*\prime}(w)^{2j}\Delta^{m+j}\big) \\
\lambda\Delta f^{*\prime}(x) P_{2m-1}(w,\Delta) &= \lambda \sum_{j=1}^{m-1} c_{2m-1,j}(w) f^{*\prime}(w)^{2j}\Delta^{m+j}.
\end{aligned}
$$

We set

$$
\begin{aligned}
P_{2m}(w,\Delta) &= \sum_{j=1}^{m-2} \big(c_{2m-1,j+1}(w)' f^{*\prime}(w) + c_{2m-1,j+1}(w)(2j+1) f^{*\prime\prime}(w)\big) \\
&\qquad \times f^{*\prime}(w)^{2j}\Delta^{m+j} \\
Q_{2m}(w,\Delta) &= \partial_w Q_{2m-1}(w,\Delta) + \lambda\Delta f^{*\prime}(x) Q_{2m-1}(w,\Delta)
\end{aligned}
$$

where P_{2m} have the desired property and from the property at rank $m-1$ the degree of Q_{2m} is lower than m. Similar computations give the result for P_{2m+1} and Q_{2m+1}. We complete on the proof with (29), $f^*(0)=1$ and using that f is centered: $f^{*\prime}(0)=0$. It follows that

$$\mathbb{E}\big[X_\Delta^{2m}\big]\leq \mathfrak{C}_1\Delta^m \qquad \text{and} \qquad \big|\mathbb{E}\big[X_\Delta^{2m+1}\big]\big|\leq \mathfrak{C}_2\Delta^m,$$

where $\mathfrak{C}_1$ and $\mathfrak{C}_2$ continuously depend on λ.

Proof of Lemma 6

We adopt the same methodology as for the proof of Theorem 3. Computations are quite similar we do not develop all of them. Each experiment is the $\lfloor T\Delta_T^{-1}\rfloor$-fold product of independent and identically distributed random variables the result is implied by (see Section 5.3 the proof of Lemma 1)

$$\big\|\mathbb{P}_\theta-\mathbb{Q}_\theta\big\|_{TV}=o\big((T/\Delta_T)^{-1}\big),$$

uniformly over the compact set Θ. Let us further denote by $p_{\Delta_T,\theta}$ and $q_{\Delta_T,\theta}$ the densities of Y_{Δ_T} and of V_{Δ_T} respectively, which can be decomposed as follows

$$p_{\Delta_T,\theta}(x)=e^{-\lambda\Delta_T}\delta_0\big(x-\tfrac{\lambda\Delta_T}{\beta}\big)+(1-e^{-\lambda\Delta_T})\widetilde{p}_{\Delta_T,\theta}(x) \tag{30}$$

$$q_{\Delta_T,\theta}(x)=e^{-\frac{8}{9}\lambda\Delta_T}\delta_0(x)+(1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}_{\Delta_T,\theta}(x) \tag{31}$$

where $\widetilde{p}_{\Delta_T,\theta}$ and $\widetilde{q}_{\Delta_T,\theta}$ are absolutely continuous with respect to the Lebesgue measure. For the reader convenience we set $\widetilde{p}:=\widetilde{p}_{\Delta_T,\theta}$ and $\widetilde{q}:=\widetilde{q}_{\Delta_T,\theta}$. Then, we have that $2\|\mathbb{P}_\theta-\mathbb{Q}_\theta\|_{TV}$ equals

$$\int_{\mathbb{R}}\big|(1-e^{-\lambda\Delta_T})\widetilde{p}(x)-(1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}(x)\big|dx+e^{-\frac{8}{9}\lambda\Delta_T}-e^{-\lambda\Delta_T},$$

where $e^{-\frac{8}{9}\lambda\Delta_T}-e^{-\lambda\Delta_T}$ is $o(\lfloor T\Delta_T^{-1}\rfloor^{-1})$ as Δ is of the order of T^α for $\alpha>0$. Applying successively the triangle inequality and Cauchy-Schwarz inequality we get

$$\int_{\mathbb{R}}\big|(1-e^{-\lambda\Delta_T})\widetilde{p}(x)-(1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}(x)\big|dx\leq I+II+III,$$

where for any $\eta>0$,

$$I=\sqrt{2\eta}\Big(\int_{\mathbb{R}}\big((1-e^{-\lambda\Delta_T})\widetilde{p}(x)-(1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}(x)\big)^2dx\Big)^{1/2},$$

$$II=\mathbb{P}_\theta\big(|Y_{\Delta_T}|\geq\eta\big) \qquad\qquad III=\mathbb{P}_\theta\big(|V_{\Delta_T}|\geq\eta\big).$$

Set $\eta_T=\kappa\sqrt{\Delta_T\log(T/\Delta_T)}$, we show that for $\kappa^2>3\lambda$, I, II and III are $o((T/\Delta_T^{-1}))$.

Bounding terms II and III. The argument used in the proof of Theorem 3 to bound the similar terms II and III also holds here. Then II and III are $o((T/\Delta_T)^{-1})$.

Bounding term I. We apply the Plancherel theorem to the integral in I, we denote by $\widetilde{p}^*$ and $\widetilde{q}^*$ the Fourier transforms of $\widetilde{p}$ and $\widetilde{q}$ respectively. They are computed with (30), (31) and the Lévy-Kintchine formula. We introduce the decomposition

$$\int_{\mathbb{R}} \big((1-e^{-\lambda\Delta_T})\widetilde{p}(x) - (1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}(x)\big)^2 dx \leq \; IV + V + VI,$$

with for any $\rho \geq 0$ and after replacing ξ by $\xi/\sqrt{\Delta_T}$

$$\begin{aligned}
IV &= \int_{|\xi|\leq\rho\sqrt{\Delta_T}} \big|(1-e^{-\lambda\Delta_T})\widetilde{p}^*\big(\tfrac{\xi}{\sqrt{\Delta_T}}\big) - (1-e^{-\frac{8}{9}\lambda\Delta_T})\widetilde{q}^*\big(\tfrac{\xi}{\sqrt{\Delta_T}}\big)\big|^2 \frac{d\xi}{\sqrt{\Delta_T}},\\
V &= \int_{|\xi|\geq\rho} \Big|\exp\Big(\lambda\Delta_T\big(\frac{1}{1-i\xi/\beta}-1\big)\Big) - e^{-\lambda\Delta_T}\Big|^2 d\xi,\\
VI &= \int_{|\xi|\geq\rho} \Big|\exp\Big(\tfrac{8}{9}\lambda\Delta_T\big(\frac{1}{1-i3\xi/(2\beta)}e^{-i3\xi/(2\beta)}-1\big)\Big) - e^{-\frac{8}{9}\lambda\Delta_T}\Big|^2 d\xi.
\end{aligned}$$

Bounding term IV. A first order expansion (see Remark 9) gives that IV is less than

$$\int_{|\xi|\leq\rho\sqrt{\Delta_T}} e^{-2\frac{\lambda\xi^2}{\beta^2}} \frac{\xi^8\alpha^2\big(\frac{\xi}{\sqrt{\Delta_T}}\big)}{\Delta_T^2} e^{2\frac{\xi^4}{\Delta_T}\alpha\big(\frac{\xi}{\sqrt{\Delta_T}}\big)} \frac{d\xi}{\sqrt{\Delta_T}} + 2\rho\sqrt{\Delta_T}(e^{-\frac{8}{9}\lambda\Delta_T} - e^{-\lambda\Delta_T})$$

for some bounded function $\xi \rightsquigarrow \alpha(\xi)$. Set $\overline{\alpha} = \sup_\xi |\alpha(\xi)|$, we obtain that IV is bounded by a constant times

$$\int_{\mathbb{R}} e^{-2(\lambda-\rho^2\overline{\alpha})\frac{\xi^2}{\beta^2}} \frac{\xi^8\overline{\alpha}^2}{\Delta_T^2}\frac{d\xi}{\sqrt{\Delta_T}}.$$

Choosing ρ such that $\lambda - \rho^2\overline{\alpha} > 0$, gives IV of order $\Delta_T^{-5/2}$.

Bounding terms V ad VI. Since

$$\begin{aligned}
V &= e^{-\lambda\Delta_T}\int_{|\xi|\geq\rho} \Big|\exp\big(\lambda\Delta_T/(1-\tfrac{i\xi}{\beta})\big) - 1\Big|^2 d\xi\\
VI &= e^{-\frac{8}{9}\lambda\Delta_T}\int_{|\xi|\geq\rho} \Big|\exp\big(\tfrac{8}{9}\lambda\Delta_T e^{-i3\xi/(2\beta)}/(1-\tfrac{i3\xi}{2\beta})\big) - 1\Big|^2 d\xi,
\end{aligned}$$

computations developed in the proof of Theorem 3 (to bound the analogous terms V and VI) holds for $C=\beta$ ($C = 16\beta/27$) for term V (for term VI), any $A > 2C$ and any $\rho > 0$ leading to $a = 1/\sqrt{(1+\frac{\rho^2}{\beta^2})} < 1$. We derive that V and VI are of the right order.

Completion of the proof of Lemma 6. Finally, $\int_{\mathbb{R}}(p_{\lambda,\Delta_T}(x) - q_{\lambda,\Delta_T}(x))^2dx$ is dominated by I which is in $\eta_T^{1/2}\Delta_T^{-5/4}$. The choice $\eta_T = \kappa\sqrt{\Delta_T\log(T/\Delta_T)}$ and the restriction condition $T/\Delta_T^2 = o((\log(T/\Delta_T))^{-1/4})$ imply $I = o((T/\Delta_T)^{-1})$. The proof is completed taking the supremum over Θ.

Acknowledgment

The author is thankful to Marc Hoffmann for his valuable remarks on this paper.

References

[1] Alexandersson, H. (1985). A Simple Stochastic mode of a Precipitation Process. *Journal of climate and applied meteorology*, **24**, 1285–1295.

[2] Athanassoulis, G.A. and Gavriliadis, P.N. (2002). The truncated Hausdorff moment problem solved by using kernel density functions. *Probabilistic Engineering Mechanics*, **17**(3), 273–291.

[3] Brown, L.D., Carter, A.V., Low, M.G. and Zhang, C-H. (2004). Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift. *The Annals of Statistics*, **32**, 2074–2097.

[4] Cont, R., and De Larrard, A. (2013). Price dynamics in a Markovian limit order market. *SIAM Journal on Financial Mathematics,* 4(1), 1–25.

[5] Duval, C. and Hoffmann, M. (2011) Statistical inference across time scales. *Electronic Journal of Statistics*, **5**, 2004–2030.

[6] P. Embrechts, C. Klüppelberg, M. Mikosch, Modeling Extremal Events, Springer, 1997.

[7] Gerber, H.U. and Shiu, E. (1998). Pricing perpetual options for jump processes. *The North American Actuarial Journal*, **2**, 101–112.

[8] F. Guilbaud, H. Pham, Optimal high-frequency trading in a pro-rata microstructure with predictive information, Arxiv preprint 1205.3051v1 (2012).

[9] Hong, K.J. and Satchell, S. (2012). Defining Single Asset Price Momentum in Terms of a Stochastic Process. *Theoretical Economics Letters*, **2**, 274–277.

[10] Kotulski, M. (1995). Asymptotic Distributions of the Continuous-Time Random Walks: A Probabilistic Approach. *Journal of Statistical Physics*, **81**, 777–792.

[11] Le Cam, L. and Yang, L.G. (2000) *Asymptotics in Statistics: Some Basic Concepts.* 2nd edition. Springer-Verlag, New York.

[12] Levy, J.B. and Taqqu, M.S. (2000). Renewal reward processes with heavy-tailed inter-renewal times and heavy-tailed rewards. *Bernoulli*, **6**, 23–44.

[13] Masoliver, J., Montero, M., Perelló, J. and Weiss, G.H. (2008). Direct and inverse problems with some generalizations and extensions. *Arxiv preprint,* 0308017v2.

[14] Metzler, R. and Klafter, J. (2004). The restaurant at the end of the random walk: recent developments in the description of anomalous transport by fractional dynamics. *Journal of Physics A,* **37**, 161–208.

[15] T. Mikosch, Non-life Insurance Mathematics: An Introduction with the Poisson Process, Springer, 2009.

[16] Önalan, Ö. (2010). Fractional Ornstein-Uhlenbeck Processes Driven by Stable Lévy Motion in Finance. *International Research Journal of Finance and Economics*, **42**, 129-139.

[17] Rodriguez, G. and Seatzu, S. (1992). On the solution of the finite moment problem. *Journal of Mathematical Analysis and Applications*, **171**(2), 321–333.

[18] Russell, J. R., and Engle, R. F. (2005). A discrete-state continuous-time model of financial transactions prices and times. Journal of Business & Economic Statistics, 23(2).

[19] Shevtsova, I. (2007). On the accuracy of the normal approximation to the distributions of Poisson random sums. PAMM, 7(1), 2080025-2080026.

[20] Sobezyk, K. and Trębicki, J. (1990). Maximum entropy principle in stochastic dynamics. *Probabilistic Engineering Mechanics*, **5**(3), 102–110.

[21] Tagliani, A. (1999). Hausdorff moment problem and maximum entropy: A unified approach. *Applied Mathematics and Computation*, **105**(2), 291–305.

[22] Tsybakov, A.B (2008). *Introduction to Nonparametric Estimation.* Springer.

[23] Uchaikin, V.V. and Zolotarev, V.M. (1999). *Chance and stability. Stable Distributions and their Applications.* VSP, Utrecht.

[24] Watson, G.N (1922). *A Treatise on the Theory of Bessel Functions.* Cambridge University Press.