Available online at www.sciencedirect.com

ScienceDirect

Stochastic Processes and their Applications 126 (2016) 503–541

stochastic processes and their applications

www.elsevier.com/locate/spa

Asymptotic equivalence for pure jump Lévy processes with unknown Lévy density and Gaussian white noise

Ester Mariucci

Laboratoire LJK, Université Joseph Fourier UMR 5224 51, Rue des Mathématiques, Saint Martin d'Hères BP 53 38041 Grenoble Cedex 09, France

Received 17 March 2015; received in revised form 7 September 2015; accepted 8 September 2015
Available online 16 September 2015

Abstract

The aim of this paper is to establish a global asymptotic equivalence between the experiments generated by the discrete (high frequency) or continuous observation of a path of a Lévy process and a Gaussian white noise experiment observed up to a time T, with T tending to ∞. These approximations are given in the sense of the Le Cam distance, under some smoothness conditions on the unknown Lévy density. All the asymptotic equivalences are established by constructing explicit Markov kernels that can be used to reproduce one experiment from the other.
© 2015 Elsevier B.V. All rights reserved.

MSC: 62B15; 62G20; 60G51

Keywords: Nonparametric experiments; Le Cam distance; Asymptotic equivalence; Lévy processes

1. Introduction

Lévy processes are a fundamental tool in modeling situations, like the dynamics of asset prices and weather measurements, where sudden changes in values may happen. For that reason they are widely employed, among many other fields, in mathematical finance. To name a simple example, the price of a commodity at time t is commonly given as an exponential function of a Lévy process. In general, exponential Lévy models are proposed for their ability to take into

E-mail address: Ester.Mariucci@imag.fr.

http://dx.doi.org/10.1016/j.spa.2015.09.009
0304-4149/© 2015 Elsevier B.V. All rights reserved.

account several empirical features observed in the returns of assets such as heavy tails, high-kurtosis and asymmetry (see [15] for an introduction to financial applications).

From a mathematical point of view, Lévy processes are a natural extension of the Brownian motion which preserves the tractable statistical properties of its increments, while relaxing the continuity of paths. The jump dynamics of a Lévy process is dictated by its Lévy density, say f. If f is continuous, its value at a point x_0 determines how frequent jumps of size close to x_0 are to occur per unit time. Concretely, if X is a pure jump Lévy process with Lévy density f, then the function f is such that

$$\int_A f(x)dx = \frac{1}{t}\mathbb{E}\left[\sum_{s\le t}\mathbb{I}_A(\Delta X_s)\right],$$

for any Borel set A and $t > 0$. Here, $\Delta X_s \equiv X_s - X_{s^-}$ denotes the magnitude of the jump of X at time s and $\mathbb{I}_A$ is the characteristic function. Thus, the Lévy measure

$$\nu(A) := \int_A f(x)dx,$$

is the average number of jumps (per unit time) whose magnitudes fall in the set A. Understanding the jumps behavior, therefore requires to estimate the Lévy measure. Several recent works have treated this problem, see e.g. [2] for an overview.

When the available data consists of the whole trajectory of the process during a time interval $[0, T]$, the problem of estimating f may be reduced to estimating the intensity function of an inhomogeneous Poisson process (see, e.g. [23,42]). However, a continuous-time sampling is never available in practice and thus the relevant problem is that of estimating f based on discrete sample data $X_{t_0}, \dots, X_{t_n}$ during a time interval $[0, T_n]$. In that case, the jumps are latent (unobservable) variables and that clearly adds to the difficulty of the problem. From now on we will place ourselves in a high-frequency setting, that is we assume that the sampling interval $\Delta_n = t_i - t_{i-1}$ tends to zero as n goes to infinity. Such a high-frequency based statistical approach has played a central role in the recent literature on nonparametric estimation for Lévy processes (see e.g. [22,13,14,1,19]). Moreover, in order to make consistent estimation possible, we will also ask the observation time T_n to tend to infinity in order to allow the identification of the jump part in the limit.

Our aim is to prove that, under suitable hypotheses, estimating the Lévy density f is equivalent to estimating the drift of an adequate Gaussian white noise model. In general, asymptotic equivalence results for statistical experiments provide a deeper understanding of statistical problems and allow to single out their main features. The idea is to pass via asymptotic equivalence to another experiment which is easier to analyze. By definition, two sequences of experiments $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$, defined on possibly different sample spaces, but with the same parameter set, are asymptotically equivalent if the Le Cam distance $\Delta(\mathscr{P}_{1,n}, \mathscr{P}_{2,n})$ tends to zero. For $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \big(P_{i,\theta} : \theta \in \Theta\big))$, $i = 1, 2$, $\Delta(\mathscr{P}_1, \mathscr{P}_2)$ is the symmetrization of the deficiency $\delta(\mathscr{P}_1, \mathscr{P}_2)$ where

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} \big\| K P_{1,\theta} - P_{2,\theta} \big\|_{TV}.$$

Here the infimum is taken over all randomizations from $(\mathscr{X}_1, \mathscr{A}_1)$ to $(\mathscr{X}_2, \mathscr{A}_2)$ and $\|\cdot\|_{TV}$ denotes the total variation distance. Roughly speaking, the Le Cam distance quantifies how much one fails to reconstruct (with the help of a randomization) a model from the other one and vice versa. Therefore, we say that $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ can be interpreted as "the models $\mathscr{P}_1$ and $\mathscr{P}_2$

contain the same amount of information about the parameter θ". The general definition of randomization is quite involved but, in the most frequent examples (namely when the sample spaces are Polish and the experiments dominated), it reduces to that of a Markov kernel. One of the most important features of the Le Cam distance is that it can be also interpreted in terms of statistical decision theory (see [32,33]; a short review is presented in the Appendix). As a consequence, saying that two statistical models are equivalent means that any statistical inference procedure can be transferred from one model to the other in such a way that the asymptotic risk remains the same, at least for bounded loss functions. Also, as soon as two models, $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$, that share the same parameter space Θ are proved to be asymptotically equivalent, the same result automatically holds for the restrictions of both $\mathscr{P}_{1,n}$ and $\mathscr{P}_{2,n}$ to a smaller subclass of Θ.

Historically, the first results of asymptotic equivalence in a nonparametric context date from 1996 and are due to [5,39]. The first two authors have shown the asymptotic equivalence of nonparametric regression and a Gaussian white noise model while the third one those of density estimation and white noise. Over the years many generalizations of these results have been proposed such as [3,28,43,9,11,41,12,37,46] for nonparametric regression or [10,31,4] for nonparametric density estimation models. Another very active field of study is that of diffusion experiments. The first result of equivalence between diffusion models and Euler scheme was established in 1998, see [38]. In later papers generalizations of this result have been considered (see [24,35]). Among others we can also cite equivalence results for generalized linear models [27], time series [29,38], diffusion models [18,25,17,16], GARCH model [7], functional linear regression [36], spectral density estimation [26] and volatility estimation [40]. Negative results are somewhat harder to come by; the most notable among them are [20,6,48]. There is however a lack of equivalence results concerning processes with jumps. A first result in this sense is [34] in which global asymptotic equivalences between the experiments generated by the discrete or continuous observation of a path of a Lévy process and a Gaussian white noise experiment are established. More precisely, in that paper, we have shown that estimating the drift function h from a continuously or discretely (high frequency) time inhomogeneous jump–diffusion process:

$$X_t = \int_0^t h(s)ds + \int_0^t \sigma(s)dW_s + \sum_{i=1}^{N_t} Y_i, \quad t \in [0, T_n], \tag{1}$$

is asymptotically equivalent to estimate h in the Gaussian model:

$$dy_t = h(t)dt + \sigma(t)dW_t, \quad t \in [0, T_n].$$

Here we try to push the analysis further and we focus on the case in which the considered parameter is the Lévy density and $X = (X_t)$ is a pure jump Lévy process (see [8] for the interest of such a class of processes when modeling asset returns). More in detail, we consider the problem of estimating the Lévy density (with respect to a fixed, possibly infinite, Lévy measure ν_0 concentrated on $I \subseteq \mathbb{R}$) $f := \frac{d\nu}{d\nu_0} : I \to \mathbb{R}$ from a continuously or discretely observed pure jump Lévy process X with possibly infinite Lévy measure. Here $I \subseteq \mathbb{R}$ denotes a possibly infinite interval and ν_0 is supposed to be absolutely continuous with respect to Lebesgue with a strictly positive density $g := \frac{d\nu_0}{d\mathrm{Leb}}$. In the case where ν is of finite variation one may write:

$$X_t = \sum_{0<s\leq t} \Delta X_s \tag{2}$$

or, equivalently, X has a characteristic function given by:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\bigg(-t\bigg(\int_I (1-e^{iuy})\nu(dy)\bigg)\bigg).$$

We suppose that the function f belongs to some a priori set $\mathscr{F}$, nonparametric in general. The discrete observations are of the form X_{t_i}, where $t_i = T_n\frac{i}{n}$, $i = 0,\dots,n$ with $T_n = n\Delta_n \to \infty$ and $\Delta_n \to 0$ as n goes to infinity. We will denote by $\mathscr{P}_n^{\nu_0}$ the statistical model associated with the continuous observation of a trajectory of X until time T_n (which is supposed to go to infinity as n goes to infinity) and by $\mathscr{Q}_n^{\nu_0}$ the one associated with the observation of the discrete data $(X_{t_i})_{i=0}^n$. The aim of this paper is to prove that, under adequate hypotheses on $\mathscr{F}$ (for example, f must be bounded away from zero and infinity; see Section 2.1 for a complete definition), the models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are both asymptotically equivalent to a sequence of Gaussian white noise models of the form:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}\frac{dW_t}{\sqrt{g(t)}}, \quad t \in I.$$

As a corollary, we then get the asymptotic equivalence between $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$. The main results are precisely stated as Theorems 2.5 and 2.6. A particular case of special interest arises when X is a compound Poisson process, $\nu_0 \equiv \mathrm{Leb}([0,1])$ and $\mathscr{F} \subseteq \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ where, for fixed $\gamma \in (0,1]$ and K, κ, M strictly positive constants, $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ is a class of continuously differentiable functions on I defined as follows:

$$\mathscr{F}^I_{(\gamma,K,\kappa,M)} = \Big\{ f : \kappa \le f(x) \le M,\ |f'(x) - f'(y)| \le K|x-y|^\gamma,\ \forall x, y \in I \Big\}. \tag{3}$$

In this case, the statistical models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are both equivalent to the Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}dW_t, \quad t \in [0,1].$$

See Example 3.1 for more details. By a theorem of Brown and Low in [5], we obtain, a posteriori, an asymptotic equivalence with the regression model

$$Y_i = \sqrt{f\Big(\frac{i}{T_n}\Big)} + \frac{1}{2\sqrt{T_n}}\xi_i, \quad \xi_i \sim \mathscr{N}(0,1),\ i = 1,\dots,[T_n].$$

Note that a similar form of a Gaussian shift was found to be asymptotically equivalent to a nonparametric density estimation experiment, see [39]. Let us mention that we also treat some explicit examples where ν_0 is neither finite nor compactly-supported (see Examples 3.2 and 3.3).

Without entering into any detail, we remark here that the methods are very different from those in [34]. In particular, since f belongs to the discontinuous part of a Lévy process, rather than its continuous part, the Girsanov-type changes of measure are irrelevant here. We thus need new instruments, like the Esscher changes of measure.

Our proof is based on the construction, for any given Lévy measure ν, of two adequate approximations $\hat{\nu}_m$ and $\bar{\nu}_m$ of ν: the idea of discretizing the Lévy density already appeared in an earlier work with P. Étoré and S. Louhichi, [21]. The present work is also inspired by the papers [10] (for a multinomial approximation), [4] (for passing from independent Poisson variables to independent normal random variables) and [34] (for a Bernoulli approximation).

This method allows us to construct explicit Markov kernels that lead from one model to the other; these may be applied in practice to transfer minimax estimators.

The paper is organized as follows: Sections 2.1 and 2.2 are devoted to make the parameter space and the considered statistical experiments precise. The main results are given in Section 2.3, followed by Section 3 in which some examples can be found. The proofs are postponed to Section 4. The paper includes Appendix recalling the definition and some useful properties of the Le Cam distance as well as of Lévy processes.

2. Assumptions and main results

2.1. The parameter space

Consider a (possibly infinite) Lévy measure ν_0 concentrated on a possibly infinite interval $I \subseteq \mathbb{R}$, admitting a density $g > 0$ with respect to Lebesgue. The parameter space of the experiments we are concerned with is a class of functions $\mathscr{F} = \mathscr{F}^{\nu_0, I}$ defined on I that form a class of Lévy densities with respect to ν_0: for each $f \in \mathscr{F}$, let ν (resp. $\hat{\nu}_m$) be the Lévy measure having f (resp. $\hat{f}_m$) as a density with respect to ν_0 where, for every $f \in \mathscr{F}$, $\hat{f}_m(x)$ is defined as follows.

Suppose first $x > 0$. Given a positive integer depending on n, $m = m_n$, let $J_j := (v_{j-1}, v_j]$ where $v_1 = \varepsilon_m \geq 0$ and v_j are chosen in such a way that

$$\mu_m := \nu_0(J_j) = \frac{\nu_0\big((I\backslash[0, \varepsilon_m]) \cap \mathbb{R}_+\big)}{m-1}, \quad \forall j = 2, \dots, m. \tag{4}$$

In the sequel, for the sake of brevity, we will only write m without making explicit the dependence on n. Define $x_j^* := \frac{\int_{J_j} x\nu_0(dx)}{\mu_m}$ and introduce a sequence of functions $0 \leq V_j \leq \frac{1}{\mu_m}$, $j = 2, \dots, m$ supported on $[x_{j-1}^*, x_{j+1}^*]$ if $j = 3, \dots, m-1$, on $[\varepsilon_m, x_3^*]$ if $j = 2$ and on $(I\backslash[0, x_{m-1}^*]) \cap \mathbb{R}_+$ if $j = m$. The V_j's are defined recursively in the following way.

- V_2 is equal to $\frac{1}{\mu_m}$ on the interval $(\varepsilon_m, x_2^*]$ and on the interval $(x_2^*, x_3^*]$ it is chosen so that it is continuous (in particular, $V_2(x_2^*) = \frac{1}{\mu_m}$), $\int_{x_2^*}^{x_3^*} V_2(y)\nu_0(dy) = \frac{\nu_0((x_2^*, v_2])}{\mu_m}$ and $V_2(x_3^*) = 0$.
- For $j = 3, \dots, m-1$ define V_j as the function $\frac{1}{\mu_m} - V_{j-1}$ on the interval $[x_{j-1}^*, x_j^*]$. On $[x_j^*, x_{j+1}^*]$ choose V_j continuous and such that $\int_{x_j^*}^{x_{j+1}^*} V_j(y)\nu_0(dy) = \frac{\nu_0((x_j^*, v_j])}{\mu_m}$ and $V_j(x_{j+1}^*) = 0$.
- Finally, let V_m be the function supported on $(I\backslash[0, x_{m-1}^*]) \cap \mathbb{R}_+$ such that

$$V_m(x) = \frac{1}{\mu_m} - V_{m-1}(x), \quad \text{for } x \in [x_{m-1}^*, x_m^*],$$
$$V_m(x) = \frac{1}{\mu_m}, \quad \text{for } x \in (I\backslash[0, x_m^*]) \cap \mathbb{R}_+.$$

(It is immediate to check that such a choice is always possible.) Observe that, by construction,

$$\sum_{j=2}^{m} V_j(x)\mu_m = 1, \quad \forall x \in (I\backslash[0, \varepsilon_m]) \cap \mathbb{R}_+ \quad \text{and} \quad \int_{(I\backslash[0,\varepsilon_m])\cap\mathbb{R}_+} V_j(y)\nu_0(dy) = 1.$$

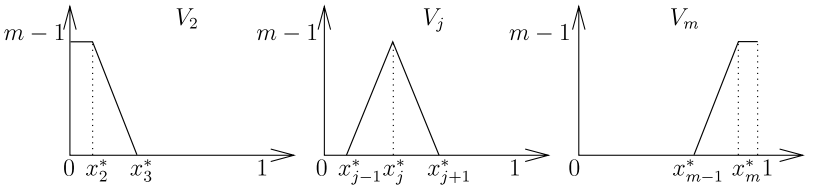
Analogously, define $\mu_m^- = \frac{\nu_0\left((I\setminus[-\varepsilon_m,0])\cap\mathbb{R}_-\right)}{m-1}$ and $J_{-m},\dots,J_{-2}$ such that $\nu_0(J_{-j}) = \mu_m^-$ for all j. Then, for $x < 0$, x^*_{-j} is defined as x_j^* by using J_{-j} and μ_m^- instead of J_j and μ_m and the V_{-j}'s are defined with the same procedure as the V_j's, starting from V_{-2} and proceeding by induction.

Define

$$\hat{f}_m(x) = \mathbb{I}_{[-\varepsilon_m,\varepsilon_m]}(x) + \sum_{j=2}^{m}\left(V_j(x)\int_{J_j} f(y)\nu_0(dy) + V_{-j}(x)\int_{J_{-j}} f(y)\nu_0(dy)\right). \tag{5}$$

The definitions of the V_j's above are modeled on the following example:

Example 2.1. Let ν_0 be the Lebesgue measure on $[0,1]$ and $\varepsilon_m = 0$. Then $v_j = \frac{j-1}{m-1}$ and $x_j^* = \frac{2j-3}{2m-2}$, $j = 2,\dots,m$. The standard choice for V_j (based on the construction by [10]) is given by the piecewise linear functions interpolating the values in the points x_j^* specified above:



Remark 2.2. The function $\hat{f}_m$ has been defined in such a way that the rate of convergence of the L_2 norm between the restriction of f and $\hat{f}_m$ on $I\setminus[-\varepsilon_m,\varepsilon_m]$ is compatible with the rate of convergence of the other quantities appearing in the statements of Theorems 2.5 and 2.6. For that reason, as in [10], we have not chosen a piecewise constant approximation of f but an approximation that is, at least in the simplest cases, a piecewise linear approximation of f. Such a choice allows us to gain an order of magnitude on the convergence rate of $\|f - \hat{f}_m\|_{L_2(\nu_0|I\setminus[-\varepsilon_m,\varepsilon_m])}$ at least when $\mathscr{F}$ is a class of sufficiently smooth functions.

We now explain the assumptions we will need to make on the parameter $f \in \mathscr{F} = \mathscr{F}^{\nu_0,I}$. The superscripts ν_0 and I will be suppressed whenever this can lead to no confusion. We require that:

(H1) There exist constants $\kappa, M > 0$ such that $\kappa \le f(y) \le M$, for all $y \in I$ and $f \in \mathscr{F}$.

For every integer $m = m_n$, we can consider $\widehat{\sqrt{f}}_m$, the approximation of $\sqrt{f}$ constructed as $\hat{f}_m$ above, i.e. $\widehat{\sqrt{f}}_m(x) = \mathbb{I}_{[-\varepsilon_m,\varepsilon_m]}(x) + \sum_{\substack{j=-m,\dots,m\\ j\neq -1,0,1.}} V_j(x)\int_{J_j}\sqrt{f(y)}\nu_0(dy)$, and introduce the quantities:

$$A_m^2(f) := \int_{I\setminus[-\varepsilon_m,\varepsilon_m]}\left(\widehat{\sqrt{f}}_m(y) - \sqrt{f(y)}\right)^2\nu_0(dy),$$

$$B_m^2(f) := \sum_{\substack{j=-m,\dots,m\\ j\neq -1,0,1.}}\left(\int_{J_j}\frac{\sqrt{f(y)}}{\sqrt{\nu_0(J_j)}}\nu_0(dy) - \sqrt{\nu(J_j)}\right)^2,$$

$$C_m^2(f) := \int_{-\varepsilon_m}^{\varepsilon_m}\left(\sqrt{f(t)} - 1\right)^2\nu_0(dt).$$

The conditions defining the parameter space $\mathscr{F}$ are expressed by asking that the quantities introduced above converge quickly enough to zero. To state the assumptions of Theorem 2.5 precisely, we will assume the existence of sequences of discretizations $m = m_n \to \infty$, of positive numbers $\varepsilon_m = \varepsilon_{m_n} \to 0$ and of functions V_j, $j = \pm 2, \dots, \pm m$, such that:

(C1) $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \int_{I\setminus(-\varepsilon_m,\varepsilon_m)} \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx) = 0.$

(C2) $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \big(A_m^2(f) + B_m^2(f) + C_m^2(f)\big) = 0.$

Remark in particular that Condition (C2) implies the following:

(H2) $\sup_{f\in\mathscr{F}} \int_I (\sqrt{f(y)} - 1)^2 \nu_0(dy) \le L,$

where $L = \sup_{f\in\mathscr{F}} \int_{-\varepsilon_m}^{\varepsilon_m} (\sqrt{f(x)} - 1)^2 \nu_0(dx) + (\sqrt{M} + 1)^2 \nu_0\big(I\setminus(-\varepsilon_m, \varepsilon_m)\big)$, for any choice of m such that the quantity in the limit appearing in Condition (C2) is finite.

Theorem 2.6 has slightly stronger hypotheses, defining possibly smaller parameter spaces: we will assume the existence of sequences m_n, ε_m and V_j, $j = \pm 2, \dots, \pm m$ (possibly different from the ones above) such that Condition (C1) is verified and the following stronger version of Condition (C2) holds:

(C2′) $\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}} \big(A_m^2(f) + B_m^2(f) + nC_m^2(f)\big) = 0.$

Finally, some of our results have a more explicit statement under the hypothesis of finite variation which we state as:

(FV) $\int_I (|x| \wedge 1)\nu_0(dx) < \infty.$

Remark 2.3. The Condition (C1) and those involving the quantities $A_m(f)$ and $B_m(f)$ all concern similar but slightly different approximations of f. In concrete examples, they may all be expected to have the same rate of convergence but to keep the greatest generality we preferred to state them separately. On the other hand, conditions on the quantity $C_m(f)$ are purely local around zero, requiring, for each $f \in \mathscr{F}$, that $f(x)$ tends to 1 quickly enough as x tends to 0.

Examples 2.4. To get a grasp on Conditions (C1), (C2) we analyze here three different examples according to the different behavior of ν_0 near $0 \in I$. In all of these cases the parameter space $\mathscr{F}^{\nu_0, I}$ will be a subclass of $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ defined as in (3). Recall that the conditions (C1), (C2) and (C2′) depend on the choice of sequences m_n, ε_m and functions V_j. For the first two of the three examples, where $I = [0, 1]$, we will make the standard choice for V_j of triangular and trapezoidal functions, similarly to those in Example 2.1. Namely, for $j = 3, \dots, m-1$ we have

$$V_j(x) = \mathbb{I}_{(x^*_{j-1}, x^*_j]}(x) \frac{x - x^*_{j-1}}{x^*_j - x^*_{j-1}} \frac{1}{\mu_m} + \mathbb{I}_{(x^*_j, x^*_{j+1}]}(x) \frac{x^*_{j+1} - x}{x^*_{j+1} - x^*_j} \frac{1}{\mu_m}; \tag{6}$$

the two extremal functions V_2 and V_m are chosen so that $V_2 \equiv \frac{1}{\mu_m}$ on $(\varepsilon_m, x^*_2]$ and $V_m \equiv \frac{1}{\mu_m}$ on $(x^*_m, 1]$. In the second example, where ν_0 is infinite, one is forced to take $\varepsilon_m > 0$ and to keep in mind that the x^*_j are not uniformly distributed on $[\varepsilon_m, 1]$. Proofs of all the statements here can be found in Section 5.2.

1. The finite case: $\nu_0 \equiv \mathrm{Leb}([0, 1])$.

In this case we are free to choose $\mathscr{F}^{\mathrm{Leb},[0,1]} = \mathscr{F}^{[0,1]}_{(\gamma,K,\kappa,M)}$. Indeed, as ν_0 is finite, there is no need to single out the first interval $J_1 = [0, \varepsilon_m]$, so that $C_m(f)$ does not enter in the proofs and the definitions of $A_m(f)$ and $B_m(f)$ involve integrals on the whole of $[0, 1]$. Also, the choice of

the V_j's as in (6) guarantees that $\int_0^1 V_j(x)dx = 1$. Then, the quantities $\|f - \hat{f}_m\|_{L_2([0,1])}$, $A_m(f)$ and $B_m(f)$ all have the same rate of convergence, which is given by:

$$\sqrt{\int_0^1 \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx)} + A_m(f) + B_m(f) = O\Big(m^{-\gamma-1} + m^{-\frac{3}{2}}\Big),$$

uniformly on f. See Section 5.2 for a proof.

2. The finite variation case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-1}\mathbb{I}_{[0,1]}(x)$.

In this case, the parameter space $\mathscr{F}^{\nu_0,[0,1]}$ is a proper subset of $\mathscr{F}^{[0,1]}_{(\gamma,K,\kappa,M)}$. Indeed, as we are obliged to choose $\varepsilon_m > 0$, we also need to impose that $C_m(f) = o\big(\frac{1}{n\sqrt{\Delta_n}}\big)$, with uniform constants with respect to f, that is, that all $f \in \mathscr{F}$ converge to 1 quickly enough as $x \to 0$. Choosing $\varepsilon_m = m^{-1-\alpha}$, $\alpha > 0$ we have that $\mu_m = \frac{\ln(\varepsilon_m^{-1})}{m-1}$, $v_j = \varepsilon_m^{\frac{m-j}{m-1}}$ and $x_j^* = \frac{(v_j - v_{j-1})}{\mu_m}$. In particular, $\max_j |v_{j-1} - v_j| = |v_m - v_{m-1}| = O\Big(\frac{\ln m}{m}\Big)$. Also in this case one can prove that the standard choice of V_j described above leads to $\int_{\varepsilon_m}^1 V_j(x)\frac{dx}{x} = 1$. Again, the quantities $\|f - \hat{f}_m\|_{L_2(\nu_0|I\setminus[0,\varepsilon_m])}$, $A_m(f)$ and $B_m(f)$ have the same rate of convergence given by:

$$\sqrt{\int_{\varepsilon_m}^1 \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx)} + A_m(f) + B_m(f) = O\left(\left(\frac{\ln m}{m}\right)^{\gamma+1}\sqrt{\ln(\varepsilon_m^{-1})}\right), \tag{7}$$

uniformly on f. The condition on $C_m(f)$ depends on the behavior of f near 0. For example, it is ensured if one considers a parametric family of the form $f(x) = e^{-\lambda x}$ with a bounded $\lambda > 0$. See Section 5.2 for a proof.

3. The infinite variation, non-compactly supported case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-2}\mathbb{I}_{\mathbb{R}_+}(x)$.

This example involves significantly more computations than the preceding ones, since the classical triangular choice for the functions V_j would not have integral equal to 1 (with respect to ν_0), and the support is not compact. The parameter space $\mathscr{F}^{\nu_0,[0,\infty)}$ can still be chosen as a proper subclass of $\mathscr{F}^{[0,\infty)}_{(\gamma,K,\kappa,M)}$, again by imposing that $C_m(f)$ converges to zero quickly enough (more details about this condition are discussed in Example 3.3). We divide the interval $[0,\infty)$ in m intervals $J_j = [v_{j-1}, v_j)$ with:

$$v_0 = 0; \qquad v_1 = \varepsilon_m; \qquad v_j = \frac{\varepsilon_m(m-1)}{m-j}; \qquad v_m = \infty; \qquad \mu_m = \frac{1}{\varepsilon_m(m-1)}.$$

To deal with the non-compactness problem, we choose some "horizon" $H(m)$ that goes to infinity slowly enough as m goes to infinity and we bound the L_2 distance between f and $\hat{f}_m$ for $x > H(m)$ by $2\sup_{x\geq H(m)}\frac{f(x)^2}{H(m)}$. We have:

$$\|f - \hat{f}_m\|^2_{L_2(\nu_0|I\setminus[0,\varepsilon_m])} + A_m^2(f) + B_m^2(f) = O\left(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}} + \sup_{x\geq H(m)}\frac{f(x)^2}{H(m)}\right).$$

In the general case where the best estimate for $\sup_{x\geq H(m)} f(x)^2$ is simply given by M^2, an optimal choice for $H(m)$ is $\sqrt{\varepsilon_m m}$, that gives a rate of convergence:

$$\|f - \hat{f}_m\|^2_{L_2(\nu_0|I\setminus[0,\varepsilon_m])} + A_m^2(f) + B_m^2(f) = O\left(\frac{1}{\sqrt{\varepsilon_m m}}\right),$$

independently of γ. See Section 5.2 for a proof.

2.2. Definition of the experiments

Let $(x_t)_{t\geq 0}$ be the canonical process on the Skorokhod space $(D, \mathscr{D})$ and denote by $P^{(b,0,\nu)}$ the law induced on $(D, \mathscr{D})$ by a Lévy process with characteristic triplet $(b, 0, \nu)$. We will write $P_t^{(b,0,\nu)}$ for the restriction of $P^{(b,0,\nu)}$ to the σ-algebra $\mathscr{D}_t$ generated by $\{x_s : 0 \leq s \leq t\}$ (see Appendix A.2 for the precise definitions). Let $Q_t^{(b,0,\nu)}$ be the marginal law at time t of a Lévy process with characteristic triplet $(b, 0, \nu)$. In the case where $\int_{|y|\leq 1} |y|\nu(dy) < \infty$ we introduce the notation $\gamma^\nu := \int_{|y|\leq 1} y\nu(dy)$; then, Condition (H2) guarantees the finiteness of $\gamma^{\nu-\nu_0}$ (see Remark 33.3 in [44] for more details).

Recall that we introduced the discretization $t_i = T_n \frac{i}{n}$ of $[0, T_n]$ and denote by $\mathbf{Q}_n^{(\gamma^{\nu-\nu_0},0,\nu)}$ the laws of the $n+1$ marginals of $(x_t)_{t\geq 0}$ at times t_i, $i = 0, \ldots, n$. We will consider the following statistical models, depending on a fixed, possibly infinite, Lévy measure ν_0 concentrated on I (clearly, the models with the subscript FV are meaningful only under the assumption (FV)):

$$\mathscr{P}_{n,FV}^{\nu_0} = \left(D, \mathscr{D}_{T_n}, \left\{P_{T_n}^{(\gamma^\nu,0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\right\}\right),$$

$$\mathscr{Q}_{n,FV}^{\nu_0} = \left(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}), \left\{\mathbf{Q}_n^{(\gamma^\nu,0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\right\}\right),$$

$$\mathscr{P}_n^{\nu_0} = \left(D, \mathscr{D}_{T_n}, \left\{P_{T_n}^{(\gamma^{\nu-\nu_0},0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\right\}\right),$$

$$\mathscr{Q}_n^{\nu_0} = \left(\mathbb{R}^{n+1}, \mathscr{B}(\mathbb{R}^{n+1}), \left\{\mathbf{Q}_n^{(\gamma^{\nu-\nu_0},0,\nu)} : f := \frac{d\nu}{d\nu_0} \in \mathscr{F}^{\nu_0,I}\right\}\right).$$

Finally, let us introduce the Gaussian white noise model that will appear in the statement of our main results. For that, let us denote by $(C(I), \mathscr{C})$ the space of continuous mappings from I into $\mathbb{R}$ endowed with its standard filtration, by g the density of ν_0 with respect to the Lebesgue measure. We will require $g > 0$ and let $\mathbb{W}_n^f$ be the law induced on $(C(I), \mathscr{C})$ by the stochastic process satisfying:

$$dy_t = \sqrt{f(t)}dt + \frac{dW_t}{2\sqrt{T_n}\sqrt{g(t)}}, \quad t \in I, \tag{8}$$

where $(W_t)_{t\in\mathbb{R}}$ denotes a Brownian motion on $\mathbb{R}$ with $W_0 = 0$. Then we set:

$$\mathscr{W}_n^{\nu_0} = \left(C(I), \mathscr{C}, \{\mathbb{W}_n^f : f \in \mathscr{F}^{\nu_0,I}\}\right).$$

Observe that when ν_0 is a finite Lévy measure, then $\mathscr{W}_n^{\nu_0}$ is equivalent to the statistical model associated with the continuous observation of a process $(\tilde{y}_t)_{t\in I}$ defined by:

$$d\tilde{y}_t = \sqrt{f(t)g(t)}dt + \frac{dW_t}{2\sqrt{T_n}}, \quad t \in I.$$

2.3. Main results

Using the notation introduced in Section 2.1, we now state our main results. For brevity of notation, we will denote by $H(f, \hat{f}_m)$ (resp. $L_2(f, \hat{f}_m)$) the Hellinger distance (resp. the L_2

distance) between the Lévy measures ν and $\hat{\nu}_m$ restricted to $I\backslash[-\varepsilon_m, \varepsilon_m]$, i.e.:

$$H^2(f, \hat{f}_m) := \int_{I\backslash[-\varepsilon_m,\varepsilon_m]} \left(\sqrt{f(x)} - \sqrt{\hat{f}_m(x)}\right)^2 \nu_0(dx),$$
$$L_2(f, \hat{f}_m)^2 := \int_{I\backslash[-\varepsilon_m,\varepsilon_m]} \left(f(y) - \hat{f}_m(y)\right)^2 \nu_0(dy).$$

Observe that Condition (H1) implies (see Lemma 5.1)

$$\frac{1}{4M} L_2(f, \hat{f}_m)^2 \le H^2(f, \hat{f}_m) \le \frac{1}{4\kappa} L_2(f, \hat{f}_m)^2.$$

Theorem 2.5. *Let ν_0 be a known Lévy measure concentrated on a (possibly infinite) interval $I \subseteq \mathbb{R}$ and having strictly positive density with respect to the Lebesgue measure. Let us choose a parameter space $\mathscr{F}^{\nu_0, I}$ such that there exist a sequence $m = m_n$ of integers, functions V_j, $j = \pm 2, \dots, \pm m$ and a sequence $\varepsilon_m \to 0$ as $m \to \infty$ such that Conditions* (H1), (C1), (C2) *are satisfied for $\mathscr{F} = \mathscr{F}^{\nu_0, I}$. Then, for n big enough we have:*

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\left(\sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} \left(A_m(f) + B_m(f) + C_m(f)\right)\right) + O\left(\sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} L_2(f, \hat{f}_m) + \sqrt{\frac{m}{n\Delta_n}\left(\frac{1}{\mu_m} + \frac{1}{\mu_{\bar{m}}}\right)}\right). \tag{9}$$

Theorem 2.6. *Let ν_0 be a known Lévy measure concentrated on a (possibly infinite) interval $I \subseteq \mathbb{R}$ and having strictly positive density with respect to the Lebesgue measure. Let us choose a parameter space $\mathscr{F}^{\nu_0, I}$ such that there exist a sequence $m = m_n$ of integers, functions V_j, $j = \pm 2, \dots, \pm m$ and a sequence $\varepsilon_m \to 0$ as $m \to \infty$ such that Conditions* (H1), (C1), *(C2′) are satisfied for $\mathscr{F} = \mathscr{F}^{\nu_0, I}$. Then, for n big enough we have:*

$$\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\left(\nu_0\left(I\backslash[-\varepsilon_m, \varepsilon_m]\right)\sqrt{n\Delta_n^2} + \frac{m\ln m}{\sqrt{n}} + \sqrt{n\sqrt{\Delta_n} \sup_{f\in\mathscr{F}} C_m(f)}\right) + O\left(\sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} \left(A_m(f) + B_m(f) + H(f, \hat{f}_m)\right)\right). \tag{10}$$

Corollary 2.7. *Let ν_0 be as above and let us choose a parameter space $\mathscr{F}^{\nu_0, I}$ so that there exist sequences m_n', ε_m', V_j' and m_n'', ε_m'', V_j'' such that:*

- *Conditions* (H1), (C1) *and* (C2) *hold for m_n', ε_m', V_j', and $\frac{m'}{n\Delta_n}\left(\frac{1}{\mu_{m'}} + \frac{1}{\mu_{\bar{m}'}}\right)$ tends to zero.*
- *Conditions* (H1), (C1) *and* (C2′) *hold for m_n'', ε_m'', V_j'', and $\nu_0\left(I\backslash[-\varepsilon_{m''}, \varepsilon_{m''}]\right)\sqrt{n\Delta_n^2} + \frac{m''\ln m''}{\sqrt{n}}$ tends to zero.*

Then the statistical models $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ are asymptotically equivalent:

$$\lim_{n\to\infty} \Delta(\mathscr{P}_n^{\nu_0}, \mathscr{Q}_n^{\nu_0}) = 0.$$

If, in addition, the Lévy measures have finite variation, i.e. if we assume (FV), then the same results hold replacing $\mathscr{P}_n^{\nu_0}$ and $\mathscr{Q}_n^{\nu_0}$ by $\mathscr{P}_{n,FV}^{\nu_0}$ and $\mathscr{Q}_{n,FV}^{\nu_0}$, respectively (see Lemma A.14).

3. Examples

We will now analyze three different examples, underlining the different behaviors of the Lévy measure ν_0 (respectively, finite, infinite with finite variation and infinite with infinite variation). The three chosen Lévy measures are $\mathbb{I}_{[0,1]}(x)dx$, $\mathbb{I}_{[0,1]}(x)\frac{dx}{x}$ and $\mathbb{I}_{\mathbb{R}_+}(x)\frac{dx}{x^2}$. In all three cases we assume the parameter f to be uniformly bounded and with uniformly γ-Hölder derivatives: we will describe adequate subclasses $\mathscr{F}^{\nu_0,I} \subseteq \mathscr{F}^{I}_{(\gamma,K,\kappa,M)}$ defined as in (3). It seems very likely that the same results that are highlighted in these examples hold true for more general Lévy measures; however, we limit ourselves to these examples in order to be able to explicitly compute the quantities involved (v_j, x_j^*, etc.) and hence estimate the distance between f and $\hat{f}_m$ as in Examples 2.4.

In the first of the three examples, where ν_0 is the Lebesgue measure on $I = [0,1]$, we are considering the statistical models associated with the discrete and continuous observation of a compound Poisson process with Lévy density f. Observe that $\mathscr{W}_n^{\mathrm{Leb}}$ reduces to the statistical model associated with the continuous observation of a trajectory from:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}}dW_t, \quad t \in [0,1].$$

In this case we have:

Example 3.1 (*Finite Lévy Measure*)**.** Let ν_0 be the Lebesgue measure on $I = [0,1]$ and let $\mathscr{F} = \mathscr{F}^{\mathrm{Leb},[0,1]}$ be any subclass of $\mathscr{F}^{[0,1]}_{(\gamma,K,\kappa,M)}$ for some strictly positive constants K, κ, M and $\gamma \in (0,1]$. Then:

$$\lim_{n\to\infty} \Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}^{\mathrm{Leb}}_n) = 0 \quad \text{and} \quad \lim_{n\to\infty} \Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}^{\mathrm{Leb}}_n) = 0.$$

More precisely,

$$\Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}^{\mathrm{Leb}}_n) = \begin{cases} O\left((n\Delta_n)^{-\frac{\gamma}{4+2\gamma}}\right) & \text{if } \gamma \in \left(0, \frac{1}{2}\right], \\ O\left((n\Delta_n)^{-\frac{1}{10}}\right) & \text{if } \gamma \in \left(\frac{1}{2}, 1\right]. \end{cases}$$

In the case where $\Delta_n = n^{-\beta}$, $\frac{1}{2} < \beta < 1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}^{\mathrm{Leb}}_n)$ is

$$\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}^{\mathrm{Leb}}_n) = \begin{cases} O\left(n^{-\frac{\gamma+\beta}{4+2\gamma}} \ln n\right) & \text{if } \gamma \in \left(0, \frac{1}{2}\right) \text{ and } \frac{2+2\gamma}{3+2\gamma} \le \beta < 1, \\ O\left(n^{\frac{1}{2}-\beta} \ln n\right) & \text{if } \gamma \in \left(0, \frac{1}{2}\right) \text{ and } \frac{1}{2} < \beta < \frac{2+2\gamma}{3+2\gamma}, \\ O\left(n^{-\frac{2\beta+1}{10}} \ln n\right) & \text{if } \gamma \in \left[\frac{1}{2}, 1\right] \text{ and } \frac{3}{4} \le \beta < 1, \\ O\left(n^{\frac{1}{2}-\beta} \ln n\right) & \text{if } \gamma \in \left[\frac{1}{2}, 1\right] \text{ and } \frac{1}{2} < \beta < \frac{3}{4}. \end{cases}$$

See Section 5.3 for a proof.

Example 3.2 (*Infinite Lévy Measure with Finite Variation*)**.** Let X be a truncated Gamma process with (infinite) Lévy measure of the form:

$$\nu(A) = \int_A \frac{e^{-\lambda x}}{x} dx, \quad A \in \mathscr{B}([0,1]).$$

Here $\mathscr{F}^{\nu_0,I}$ is a 1-dimensional parametric family in λ, assuming that there exists a known constant λ_0 such that $0 < \lambda \leq \lambda_0 < \infty$, $f(t) = e^{-\lambda t}$ and $d\nu_0(x) = \frac{1}{x}dx$. In particular, f is Lipschitz, i.e. $\mathscr{F}^{\nu_0,[0,1]} \subset \mathscr{F}^{[0,1]}_{(\gamma=1,K,\kappa,M)}$. The discrete or continuous observations (up to time T_n) of X are asymptotically equivalent to $\mathscr{W}_n^{\nu_0}$, the statistical model associated with the observation of a trajectory of the process (y_t):

$$dy_t = \sqrt{f(t)}dt + \frac{\sqrt{t}dW_t}{2\sqrt{T_n}}, \quad t \in [0,1].$$

More precisely, in the case where $\Delta_n = n^{-\beta}$, $\frac{1}{2} < \beta < 1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}^{\nu_0}_{n,FV}, \mathscr{W}_n^{\nu_0})$ is

$$\Delta(\mathscr{Q}^{\nu_0}_{n,FV}, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\left(n^{\frac{1}{2}-\beta}\ln n\right) & \text{if } \dfrac{1}{2} < \beta \leq \dfrac{9}{10} \\ O\left(n^{-\frac{1+2\beta}{7}}\ln n\right) & \text{if } \dfrac{9}{10} < \beta < 1. \end{cases}$$

Concerning the continuous setting we have:

$$\Delta(\mathscr{P}^{\nu_0}_{n,FV}, \mathscr{W}_n^{\nu_0}) = O\left(n^{\frac{\beta-1}{6}}(\ln n)^{\frac{5}{2}}\right) = O\left(T_n^{-\frac{1}{6}}(\ln T_n)^{\frac{5}{2}}\right).$$

See Section 5.4 for a proof.

Example 3.3 (*Infinite Lévy Measure, Infinite Variation*)**.** Let X be a pure jump Lévy process with infinite Lévy measure of the form:

$$\nu(A) = \int_A \frac{2 - e^{-\lambda x^3}}{x^2}dx, \quad A \in \mathscr{B}(\mathbb{R}^+).$$

Again, we are considering a parametric family in $\lambda > 0$, assuming that the parameter stays bounded below a known constant λ_0. Here, $f(t) = 2 - e^{-\lambda t^3}$, hence $1 \leq f(t) \leq 2$, for all $t \geq 0$, and f is Lipschitz, i.e. $\mathscr{F}^{\nu_0,\mathbb{R}_+} \subset \mathscr{F}^{\mathbb{R}_+}_{(\gamma=1,K,\kappa,M)}$. The discrete or continuous observations (up to time T_n) of X are asymptotically equivalent to the statistical model associated with the observation of a trajectory of the process (y_t):

$$dy_t = \sqrt{f(t)}dt + \frac{tdW_t}{2\sqrt{T_n}}, \quad t \geq 0.$$

More precisely, in the case where $\Delta_n = n^{-\beta}$, $0 < \beta < 1$, an upper bound for the rate of convergence of $\Delta(\mathscr{Q}^{\nu_0}_n, \mathscr{W}_n^{\nu_0})$ is

$$\Delta(\mathscr{Q}^{\nu_0}_n, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\left(n^{\frac{1}{2}-\frac{2}{3}\beta}\right) & \text{if } \dfrac{3}{4} < \beta < \dfrac{12}{13} \\ O\left(n^{-\frac{1}{6}+\frac{\beta}{18}}(\ln n)^{\frac{7}{6}}\right) & \text{if } \dfrac{12}{13} \leq \beta < 1. \end{cases}$$

In the continuous setting, we have

$$\Delta(\mathscr{P}^{\nu_0}_n, \mathscr{W}_n^{\nu_0}) = O\left(n^{\frac{3\beta-3}{34}}(\ln n)^{\frac{7}{6}}\right) = O\left(T_n^{-\frac{3}{34}}(\ln T_n)^{\frac{7}{6}}\right).$$

See Section 5.5 for a proof.

4. Proofs of the main results

In order to simplify notations, the proofs will be presented in the case $I \subseteq \mathbb{R}^+$. Nevertheless, this allows us to present all the main difficulties, since they can only appear near 0. To prove Theorems 2.5 and 2.6 we need to introduce several intermediate statistical models. In that regard, let us denote by Q_j^f the law of a Poisson random variable with mean $T_n\nu(J_j)$ (see (4) for the definition of J_j). We will denote by $\mathscr{L}_m$ the statistical model associated with the family of probabilities $\{\bigotimes_{j=2}^m Q_j^f : f \in \mathscr{F}\}$:

$$\mathscr{L}_m = \left(\bar{\mathbb{N}}^{m-1}, \mathcal{P}(\bar{\mathbb{N}}^{m-1}), \left\{\bigotimes_{j=2}^m Q_j^f : f \in \mathscr{F}\right\}\right). \tag{11}$$

By N_j^f we mean the law of a Gaussian random variable $\mathcal{N}(2\sqrt{T_n\nu(J_j)}, 1)$ and by $\mathscr{N}_m$ the statistical model associated with the family of probabilities $\{\bigotimes_{j=2}^m N_j^f : f \in \mathscr{F}\}$:

$$\mathscr{N}_m = \left(\mathbb{R}^{m-1}, \mathscr{B}(\mathbb{R}^{m-1}), \left\{\bigotimes_{j=2}^m N_j^f : f \in \mathscr{F}\right\}\right). \tag{12}$$

For each $f \in \mathscr{F}$, let $\bar{\nu}_m$ be the measure having $\bar{f}_m$ as a density with respect to ν_0 where, for every $f \in \mathscr{F}$, $\bar{f}_m$ is defined as follows.

$$\bar{f}_m(x) := \begin{cases} 1 & \text{if } x \in J_1, \\ \dfrac{\nu(J_j)}{\nu_0(J_j)} & \text{if } x \in J_j, \quad j = 2, \dots, m. \end{cases} \tag{13}$$

Furthermore, define

$$\bar{\mathscr{P}}_n^{\nu_0} = \left(D, \mathscr{D}_{T_n}, \left\{P_{T_n}^{(\gamma^{\bar{\nu}_m - \nu_0}, 0, \bar{\nu}_m)} : \frac{d\bar{\nu}_m}{d\nu_0} \in \mathscr{F}\right\}\right). \tag{14}$$

4.1. Proof of Theorem 2.5

We begin by a series of lemmas that will be needed in the proof. Before doing so, let us underline the scheme of the proof. We recall that the goal is to prove that estimating $f = \frac{d\nu}{d\nu_0}$ from the continuous observation of a Lévy process $(X_t)_{t\in[0,T_n]}$ without Gaussian part and having Lévy measure ν is asymptotically equivalent to estimating f from the Gaussian white noise model:

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n g(t)}} dW_t, \quad g = \frac{d\nu_0}{d\mathrm{Leb}}, \ t \in I.$$

Also, recall the definition of $\hat{\nu}_m$ given in (5) and read $\mathscr{P}_1 \overset{\Delta}{\Longleftrightarrow} \mathscr{P}_2$ as $\mathscr{P}_1$ is asymptotically equivalent to $\mathscr{P}_2$. Then, we can outline the proof in the following way.

- Step 1: $P_{T_n}^{(\gamma^{\nu-\nu_0}, 0, \nu)} \overset{\Delta}{\Longleftrightarrow} P_{T_n}^{(\gamma^{\hat{\nu}_m - \nu_0}, 0, \hat{\nu}_m)}$;
- Step 2: $P_{T_n}^{(\gamma^{\hat{\nu}_m - \nu_0}, 0, \hat{\nu}_m)} \overset{\Delta}{\Longleftrightarrow} \bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j))$ (Poisson approximation).
  Here $\bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j))$ represents a statistical model associated with the observation of $m-1$ independent Poisson r.v. of parameters $T_n\nu(J_j)$;

- Step 3: $\bigotimes_{j=2}^m \mathscr{P}(T_n\nu(J_j)) \overset{\Delta}{\Longleftrightarrow} \bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)}, 1)$ (Gaussian approximation);
- Step 4: $\bigotimes_{j=2}^m \mathscr{N}(2\sqrt{T_n\nu(J_j)}, 1) \overset{\Delta}{\Longleftrightarrow} (y_t)_{t\in I}$.

Lemmas 4.1–4.3 are the key ingredients of Step 2.

Lemma 4.1. *Let $\bar{\mathscr{P}}_n^{\nu_0}$ and $\mathscr{L}_m$ be the statistical models defined in (14) and (11), respectively. Under the Assumption* (H2) *we have:*

$$\Delta(\bar{\mathscr{P}}_n^{\nu_0}, \mathscr{L}_m) = 0, \quad \text{for all } m.$$

Proof. Denote by $\bar{\mathbb{N}} = \mathbb{N} \cup \{\infty\}$ and consider the statistics $S : (D, \mathscr{D}_{T_n}) \to \left(\bar{\mathbb{N}}^{m-1}, \mathcal{P}(\bar{\mathbb{N}}^{m-1})\right)$ defined by

$$S(x) = \left(N_{T_n}^{x;\,2}, \dots, N_{T_n}^{x;\,m}\right) \quad \text{with } N_{T_n}^{x;\,j} = \sum_{r\le T_n} \mathbb{I}_{J_j}(\Delta x_r). \tag{15}$$

An application of Theorem A.12 to $P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}$ and $P_{T_n}^{(0,0,\nu_0)}$, yields

$$\frac{dP_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}}{dP_{T_n}^{(0,0,\nu_0)}}(x) = \exp\left(\sum_{j=2}^m \left(\ln\left(\frac{\nu(J_j)}{\nu_0(J_j)}\right)\right) N_{T_n}^{x;j} - T_n \int_I (\bar{f}_m(y) - 1)\nu_0(dy)\right).$$

Hence, by means of the Fisher factorization theorem, we conclude that S is a sufficient statistics for $\bar{\mathscr{P}}_n^{\nu_0}$. Furthermore, under $P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}$, the random variables $N_{T_n}^{x;j}$ have Poisson distributions Q_j^f with means $T_n\nu(J_j)$. Then, by means of Property A.7, we get $\Delta(\bar{\mathscr{P}}_n^{\nu_0}, \mathscr{L}_m) = 0$, for all m. □

Let us denote by $\hat{Q}_j^f$ the law of a Poisson random variable with mean $T_n \int_{J_j} \hat{f}_m(y)\nu_0(dy)$ and let $\hat{\mathscr{L}}_m$ be the statistical model associated with the family of probabilities $\{\bigotimes_{j=2}^m \hat{Q}_j^f : f \in \mathscr{F}\}$.

Lemma 4.2.

$$\Delta(\mathscr{L}_m, \hat{\mathscr{L}}_m) \le \sup_{f\in\mathscr{F}} \sqrt{\frac{T_n}{\kappa} \int_{I\setminus[0,\varepsilon_m]} \left(f(y) - \hat{f}_m(y)\right)^2 \nu_0(dy)}.$$

Proof. By means of Facts A.2–A.4, we get:

$$\begin{aligned}
\Delta(\mathscr{L}_m, \hat{\mathscr{L}}_m) &\le \sup_{f\in\mathscr{F}} H\left(\bigotimes_{j=2}^m Q_j^f, \bigotimes_{j=2}^m \hat{Q}_j^f\right) \\
&\le \sup_{f\in\mathscr{F}} \sqrt{\sum_{j=2}^m 2H^2(Q_j^f, \hat{Q}_j^f)} \\
&= \sup_{f\in\mathscr{F}} \sqrt{2} \sqrt{\sum_{j=2}^m \left(1 - \exp\left(-\frac{T_n}{2}\left[\sqrt{\int_{J_j} \hat{f}_m(y)\nu_0(dy)} - \sqrt{\int_{J_j} f(y)\nu_0(dy)}\right]^2\right)\right)}.
\end{aligned}$$

By making use of the fact that $1 - e^{-x} \le x$ for all $x \ge 0$ and the equality $\sqrt{a} - \sqrt{b} = \frac{a-b}{\sqrt{a}+\sqrt{b}}$ combined with the lower bound $f \ge \kappa$ (that also implies $\hat{f}_m \ge \kappa$) and finally the

Cauchy–Schwarz inequality, we obtain:

$$\begin{aligned}
&1-\exp\left(-\frac{T_n}{2}\left[\sqrt{\int_{J_j}\hat{f}_m(y)\nu_0(dy)}-\sqrt{\int_{J_j}f(y)\nu_0(dy)}\right]^2\right)\\
&\quad\le\frac{T_n}{2}\left[\sqrt{\int_{J_j}\hat{f}_m(y)\nu_0(dy)}-\sqrt{\int_{J_j}f(y)\nu_0(dy)}\right]^2\\
&\quad\le\frac{T_n}{2}\frac{\left(\int_{J_j}(f(y)-\hat{f}_m(y))\nu_0(dy)\right)^2}{\kappa\nu_0(J_j)}\\
&\quad\le\frac{T_n}{2\kappa}\int_{J_j}\left(f(y)-\hat{f}_m(y)\right)^2\nu_0(dy).
\end{aligned}$$

Hence,

$$H\left(\bigotimes_{j=2}^m Q_j^f,\bigotimes_{j=2}^m\hat{Q}_j^f\right)\le\sqrt{\frac{T_n}{\kappa}\int_{I\setminus[0,\varepsilon_m]}\left(f(y)-\hat{f}_m(y)\right)^2\nu_0(dy)}.\quad\square$$

Lemma 4.3. *Let $\hat{\nu}_m$ and $\bar{\nu}_m$ the Lévy measures defined as in (5) and (13), respectively. For every $f\in\mathscr{F}$, there exists a Markov kernel K such that*

$$K\,P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)}=P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)}.$$

Proof. By construction, $\bar{\nu}_m$ and $\hat{\nu}_m$ coincide on $[0,\varepsilon_m]$. Let us denote by $\bar{\nu}_m^{\mathrm{res}}$ and $\hat{\nu}_m^{\mathrm{res}}$ the restriction on $I\setminus[0,\varepsilon_m]$ of $\bar{\nu}_m$ and $\hat{\nu}_m$ respectively, then it is enough to prove: $K\,P_{T_n}^{(\gamma^{\bar{\nu}_m^{\mathrm{res}}-\nu_0},0,\bar{\nu}_m^{\mathrm{res}})}=P_{T_n}^{(\gamma^{\hat{\nu}_m^{\mathrm{res}}-\nu_0},0,\hat{\nu}_m^{\mathrm{res}})}$. First of all, let us observe that the kernel M:

$$M(x,A)=\sum_{j=2}^m\mathbb{I}_{J_j}(x)\int_A V_j(y)\nu_0(dy),\quad x\in I\setminus[0,\varepsilon_m],\ A\in\mathscr{B}(I\setminus[0,\varepsilon_m])$$

is defined in such a way that $M\bar{\nu}_m^{\mathrm{res}}=\hat{\nu}_m^{\mathrm{res}}$. Indeed, for all $A\in\mathscr{B}(I\setminus[0,\varepsilon_m])$,

$$\begin{aligned}
M\bar{\nu}_m^{\mathrm{res}}(A)&=\sum_{j=2}^m\int_{J_j}M(x,A)\bar{\nu}_m^{\mathrm{res}}(dx)=\sum_{j=2}^m\int_{J_j}\left(\int_A V_j(y)\nu_0(dy)\right)\bar{\nu}_m^{\mathrm{res}}(dx)\\
&=\sum_{j=2}^m\left(\int_A V_j(y)\nu_0(dy)\right)\nu(J_j)=\int_A\hat{f}_m(y)\nu_0(dy)=\hat{\nu}_m^{\mathrm{res}}(A).
\end{aligned}\tag{16}$$

Observe that $(\gamma^{\bar{\nu}_m^{\mathrm{res}}-\nu_0},0,\bar{\nu}_m^{\mathrm{res}})$ and $(\gamma^{\hat{\nu}_m^{\mathrm{res}}-\nu_0},0,\hat{\nu}_m^{\mathrm{res}})$ are Lévy triplets associated with compound Poisson processes since $\bar{\nu}_m^{\mathrm{res}}$ and $\hat{\nu}_m^{\mathrm{res}}$ are finite Lévy measures. The Markov kernel K interchanging the laws of the Lévy processes is constructed explicitly in the case of compound Poisson processes. Indeed if $\bar{X}$ is the compound Poisson process having Lévy measure $\bar{\nu}_m^{\mathrm{res}}$, then $\bar{X}_t=\sum_{i=1}^{N_t}\bar{Y}_i$, where N_t is a Poisson process of intensity $\iota_m:=\bar{\nu}_m^{\mathrm{res}}(I\setminus[0,\varepsilon_m])$ and the $\bar{Y}_i$ are i.i.d. random variables with probability law $\frac{1}{\iota_m}\bar{\nu}_m^{\mathrm{res}}$. Moreover, given a trajectory of $\bar{X}$, both the trajectory $(n_t)_{t\in[0,T_n]}$ of the Poisson process $(N_t)_{t\in[0,T_n]}$ and the realizations $\bar{y}_i$

of $\bar{Y}_i$, $i = 1, \dots, n_{T_n}$ are uniquely determined. This allows us to construct n_{T_n} i.i.d. random variables $\hat{Y}_i$ as follows: for every realization $\bar{y}_i$ of $\bar{Y}_i$, we define the realization $\hat{y}_i$ of $\hat{Y}_i$ by throwing it according to the probability law $M(\bar{y}_i, \cdot)$. Hence, thanks to (16), $(\hat{Y}_i)_i$ are i.i.d. random variables with probability law $\frac{1}{\iota_m}\hat{\nu}_m^{\text{res}}$. The desired Markov kernel K (defined on the Skorokhod space) is then given by:

$$K : (\bar{X}_t)_{t\in[0,T_n]} \longmapsto \left(\hat{X}_t := \sum_{i=1}^{N_t} \hat{Y}_i\right)_{t\in[0,T_n]}.$$

Finally, observe that, since

$$\iota_m = \int_{I\setminus[0,\varepsilon_m]} \bar{f}_m(y)\nu_0(dy) = \int_{I\setminus[0,\varepsilon_m]} f(y)\nu_0(dy) = \int_{I\setminus[0,\varepsilon_m]} \hat{f}_m(y)\nu_0(dy),$$

$(\hat{X}_t)_{t\in[0,T_n]}$ is a compound Poisson process with Lévy measure $\hat{\nu}_m^{\text{res}}$. □

Let us now state two lemmas needed to understand Step 4.

Lemma 4.4. *Denote by $\mathscr{W}_m^{\#}$ the statistical model associated with the continuous observation of a trajectory from the Gaussian white noise:*

$$dy_t = \sqrt{f(t)}dt + \frac{1}{2\sqrt{T_n}\sqrt{g(t)}}dW_t, \quad t \in I\setminus[0,\varepsilon_m].$$

Then, according with the notation introduced in Section 2.1 and at the beginning of Section 4, we have

$$\Delta(\mathscr{N}_m, \mathscr{W}_m^{\#}) \le 2\sqrt{T_n}\sup_{f\in\mathscr{F}}\big(A_m(f) + B_m(f)\big).$$

Proof. As a preliminary remark observe that $\mathscr{W}_m^{\#}$ is equivalent to the model that observes a trajectory from:

$$d\bar{y}_t = \sqrt{f(t)}g(t)dt + \frac{\sqrt{g(t)}}{2\sqrt{T_n}}dW_t, \quad t \in I\setminus[0,\varepsilon_m].$$

Let us denote by $\bar{Y}_j$ the increments of the process $(\bar{y}_t)$ over the intervals J_j, $j = 2, \dots, m$, i.e.

$$\bar{Y}_j := \bar{y}_{v_j} - \bar{y}_{v_{j-1}} \sim \mathscr{N}\left(\int_{J_j}\sqrt{f(y)}\nu_0(dy), \frac{\nu_0(J_j)}{4T_n}\right)$$

and denote by $\bar{\mathscr{N}}_m$ the statistical model associated with the distributions of these increments. As an intermediate result, we will prove that

$$\Delta(\mathscr{N}_m, \bar{\mathscr{N}}_m) \le 2\sqrt{T_n}\sup_{f\in\mathscr{F}} B_m(f), \quad \text{for all } m. \tag{17}$$

To that aim, remark that the experiment $\bar{\mathscr{N}}_m$ is equivalent to observing $m-1$ independent Gaussian random variables of means $\frac{2\sqrt{T_n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy)$, $j = 2, \dots, m$ and variances identically 1, name this last experiment $\mathscr{N}_m^{\#}$. Hence, using also Property A.1, Facts A.2 and A.5

we get:

$$\Delta(\mathscr{N}_m, \bar{\mathscr{N}_m}) \le \Delta(\mathscr{N}_m, \mathscr{N}_m^{\#}) \le \sqrt{\sum_{j=2}^{m}\left(\frac{2\sqrt{T_n}}{\sqrt{\nu_0(J_j)}}\int_{J_j}\sqrt{f(y)}\nu_0(dy) - 2\sqrt{T_n\nu(J_j)}\right)^2}.$$

Since it is clear that $\delta(\mathscr{W}_m^{\#}, \bar{\mathscr{N}_m}) = 0$, in order to bound $\Delta(\mathscr{N}_m, \mathscr{W}_m^{\#})$ it is enough to bound $\delta(\bar{\mathscr{N}_m}, \mathscr{W}_m^{\#})$. Using similar ideas as in [10, Section 8.2], we define a new stochastic process as:

$$Y_t^* = \sum_{j=2}^{m}\bar{Y}_j\int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy) + \frac{1}{2\sqrt{T_n}}\sum_{j=2}^{m}\sqrt{\nu_0(J_j)}B_j(t), \quad t \in I\backslash[0,\varepsilon_m],$$

where the $(B_j(t))$ are independent centered Gaussian processes independent of (W_t) and with variances

$$\mathrm{Var}(B_j(t)) = \int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy) - \left(\int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy)\right)^2.$$

These processes can be constructed from a standard Brownian bridge $\{B(s), s \in [0,1]\}$, independent of (W_t), via

$$B_i(t) = B\left(\int_{\varepsilon_m}^{t}V_i(y)\nu_0(dy)\right).$$

By construction, (Y_t^*) is a Gaussian process with mean and variance given by, respectively:

$$\mathbb{E}[Y_t^*] = \sum_{j=2}^{m}\mathbb{E}[\bar{Y}_j]\int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy) = \sum_{j=2}^{m}\left(\int_{J_j}\sqrt{f(y)}\nu_0(dy)\right)\int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy),$$

$$\begin{aligned}\mathrm{Var}[Y_t^*] &= \sum_{j=2}^{m}\mathrm{Var}[\bar{Y}_j]\left(\int_{\varepsilon_m}^{t}V_j(y)\nu_0(dy)\right)^2 + \frac{1}{4T_n}\sum_{j=2}^{m}\nu_0(J_j)\mathrm{Var}(B_j(t))\\ &= \frac{1}{4T_n}\int_{\varepsilon_m}^{t}\sum_{j=2}^{m}\nu_0(J_j)V_j(y)\nu_0(dy) = \frac{1}{4T_n}\int_{\varepsilon_m}^{t}\nu_0(dy) = \frac{\nu_0([\varepsilon_m,t])}{4T_n}.\end{aligned}$$

One can compute in the same way the covariance of (Y_t^*) finding that

$$\mathrm{Cov}(Y_s^*, Y_t^*) = \frac{\nu_0([\varepsilon_m,s])}{4T_n}, \quad \forall s \le t.$$

We can then deduce that

$$Y_t^* = \int_{\varepsilon_m}^{t}\widehat{\sqrt{f}}_m(y)\nu_0(dy) + \int_{\varepsilon_m}^{t}\frac{\sqrt{g(s)}}{2\sqrt{T_n}}dW_s^*, \quad t \in I\backslash[0,\varepsilon_m],$$

where (W_t^*) is a standard Brownian motion and

$$\widehat{\sqrt{f}}_m(x) := \sum_{j=2}^{m}\left(\int_{J_j}\sqrt{f(y)}\nu_0(dy)\right)V_j(x).$$

Applying Fact A.6, we get that the total variation distance between the process $(Y_t^*)_{t\in I\backslash[0,\varepsilon_m]}$ constructed from the random variables $\bar{Y}_j$, $j = 2,\dots,m$ and the Gaussian process $(\bar{y}_t)_{t\in I\backslash[0,\varepsilon_m]}$

is bounded by

$$\sqrt{4T_n \int_{I\setminus[0,\varepsilon_m]} \left(\widehat{\sqrt{f}}_m - \sqrt{f(y)}\right)^2 \nu_0(dy)},$$

which gives the term in $A_m(f)$. □

Lemma 4.5. *In accordance with the notation of Lemma 4.4, we have:*

$$\Delta(\mathscr{W}_m^{\#}, \mathscr{W}_n^{\nu_0}) = O\left(\sup_{f\in\mathscr{F}} \sqrt{T_n \int_0^{\varepsilon_m} \left(\sqrt{f(t)} - 1\right)^2 \nu_0(dt)}\right). \tag{18}$$

Proof. Clearly $\delta(\mathscr{W}_n^{\nu_0}, \mathscr{W}_m^{\#}) = 0$. To show that $\delta(\mathscr{W}_m^{\#}, \mathscr{W}_n^{\nu_0}) \to 0$, let us consider a Markov kernel $K^{\#}$ from $C(I\setminus[0,\varepsilon_m])$ to $C(I)$ defined as follows: introduce a Gaussian process, $(B_t^m)_{t\in[0,\varepsilon_m]}$ with mean equal to t and covariance

$$\mathrm{Cov}(B_s^m, B_t^m) = \int_0^{\varepsilon_m} \frac{1}{4T_n g(s)} \mathbb{I}_{[0,s]\cap[0,t]}(z)dz.$$

In particular,

$$\mathrm{Var}(B_t^m) = \int_0^t \frac{1}{4T_n g(s)} ds.$$

Consider it as a process on the whole of I by defining $B_t^m = B_{\varepsilon_m}^m \ \forall t > \varepsilon_m$. Let ω_t be a trajectory in $C(I\setminus[0,\varepsilon_m])$, which again we constantly extend to a trajectory on the whole of I. Then, we define $K^{\#}$ by sending the trajectory ω_t to the trajectory $\omega_t + B_t^m$. If we define $\tilde{\mathbb{W}}_n$ as the law induced on $C(I)$ by

$$d\tilde{y}_t = h(t)dt + \frac{dW_t}{2\sqrt{T_n g(t)}}, \quad t \in I,\ h(t) = \begin{cases} 1 & t \in [0,\varepsilon_m] \\ \sqrt{f(t)} & t \in I\setminus[0,\varepsilon_m], \end{cases}$$

then $K^{\#}\mathbb{W}_n^f|_{I\setminus[0,\varepsilon_m]} = \tilde{\mathbb{W}}_n$, where $\mathbb{W}_n^f$ is defined as in (8). By means of Fact A.6 we deduce (18). □

Proof of Theorem 2.5. The proof of the theorem follows by combining the previous lemmas together:

- Step 1: Let us denote by $\hat{\mathscr{P}}_{n,m}^{\nu_0}$ the statistical model associated with the family of probabilities $(P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)} : \frac{d\nu}{d\nu_0} \in \mathscr{F})$. Thanks to Property A.1, Fact A.2 and Theorem A.13 we have that

$$\Delta(\mathscr{P}_n^{\nu_0}, \hat{\mathscr{P}}_{n,m}^{\nu_0}) \le \sqrt{\frac{T_n}{2}} \sup_{f\in\mathscr{F}} H(f, \hat{f}_m).$$

- Step 2: On the one hand, thanks to Lemma 4.1, one has that the statistical model associated with the family of probability $(P_{T_n}^{(\gamma^{\bar{\nu}_m-\nu_0},0,\bar{\nu}_m)} : \frac{d\nu}{d\nu_0} \in \mathscr{F})$ is equivalent to $\mathscr{L}_m$. By means of Lemma 4.2 we can bound $\Delta(\mathscr{L}_m, \hat{\mathscr{L}}_m)$. On the other hand it is easy to see that $\delta(\hat{\mathscr{P}}_{n,m}^{\nu_0}, \hat{\mathscr{L}}_m) = 0$. Indeed, it is enough to consider the statistics

$$S : x \mapsto \left(\sum_{r\le T_n} \mathbb{I}_{J_2}(\Delta x_r), \ldots, \sum_{r\le T_n} \mathbb{I}_{J_m}(\Delta x_r)\right)$$

since the law of the random variable $\sum_{r\le T_n}\mathbb{I}_{J_j}(\Delta x_r)$ under $P_{T_n}^{(\gamma^{\hat{\nu}_m-\nu_0},0,\hat{\nu}_m)}$ is Poisson of parameter $T_n\int_{J_j}\hat{f}_m(y)\nu_0(dy)$ for all $j=2,\dots,m$. Finally, Lemmas 4.1 and 4.3 allow us to conclude that $\delta(\mathscr{L}_m,\hat{\mathscr{P}}_{n,m}^{\nu_0})=0$. Collecting all the pieces together, we get

$$\Delta(\hat{\mathscr{P}}_{n,m}^{\nu_0},\mathscr{L}_m)\le\sup_{f\in\mathscr{F}}\sqrt{\frac{T_n}{\kappa}\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat{f}_m(y)\big)^2\nu_0(dy)}.$$

- Step 3: Applying Theorem A.9 and Fact A.3 we can pass from the Poisson approximation given by $\mathscr{L}_m$ to a Gaussian one obtaining

$$\Delta(\mathscr{L}_m,\mathscr{N}_m)=C\sup_{f\in\mathscr{F}}\sqrt{\sum_{j=2}^m\frac{2}{T_n\nu(J_j)}}\le C\sqrt{\sum_{j=2}^m\frac{2\kappa}{T_n\nu_0(J_j)}}=C\sqrt{\frac{(m-1)2\kappa}{T_n\mu_m}}.$$

- Step 4: Finally, Lemmas 4.4 and 4.5 allow us to conclude that:

$$\Delta(\mathscr{P}_n^{\nu_0},\mathscr{W}_n^{\nu_0})=O\Big(\sqrt{T_n}\sup_{f\in\mathscr{F}}\big(A_m(f)+B_m(f)+C_m\big)\Big)$$
$$+O\Big(\sqrt{T_n}\sup_{f\in\mathscr{F}}\sqrt{\int_{I\setminus[0,\varepsilon_m]}\big(f(y)-\hat{f}_m(y)\big)^2\nu_0(dy)}+\sqrt{\frac{m}{T_n\mu_m}}\Big).\quad\square$$

4.2. *Proof of Theorem 2.6*

Again, before stating some technical lemmas, let us highlight the main ideas of the proof. We recall that the goal is to prove that estimating $f=\frac{d\nu}{d\nu_0}$ from the discrete observations $(X_{t_i})_{i=0}^n$ of a Lévy process without Gaussian component and having Lévy measure ν is asymptotically equivalent to estimating f from the Gaussian white noise model

$$dy_t=\sqrt{f(t)}dt+\frac{1}{2\sqrt{T_n g(t)}}dW_t,\quad g=\frac{d\nu_0}{d\mathrm{Leb}},\ t\in I.$$

Reading $\mathscr{P}_1\overset{\Delta}{\Longleftrightarrow}\mathscr{P}_2$ as $\mathscr{P}_1$ is asymptotically equivalent to $\mathscr{P}_2$, we have:

- Step 1. Clearly $(X_{t_i})_{i=0}^n\overset{\Delta}{\Longleftrightarrow}(X_{t_i}-X_{t_{i-1}})_{i=1}^n$. Moreover, $(X_{t_i}-X_{t_{i-1}})_i\overset{\Delta}{\Longleftrightarrow}(\epsilon_i Y_i)$ where (ϵ_i) are i.i.d. Bernoulli r.v. with parameter $\alpha=\iota_m\Delta_n e^{-\iota_m\Delta_n}$, $\iota_m:=\int_{I\setminus[0,\varepsilon_m]}f(y)\nu_0(dy)$ and $(Y_i)_i$ are i.i.d. r.v. independent of $(\epsilon_i)_{i=1}^n$ and of density $\frac{f}{\iota_m}$ with respect to $\nu_{0|_{I\setminus[0,\varepsilon_m]}}$;
- Step 2. $(\epsilon_i Y_i)_i\overset{\Delta}{\Longleftrightarrow}\mathcal{M}(n;(\gamma_j)_{j=1}^m)$, where $\mathcal{M}(n;(\gamma_j)_{j=1}^m)$ is a multinomial distribution with $\gamma_1=1-\alpha$ and $\gamma_i:=\alpha\nu(J_i)$ $i=2,\dots,m$;
- Step 3. Gaussian approximation: $\mathcal{M}(n;(\gamma_1,\dots\gamma_m))\overset{\Delta}{\Longleftrightarrow}\bigotimes_{j=2}^m\mathcal{N}(2\sqrt{T_n\nu(J_j)},1)$;
- Step 4. $\bigotimes_{j=2}^m\mathcal{N}(2\sqrt{T_n\nu(J_j)},1)\overset{\Delta}{\Longleftrightarrow}(y_t)_{t\in I}$.

Lemma 4.6. *Let ν_i, $i=1,2$, be Lévy measures such that $\nu_1\ll\nu_2$ and $b_1-b_2=\int_{|y|\le1}y(\nu_1-\nu_2)(dy)<\infty$. Then, for all $0<t<\infty$, we have:*

$$\Big\|Q_t^{(b_1,0,\mu_1)}-Q_t^{(b_2,0,\mu_2)}\Big\|_{TV}\le\sqrt{\frac{t}{2}}H(\nu_1,\nu_2).$$

Proof. For all given t, let K_t be the Markov kernel defined as $K_t(\omega, A) := \mathbb{I}_A(\omega_t)$, $\forall\ A \in \mathscr{B}(\mathbb{R})$, $\forall\ \omega \in D$. Then we have:

$$\begin{aligned}\left\| Q_t^{(b_1,0,\nu_1)} - Q_t^{(b_2,0,\nu_2)} \right\|_{TV} &= \left\| K_t P_t^{(b_1,0,\nu_1)} - K_t P_t^{(b_2,0,\nu_2)} \right\|_{TV} \\ &\leq \left\| P_t^{(b_1,0,\nu_1)} - P_t^{(b_2,0,\nu_2)} \right\|_{TV} \\ &\leq \sqrt{\frac{t}{2}} H(\nu_1, \nu_2),\end{aligned}$$

where we have used that Markov kernels reduce the total variation distance and Theorem A.13. □

Lemma 4.7. *Let $(P_i)_{i=1}^n$, $(Y_i)_{i=1}^n$ and $(\epsilon_i)_{i=1}^n$ be samples of, respectively, Poisson random variables $\mathscr{P}(\lambda_i)$, random variables with common distribution and Bernoulli random variables of parameters $\lambda_i e^{-\lambda_i}$, which are all independent. Let us denote by $Q_{(Y_i,P_i)}$ (resp. $Q_{(Y_i,\epsilon_i)}$) the law of $\sum_{j=1}^{P_i} Y_j$ (resp., $\epsilon_i Y_i$). Then:*

$$\left\| \bigotimes_{i=1}^n Q_{(Y_i,P_i)} - \bigotimes_{i=1}^n Q_{(Y_i,\epsilon_i)} \right\|_{TV} \leq 2 \sqrt{\sum_{i=1}^n \lambda_i^2}. \tag{19}$$

The proof of this lemma can be found in [34, Section 2.1].

Lemma 4.8. *Let f_m^{tr} be the truncated function defined as follows:*

$$f_m^{\mathrm{tr}}(x) = \begin{cases} 1 & \text{if } x \in [0, \varepsilon_m] \\ f(x) & \text{otherwise} \end{cases}$$

and let ν_m^{tr} (resp. ν_m^{res}) be the Lévy measure having f_m^{tr} (resp. $f|_{I\setminus[0,\varepsilon_m]}$) as a density with respect to ν_0. Denote by $\mathscr{Q}_n^{\mathrm{tr},\nu_0}$ the statistical model associated with the family of probabilities $\left(\bigotimes_{i=1}^n Q_{t_i-t_{i-1}}^{(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0},0,\nu_m^{\mathrm{tr}})} : \frac{d\nu_m^{\mathrm{tr}}}{d\nu_0} \in \mathscr{F}\right)$ and by $\mathscr{Q}_n^{\mathrm{res},\nu_0}$ the model associated with the family of probabilities $\left(\bigotimes_{i=1}^n Q_{t_i-t_{i-1}}^{(\gamma^{\nu_m^{\mathrm{res}}-\nu_0},0,\nu_m^{\mathrm{res}})} : \frac{d\nu_m^{\mathrm{res}}}{d\nu_0} \in \mathscr{F}\right)$. Then:

$$\Delta(\mathscr{Q}_n^{\mathrm{tr},\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) = 0.$$

Proof. Let us start by proving that $\delta(\mathscr{Q}_n^{\mathrm{tr},\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) = 0$. For that, let us consider two independent Lévy processes, X^{tr} and X^0, of Lévy triplets given by $\left(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0}, 0, \nu_m^{\mathrm{tr}-\nu_0}\right)$ and $\left(0, 0, \nu_0|_{[0,\varepsilon_m]}\right)$, respectively. Then it is clear (using the *Lévy–Khintchine formula*) that the random variable $X_t^{\mathrm{tr}} - X_t^0$ is a randomization of X_t^{tr} (since the law of X_t^0 does not depend on ν) having law $Q_t^{(\gamma^{\nu_m^{\mathrm{res}}-\nu_0},0,\nu_m^{\mathrm{res}})}$, for all $t \geq 0$. Similarly, one can prove that $\delta(\mathscr{Q}_n^{\mathrm{res},\nu_0}, \mathscr{Q}_n^{\mathrm{tr},\nu_0}) = 0$. □

Proof of Theorem 2.6. As a preliminary remark, observe that the model $\mathscr{Q}_n^{\nu_0}$ is equivalent to the one that observes the increments of $\left((x_t), P_{T_n}^{(\gamma^{\nu-\nu_0},0,\nu)}\right)$, that is, the model $\tilde{\mathscr{Q}}_n^{\nu_0}$ associated with the family of probabilities $\left(\bigotimes_{i=1}^n Q_{t_i-t_{i-1}}^{(\gamma^{\nu-\nu_0},0,\nu)} : \frac{d\nu}{d\nu_0} \in \mathscr{F}\right)$.

- Step 1: Facts A.2–A.3 and Lemma 4.6 allow us to write

$$\left\|\bigotimes_{i=1}^{n} Q_{\Delta_n}^{(\gamma^{\nu-\nu_0},0,\nu)} - \bigotimes_{i=1}^{n} Q_{\Delta_n}^{(\gamma^{\nu_m^{\mathrm{tr}}-\nu_0},0,\nu_m^{\mathrm{tr}})}\right\|_{TV} \le \sqrt{n\sqrt{\frac{\Delta_n}{2}}H(\nu,\nu_m^{\mathrm{tr}})}$$
$$= \sqrt{n\sqrt{\frac{\Delta_n}{2}}\sqrt{\int_0^{\varepsilon_m}\left(\sqrt{f(y)}-1\right)^2\nu_0(dy)}}.$$

Using this bound together with Lemma 4.8 and the notation therein, we get $\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{Q}_n^{\mathrm{res},\nu_0}) \le \sqrt{n\sqrt{\frac{\Delta_n}{2}}}\sup_{f\in\mathscr{F}} H(f, f_m^{\mathrm{tr}})$. Observe that ν_m^{res} is a finite Lévy measure, hence $\left((x_t), P_{T_n}^{(\gamma^{\nu_m^{\mathrm{res}}},0,\nu_m^{\mathrm{res}})}\right)$ is a compound Poisson process with intensity equal to $\iota_m := \int_{I\setminus[0,\varepsilon_m]} f(y)\nu_0(dy)$ and jumps size density $\frac{f(x)g(x)}{\iota_m}$, for all $x \in I\setminus[0,\varepsilon_m]$ (recall that we are assuming that ν_0 has a density g with respect to Lebesgue). In particular, this means that $Q_{\Delta_n}^{(\gamma^{\nu_m^{\mathrm{res}}},0,\nu_m^{\mathrm{res}})}$ can be seen as the law of the random variable $\sum_{j=1}^{P_i} Y_j$ where P_i is a Poisson variable of mean $\iota_m\Delta_n$, independent from $(Y_i)_{i\ge0}$, a sequence of i.i.d. random variables with density $\frac{fg}{\iota_m}\mathbb{I}_{I\setminus[0,\varepsilon_m]}$ with respect to Lebesgue. Remark also that ι_m is confined between $\kappa\nu_0\big(I\setminus[0,\varepsilon_m]\big)$ and $M\nu_0\big(I\setminus[0,\varepsilon_m]\big)$.

Let $(\epsilon_i)_{i\ge0}$ be a sequence of i.i.d. Bernoulli variables, independent of $(Y_i)_{i\ge0}$, with mean $\iota_m\Delta_n e^{-\iota_m\Delta_n}$. For $i = 1,\dots,n$, denote by $Q_i^{\epsilon,f}$ the law of the variable $\epsilon_i Y_i$ and by $\mathscr{Q}_n^{\epsilon}$ the statistical model associated with the observations of the vector $(\epsilon_1 Y_1,\dots,\epsilon_n Y_n)$, i.e.

$$\mathscr{Q}_n^{\epsilon} = \left(I^n, \mathscr{B}(I^n), \left\{\bigotimes_{i=1}^{n} Q_i^{\epsilon,f} : f\in\mathscr{F}\right\}\right).$$

Furthermore, denote by $\tilde{Q}_i^f$ the law of $\sum_{j=1}^{P_i} Y_j$. Then an application of Lemma 4.7 yields:

$$\left\|\bigotimes_{i=1}^{n}\tilde{Q}_i^f - \bigotimes_{i=1}^{n} Q_i^{\epsilon,f}\right\|_{TV} \le 2\iota_m\sqrt{n\Delta_n^2} \le 2M\nu_0\big(I\setminus[0,\varepsilon_m]\big)\sqrt{n\Delta_n^2}.$$

Hence, we get:

$$\Delta(\mathscr{Q}_n^{\mathrm{res},\nu_0}, \mathscr{Q}_n^{\epsilon}) = O\left(\nu_0\big(I\setminus[0,\varepsilon_m]\big)\sqrt{n\Delta_n^2}\right). \tag{20}$$

Here the O depends only on M.

- Step 2: Let us introduce the following random variables:

$$Z_1 = \sum_{j=1}^{n}\mathbb{I}_{\{0\}}(\epsilon_j Y_j); \qquad Z_i = \sum_{j=1}^{n}\mathbb{I}_{J_i}(\epsilon_j Y_j), \quad i = 2,\dots,m.$$

Observe that the law of the vector $(Z_1,\dots,Z_m)$ is multinomial $\mathcal{M}(n;\gamma_1,\dots,\gamma_m)$ where

$$\gamma_1 = 1-\iota_m\Delta_n e^{-\iota_m\Delta_n}, \qquad \gamma_i = \Delta_n e^{-\iota_m\Delta_n}\nu(J_i), \quad i = 2,\dots,m.$$

Let us denote by $\mathcal{M}_n$ the statistical model associated with the observation of $(Z_1,\dots,Z_m)$. Clearly $\delta(\mathscr{Q}_n^{\epsilon},\mathcal{M}_n) = 0$. Indeed, $\mathcal{M}_n$ is the image experiment by the random variable

$S : I^n \to \{1, \dots, n\}^m$ defined as

$$S(x_1, \dots, x_n) = \Big(\#\{j : x_j = 0\}; \#\big\{j : x_j \in J_2\big\}; \dots; \#\big\{j : x_j \in J_m\big\}\Big),$$

where $\#A$ denotes the cardinal of the set A.

We shall now prove that $\delta(\mathcal{M}_n, \mathcal{Q}_n^\epsilon) \le \sup_{f\in\mathscr{F}} \sqrt{n\Delta_n H^2(f, \hat{f}_m)}$. We start by defining a discrete random variable X^* concentrated at the points $0, x_i^*,\ i = 2, \dots, m$:

$$\mathbb{P}(X^* = y) = \begin{cases} \gamma_i & \text{if } y = x_i^*, \quad i = 1, \dots, m, \\ 0 & \text{otherwise,} \end{cases}$$

with the convention $x_1^* = 0$. It is easy to see that $\mathcal{M}_n$ is equivalent to the statistical model associated with n independent copies of X^*. Let us introduce the Markov kernel

$$K(x_i^*, A) = \begin{cases} \mathbb{I}_A(0) & \text{if } i = 1, \\ \displaystyle\int_A V_i(x) v_0(dx) & \text{otherwise.} \end{cases}$$

Denote by P^* the law of the random variable X^* and by $Q_i^{\epsilon,\hat{f}}$ the law of a random variable $\epsilon_i \hat{Y}_i$ where ϵ_i is Bernoulli independent of $\hat{Y}_i$, with mean $\iota_m \Delta_n e^{-\iota_m \Delta_n}$ and $\hat{Y}_i$ has a density $\frac{\hat{f}_m g}{\iota_m} \mathbb{I}_{I\setminus[0,\varepsilon_m]}$ with respect to Lebesgue. The same computations as in Lemma 4.3 prove that $KP^* = Q_i^{\epsilon,\hat{f}}$. Hence, thanks to Remark A.8, we get the equivalence between $\mathcal{M}_n$ and the statistical model associated with the observations of n independent copies of $\epsilon_i \hat{Y}_i$. In order to bound $\delta(\mathcal{M}_n, \mathcal{Q}_n^\epsilon)$ it is enough to bound the total variation distance between the probabilities $\bigotimes_{i=1}^n Q_i^{\epsilon,f}$ and $\bigotimes_{i=1}^n Q_i^{\epsilon,\hat{f}}$. Alternatively, we can bound the Hellinger distance between each of the $Q_i^{\epsilon,f}$ and $Q_i^{\epsilon,\hat{f}}$, thanks to Facts A.2 and A.3, which is:

$$\begin{aligned} \left\| \bigotimes_{i=1}^n Q_i^{\epsilon,f} - \bigotimes_{i=1}^n Q_i^{\epsilon,\hat{f}} \right\|_{TV} &\le \sqrt{\sum_{i=1}^n H^2\big(Q_i^{\epsilon,f}, Q_i^{\epsilon,\hat{f}}\big)} \\ &= \sqrt{\sum_{i=1}^n \frac{1-\gamma_1}{\iota} H^2(f, \hat{f}_m)} \le \sqrt{n\Delta_n H^2(f, \hat{f}_m)}. \end{aligned}$$

It follows that

$$\delta(\mathcal{M}_n, \mathcal{Q}_n^\epsilon) \le \sqrt{n\Delta_n} \sup_{f\in\mathscr{F}} H(f, \hat{f}_m).$$

- Step 3: Let us denote by $\mathcal{N}_m^*$ the statistical model associated with the observation of m independent Gaussian variables $\mathcal{N}(n\gamma_i, n\gamma_i),\ i = 1, \dots, m$. Very similar computations to those in [10] yield

$$\Delta(\mathcal{M}_n, \mathcal{N}_m^*) = O\Big(\frac{m \ln m}{\sqrt{n}}\Big).$$

In order to prove the asymptotic equivalence between $\mathcal{M}_n$ and $\mathcal{N}_m$ defined as in (12) we need to introduce some auxiliary statistical models. Let us denote by $\mathcal{A}_m$ the experiment obtained from $\mathcal{N}_m^*$ by disregarding the first component. Furthermore, let us denote by $\mathcal{N}_m^\#$ the experiment associated with the observation of $m-1$ independent Gaussian variables $\mathcal{N}(\sqrt{n\gamma_i}, \frac{1}{4}),\ i = 2, \dots, m$. First of all, let us prove that $\mathcal{N}_m^*$ and $\mathcal{A}_m$ are asymptotically

equivalent. One direction is trivial. In the other direction, the Markov kernel is given by

$$K(x_2,\dots,x_m,A)=\mathbb{E}[\mathbb{I}_A(X,x_2,\dots,x_m)],\quad A\subset\mathbb{R}^m,$$

where X is a Gaussian random variable with mean and variance both equal to n. The image experiment of $\mathcal{A}_m$ through K is the statistical model associated with the observation of m independent Gaussian random variables of the form $\mathcal{N}(n,n)\otimes\bigotimes_{i=2}^m\mathcal{N}(n\gamma_i,n\gamma_i)$. The total variation distance can then be computed explicitly: it equals the distance between the first components, for which the general formula of Fact A.5 can be used. We get the bound:

$$\Delta(\mathcal{N}_m^*,\mathcal{A}_m)\le\sup_{f\in\mathscr{F}}\sqrt{\Big(2+\frac{n}{2}\Big)(1-\gamma_1)^2}=O\Big(\nu_0\big(I\setminus[0,\varepsilon_m]\big)\Delta_n\sqrt{n}\Big).$$

Moreover, using a result contained in [10], Section 7.2, one has that

$$\Delta(\mathcal{A}_m,\mathcal{N}_m^{\#})=O\Big(\frac{m}{\sqrt{n}}\Big).$$

Finally, using Facts A.2 and A.5 we can write

$$\begin{aligned}\Delta(\mathcal{N}_m^{\#},\mathcal{N}_m)&\le\sqrt{2\sum_{i=2}^m\Big(\sqrt{T_n\nu(J_i)}-\sqrt{T_n\nu(J_i)\exp(-\iota_m\Delta_n)}\Big)^2}\\&\le\sqrt{2T_n\Delta_n^2\iota_m^3}\le\sqrt{2n\Delta_n^3M^3\big(\nu_0\big(I\setminus[0,\varepsilon_m]\big)\big)^3}.\end{aligned}$$

To sum up,

$$\Delta(\mathcal{M}_n,\mathcal{N}_m)=O\Big(\frac{m\ln m}{\sqrt{n}}+\sqrt{n\Delta_n^3\big(\nu_0\big(I\setminus[0,\varepsilon_m]\big)\big)^3}+\nu_0\big(I\setminus[0,\varepsilon_m]\big)\Delta_n\sqrt{n}\Big),$$

with the O depending only on κ and M.

- Step 4: An application of Lemmas 4.4 and 4.5 yields

$$\Delta(\mathcal{N}_m,\mathscr{W}_n^{\nu_0})\le2\sqrt{T_n}\sup_{f\in\mathscr{F}}\big(A_m(f)+B_m(f)+C_m(f)\big).\quad\square$$

5. Proofs of the examples

The purpose of this section is to give detailed proofs of Examples 2.4 and Examples 3.1–3.3. As in Section 4 we suppose $I\subseteq\mathbb{R}_+$. We start by giving some bounds for the quantities $A_m(f)$, $B_m(f)$ and $L_2(f,\hat{f}_m)$, the L_2-distance between the restriction of f and $\hat{f}_m$ on $I\setminus[0,\varepsilon_m]$.

5.1. Bounds for $A_m(f)$, $B_m(f)$, $L_2(f,\hat{f}_m)$ when $\hat{f}_m$ is piecewise linear

In this section we suppose f to be in $\mathscr{F}^I_{(\gamma,K,\kappa,M)}$ defined as in (3). We are going to assume that the V_j are given by triangular/trapezoidal functions as in (6). In particular, in this case $\hat{f}_m$ is piecewise linear.

Lemma 5.1. *Let $0<\kappa<M$ be two constants and let f_i, $i=1,2$ be functions defined on an interval J and such that $\kappa\le f_i\le M$, $i=1,2$. Then, for any measure ν_0, we have:*

$$\frac{1}{4M}\int_J\big(f_1(x)-f_2(x)\big)^2\nu_0(dx)\le\int_J\big(\sqrt{f_1(x)}-\sqrt{f_2(x)}\big)^2\nu_0(dx)$$

$$\leq \frac{1}{4\kappa}\int_J \big(f_1(x)-f_2(x)\big)^2 \nu_0(dx).$$

Proof. This simply comes from the following inequalities:

$$\frac{1}{2\sqrt{M}}|f_1(x)-f_2(x)| \leq \frac{|f_1(x)-f_2(x)|}{\sqrt{f_1(x)}+\sqrt{f_2(x)}} = \big|\sqrt{f_1(x)}-\sqrt{f_2(x)}\big|$$
$$\leq \frac{1}{2\sqrt{\kappa}}|f_1(x)-f_2(x)|. \quad \square$$

Recall that x_i^* is chosen so that $\int_{J_i}(x-x_i^*)\nu_0(dx)=0$. Consider the following Taylor expansions for $x\in J_i$:

$$f(x)=f(x_i^*)+f'(x_i^*)(x-x_i^*)+R_i(x); \qquad \hat{f}_m(x)=\hat{f}_m(x_i^*)+\hat{f}_m'(x_i^*)(x-x_i^*),$$

where $\hat{f}_m(x_i^*)=\frac{\nu(J_i)}{\nu_0(J_i)}$ and $\hat{f}_m'(x_i^*)$ is the left or right derivative in x_i^* depending whether $x<x_i^*$ or $x>x_i^*$ (as $\hat{f}_m$ is piecewise linear, no rest is involved in its Taylor expansion).

Lemma 5.2. *The following estimates hold:*

$$|R_i(x)| \leq K|\xi_i-x_i^*|^\gamma |x-x_i^*|;$$
$$\big|f(x_i^*)-\hat{f}_m(x_i^*)\big| \leq \|R_i\|_{L_\infty(\nu_0)} \quad \text{for } i=2,\dots,m-1;$$
$$\big|f(x)-\hat{f}_m(x)\big| \leq \begin{cases} 2\|R_i\|_{L_\infty(\nu_0)}+K|x_i^*-\eta_i|^\gamma|x-x_i^*| & \text{if } x\in J_i,\ i=3,\dots,m-1;\\ C|x-\tau_i| & \text{if } x\in J_i,\ i\in\{2,m\}\end{cases}$$

for some constant C and points $\xi_i\in J_i$, $\eta_i\in J_{i-1}\cup J_i\cup J_{i+1}$, $\tau_2\in J_2\cup J_3$ and $\tau_m\in J_{m-1}\cup J_m$.

Proof. By definition of R_i, we have

$$|R_i(x)| = \Big|\big(f'(\xi_i)-f'(x_i^*)\big)(x-x_i^*)\Big| \leq K|\xi_i-x_i^*|^\gamma|x-x_i^*|,$$

for some point $\xi_i\in J_i$. For the second inequality,

$$|f(x_i^*)-\hat{f}_m(x_i^*)| = \frac{1}{\nu_0(J_i)}\left|\int_{J_i}(f(x_i^*)-f(x))\nu_0(dx)\right|$$
$$= \frac{1}{\nu_0(J_i)}\left|\int_{J_i}R_i(x)\nu_0(dx)\right| \leq \|R_i\|_{L_\infty(\nu_0)},$$

where in the first inequality we have used the defining property of x_i^*. For the third inequality, let us start by proving that for all $2<i<m-1$, $\hat{f}_m'(x_i^*)=f'(\chi_i)$ for some $\chi_i\in J_i\cup J_{i+1}$ (here, we are considering right derivatives; for left ones, this would be $J_{i-1}\cup J_i$). To see that, take $x\in J_i\cap[x_i^*,x_{i+1}^*]$ and introduce the function $h(x):=f(x)-l(x)$ where

$$l(x)=\frac{x-x_i^*}{x_{i+1}^*-x_i^*}\big(\hat{f}_m(x_{i+1}^*)-\hat{f}_m(x_i^*)\big)+\hat{f}_m(x_i^*).$$

Then, using the fact that $\int_{J_i}(x-x_i^*)\nu_0(dx)=0$ joint with $\int_{J_{i+1}}(x-x_i^*)\nu_0(dx)=(x_{j+1}^*-x_j^*)\mu_m$, we get

$$\int_{J_i}h(x)\nu_0(dx)=0=\int_{J_{i+1}}h(x)\nu_0(dx).$$

In particular, by means of the mean theorem, one can conclude that there exist two points $p_i \in J_i$ and $p_{i+1} \in J_{i+1}$ such that

$$h(p_i) = \frac{\int_{J_i} h(x)\nu_0(dx)}{\nu_0(J_i)} = \frac{\int_{J_{i+1}} h(x)\nu_0(dx)}{\nu_0(J_{i+1})} = h(p_{i+1}).$$

As a consequence, we can deduce that there exists $\chi_i \in [p_i, p_{i+1}] \subseteq J_i \cup J_{i+1}$ such that $h'(\chi_i) = 0$, hence $f'(\chi_i) = l'(\chi_i) = \hat{f}'_m(x_i^*)$. When $2 < i < m-1$, the two Taylor expansions together with the fact that $\hat{f}'_m(x_i^*) = f'(\chi_i)$ for some $\chi_i \in J_i \cup J_{i+1}$, give

$$\begin{aligned} |f(x) - \hat{f}_m(x)| &\le |f(x_i^*) - \hat{f}_m(x_i^*)| + |R_i(x)| + K|x_i^* - \chi_i|^\gamma |x - x_i^*| \\ &\le 2\|R_i\|_{L_\infty(\nu_0)} + K|x_i^* - \chi_i|^\gamma |x - x_i^*| \end{aligned}$$

whenever $x \in J_i$ and $x > x_i^*$ (the case $x < x_i^*$ is handled similarly using the left derivative of $\hat{f}_m$ and $\xi_i \in J_{i-1} \cup J_i$). For the remaining cases, consider for example $i = 2$. Then $\hat{f}_m(x)$ is bounded by the minimum and the maximum of f on $J_2 \cup J_3$, hence $\hat{f}_m(x) = f(\tau)$ for some $\tau \in J_2 \cup J_3$. Since f' is bounded by $C = 2M + K$, one has $|f(x) - \hat{f}_m(x)| \le C|x - \tau|$. □

Lemma 5.3. *With the same notations as in Lemma 5.2, the estimates for $A_m^2(f)$, $B_m^2(f)$ and $L_2(f, \hat{f}_m)^2$ are as follows:*

$$\begin{aligned} L_2(f, \hat{f}_m)^2 \;&\le\; \frac{1}{4\kappa}\Bigg(\sum_{i=3}^{m} \int_{J_i} \Big(2\|R_i\|_{L_\infty(\nu_0)} + K|x_i^* - \eta_i|^\gamma |x - x_i^*|\Big)^2 \nu_0(dx)\Bigg) \\ &\quad + C^2\Big(\int_{J_2} |x - \tau_2|^2 \nu_0(dx) + \int_{J_m} |x - \tau_m|^2 \nu_0(dx)\Big). \\ A_m^2(f) &= L_2\big(\sqrt{f}, \widehat{\sqrt{f}}_m\big)^2 = O\Big(L_2(f, \hat{f}_m)^2\Big) \\ B_m^2(f) &= O\Bigg(\sum_{i=2}^{m} \frac{1}{\sqrt{\kappa}} \nu_0(J_i)(2\sqrt{M} + 1)^2 \|R_i\|^2_{L_\infty(\nu_0)}\Bigg). \end{aligned}$$

Proof. The L_2-bound is now a straightforward application of Lemmas 5.1 and 5.2. The one on $A_m(f)$ follows, since if $f \in \mathscr{F}^I_{(\gamma, K, \kappa, M)}$, then $\sqrt{f} \in \mathscr{F}^I_{(\gamma, \frac{K}{\sqrt{\kappa}}, \sqrt{\kappa}, \sqrt{M})}$. In order to bound $B_m^2(f)$ write it as:

$$B_m^2(f) = \sum_{j=1}^{m} \nu_0(J_j)\Bigg(\frac{\int_{J_j} \sqrt{f(y)}\nu_0(dy)}{\nu_0(J_j)} - \sqrt{\frac{\nu(J_j)}{\nu_0(J_j)}}\Bigg)^2 =: \sum_{j=1}^{m} \nu_0(J_j) E_j^2.$$

By the triangular inequality, let us bound E_j by $F_j + G_j$ where:

$$F_j = \left|\sqrt{\frac{\nu(J_j)}{\nu_0(J_j)}} - \sqrt{f(x_j^*)}\right| \quad \text{and} \quad G_j = \left|\sqrt{f(x_j^*)} - \frac{\int_{J_j} \sqrt{f(y)}\nu_0(dy)}{\nu_0(J_j)}\right|.$$

Using the same trick as in the proof of Lemma 5.1, we can bound:

$$F_j \le 2\sqrt{M}\left|\frac{\int_{J_j} \big(f(x) - f(x_i^*)\big)\nu_0(dx)}{\nu_0(J_j)}\right| \le 2\sqrt{M}\|R_j\|_{L_\infty(\nu_0)}.$$

On the other hand,

$$G_j = \frac{1}{\nu_0(J_j)}\left|\int_{J_j}\left(\sqrt{f(x_j^*)}-\sqrt{f(y)}\right)\nu_0(dy)\right| = \frac{1}{\nu_0(J_j)}\left|\int_{J_j}\left(\frac{f'(x_j^*)}{2\sqrt{f(x_j^*)}}(x-x_j^*)+\tilde{R}_j(y)\right)\nu_0(dy)\right| \le \|\tilde{R}_j\|_{L_\infty(\nu_0)},$$

which has the same magnitude as $\frac{1}{\kappa}\|R_j\|_{L_\infty(\nu_0)}$. □

Remark 5.4. Observe that when ν_0 is finite, there is no need for a special definition of $\hat{f}_m$ near 0, and all the estimates in Lemma 5.2 hold true replacing every occurrence of $i = 2$ by $i = 1$.

Remark 5.5. The same computations as in Lemmas 5.2 and 5.3 can be adapted to the general case where the V_j's (and hence $\hat{f}_m$) are not piecewise linear. In the general case, the Taylor expansion of $\hat{f}_m$ in x_i^* involves a rest as well, say $\hat{R}_i$, and one needs to bound this, as well.

5.2. *Proofs of Examples 2.4*

In the following, we collect the details of the proofs of Examples 2.4.

1. The finite case: $\nu_0 \equiv \mathrm{Leb}([0,1])$.

Remark that in the case where ν_0 if finite there are no convergence problems near zero and so we can consider the easier approximation of f:

$$\hat{f}_m(x) := \begin{cases} m\theta_1 & \text{if } x \in \left[0, x_1^*\right], \\ m^2\left[\theta_{j+1}(x-x_j^*)+\theta_j(x_{j+1}^*-x)\right] & \text{if } x \in (x_j^*, x_{j+1}^*] \quad j=1,\dots,m-1, \\ m\theta_m & \text{if } x \in (x_m^*, 1] \end{cases}$$

where

$$x_j^* = \frac{2j-1}{2m}, \qquad J_j = \left(\frac{j-1}{m}, \frac{j}{m}\right], \qquad \theta_j = \int_{J_j} f(x)dx, \quad j=1,\dots,m.$$

In this case we take $\varepsilon_m = 0$ and Conditions (C2) and (C2′) coincide:

$$\lim_{n\to\infty} n\Delta_n \sup_{f\in\mathscr{F}}\left(A_m^2(f)+B_m^2(f)\right) = 0.$$

Applying Lemma 5.3, we get

$$\sup_{f\in\mathscr{F}}\left(L_2(f,\hat{f}_m)+A_m(f)+B_m(f)\right) = O\left(m^{-\frac{3}{2}}+m^{-1-\gamma}\right);$$

(actually, each of the three terms on the left hand side has the same rate of convergence).

2. The finite variation case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-1}\mathbb{I}_{[0,1]}(x)$.

To prove that the standard choice of V_j described at the beginning of Examples 2.4 leads to $\int_{\varepsilon_m}^1 V_j(x)\frac{dx}{x} = 1$, it is enough to prove that this integral is independent of j, since in general $\int_{\varepsilon_m}^1 \sum_{j=2}^m V_j(x)\frac{dx}{x} = m-1$. To that aim observe that, for $j = 3,\dots,m-1$,

$$\mu_m\int_{\varepsilon_m}^1 V_j(x)\nu_0(dx) = \int_{x_{j-1}^*}^{x_j^*}\frac{x-x_{j-1}^*}{x_j^*-x_{j-1}^*}\frac{dx}{x} + \int_{x_j^*}^{x_{j+1}^*}\frac{x_{j+1}^*-x}{x_{j+1}^*-x_j^*}\frac{dx}{x}.$$

Let us show that the first addendum does not depend on j. We have

$$\int_{x^*_{j-1}}^{x^*_j} \frac{dx}{x^*_j - x^*_{j-1}} = 1 \quad \text{and} \quad -\frac{x^*_{j-1}}{x^*_j - x^*_{j-1}} \int_{x^*_{j-1}}^{x^*_j} \frac{dx}{x} = \frac{x^*_{j-1}}{x^*_j - x^*_{j-1}} \ln\Big(\frac{x^*_{j-1}}{x^*_j}\Big).$$

Since $x^*_j = \frac{v_j - v_{j-1}}{\mu_m}$ and $v_j = \varepsilon_m^{\frac{m-j}{m-1}}$, the quantities $\frac{x^*_j}{x^*_{j-1}}$ and, hence, $\frac{x^*_{j-1}}{x^*_j - x^*_{j-1}}$ do not depend on j. The second addendum and the trapezoidal functions V_2 and V_m are handled similarly. Thus, $\hat{f}_m$ can be chosen of the form

$$\hat{f}_m(x) := \begin{cases} 1 & \text{if } x \in [0, \varepsilon_m], \\ \dfrac{\nu(J_2)}{\mu_m} & \text{if } x \in (\varepsilon_m, x^*_2], \\ \dfrac{1}{x^*_{j+1} - x^*_j}\left[\dfrac{\nu(J_{j+1})}{\mu_m}(x - x^*_j) + \dfrac{\nu(J_j)}{\mu_m}(x^*_{j+1} - x)\right] & \text{if } x \in (x^*_j, x^*_{j+1}] \\ \qquad j = 2, \dots, m-1, & \\ \dfrac{\nu(J_m)}{\mu_m} & \text{if } x \in (x^*_m, 1]. \end{cases}$$

A straightforward application of Lemmas 5.2 and 5.3 gives

$$\sqrt{\int_{\varepsilon_m}^1 \Big(f(x) - \hat{f}_m(x)\Big)^2 \nu_0(dx) + A_m(f) + B_m(f)} = O\left(\left(\frac{\ln m}{m}\right)^{\gamma+1} \sqrt{\ln(\varepsilon_m^{-1})}\right),$$

as announced.

3. The infinite variation, non-compactly supported case: $\frac{d\nu_0}{d\mathrm{Leb}}(x) = x^{-2}\mathbb{I}_{\mathbb{R}_+}(x)$. Recall that we want to prove that

$$L_2(f, \hat{f}_m)^2 + A_m^2(f) + B_m^2(f) = O\left(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2\gamma}} + \sup_{x \geq H(m)} \frac{f(x)^2}{H(m)}\right),$$

for any given sequence $H(m)$ going to infinity as $m \to \infty$.

Let us start by addressing the problem that the triangular/trapezoidal choice for V_j is not doable. Introduce the following notation: $V_j = \overset{\triangle}{V}_j + A_j$, $j = 2, \dots, m$, where the $\overset{\triangle}{V}_j$'s are triangular/trapezoidal function similar to those in (6). The difference is that here, since x^*_m is not defined, $\overset{\triangle}{V}_{m-1}$ is a trapezoid, linear between x^*_{m-2} and x^*_{m-1} and constantly equal to $\frac{1}{\mu_m}$ on $[x^*_{m-1}, v_{m-1}]$ and $\overset{\triangle}{V}_m$ is supported on $[v_{m-1}, \infty)$, where it is constantly equal to $\frac{1}{\mu_m}$. Each A_j is chosen so that:

1. It is supported on $[x^*_{j-1}, x^*_{j+1}]$ (unless $j = 2$, $j = m-1$ or $j = m$; in the first case the support is $[x^*_2, x^*_3]$, in the second one it is $[x^*_{m-2}, x^*_{m-1}]$, and $A_m \equiv 0$);
2. A_j coincides with $-A_{j-1}$ on $[x^*_{j-1}, x^*_j]$, $j = 3, \dots, m-1$ (so that $\sum V_j \equiv \frac{1}{\mu_m}$) and its first derivative is bounded (in absolute value) by $\frac{1}{\mu_m(x^*_j - x^*_{j-1})}$ (so that V_j is non-negative and bounded by $\frac{1}{\mu_m}$);
3. A_j vanishes, along with its first derivatives, on x^*_{j-1}, x^*_j and x^*_{j+1}.

We claim that these conditions are sufficient to assure that $\hat{f}_m$ converges to f quickly enough. First of all, by Remark 5.5, we observe that, to have a good bound on $L_2(f, \hat{f}_m)$, the crucial property of $\hat{f}_m$ is that its first right (resp. left) derivative has to be equal to $\frac{1}{\mu_m(x^*_{j+1}-x^*_j)}$ (resp. $\frac{1}{\mu_m(x^*_j-x^*_{j-1})}$) and its second derivative has to be small enough (for example, so that the rest $\hat{R}_j$ is as small as the rest R_j of f already appearing in Lemma 5.2).

The (say) left derivatives in x^*_j of $\hat{f}_m$ are given by

$$\hat{f}'_m(x^*_j) = \big(\overset{\triangle}{V}_j{}'(x^*_j) + A'_j(x^*_j)\big)\big(\nu(J_j) - \nu(J_{j-1})\big); \qquad \hat{f}''_m(x^*_j) = A''_j(x^*_j)\big(\nu(J_j) - \nu(J_{j-1})\big).$$

Then, in order to bound $|\hat{f}''_m(x^*_j)|$ it is enough to bound $|A''_j(x^*_j)|$ because:

$$\begin{aligned}\big|\hat{f}''_m(x^*_j)\big| &\le |A''_j(x^*_j)|\left|\int_{J_j} f(x)\frac{dx}{x^2} - \int_{J_{j-1}} f(x)\frac{dx}{x^2}\right| \\ &\le |A''_j(x^*_j)| \sup_{x\in I}|f'(x)|(\ell_j + \ell_{j-1})\mu_m,\end{aligned}$$

where ℓ_j is the Lebesgue measure of J_j.

We are thus left to show that we can choose the A_j's satisfying points 1–3, with a small enough second derivative, and such that $\int_I V_j(x)\frac{dx}{x^2} = 1$. To make computations easier, we will make the following explicit choice:

$$A_j(x) = b_j(x - x^*_j)^2(x - x^*_{j-1})^2 \quad \forall x \in [x^*_{j-1}, x^*_j),$$

for some b_j depending only on j and m (the definitions on $[x^*_j, x^*_{j+1})$ are uniquely determined by the condition $A_j + A_{j+1} \equiv 0$ there).

Define $j_{\max}$ as the index such that $H(m) \in J_{j_{\max}}$; it is straightforward to check that

$$j_{\max} \sim m - \frac{\varepsilon_m(m-1)}{H(m)}; \qquad x^*_{m-k} = \varepsilon_m(m-1)\log\Big(1 + \frac{1}{k}\Big), \quad k = 1, \dots, m-2.$$

One may compute the following Taylor expansions:

$$\begin{aligned}\int_{x^*_{m-k-1}}^{x^*_{m-k}} \overset{\triangle}{V}_{m-k}(x)\nu_0(dx) &= \frac{1}{2} - \frac{1}{6k} + \frac{5}{24k^2} + O\Big(\frac{1}{k^3}\Big); \\ \int_{x^*_{m-k}}^{x^*_{m-k+1}} \overset{\triangle}{V}_{m-k}(x)\nu_0(dx) &= \frac{1}{2} + \frac{1}{6k} + \frac{1}{24k^2} + O\Big(\frac{1}{k^3}\Big).\end{aligned}$$

In particular, for $m \gg 0$ and $m-k \le j_{\max}$, so that also $k \gg 0$, all the integrals $\int_{x^*_{j-1}}^{x^*_{j+1}} \overset{\triangle}{V}_j(x)\nu_0(dx)$ are bigger than 1 (it is immediate to see that the same is true for $\overset{\triangle}{V}_2$, as well). From now on we will fix a $k \ge \frac{\varepsilon_m m}{H(m)}$ and let $j = m - k$.

Summing together the conditions $\int_I V_i(x)\nu_0(dx) = 1\ \forall i > j$ and noticing that the function $\sum_{i=j}^m V_i$ is constantly equal to $\frac{1}{\mu_m}$ on $[x^*_j, \infty)$ we have:

$$\int_{x^*_{j-1}}^{x^*_j} A_j(x)\nu_0(dx) = m - j + 1 - \frac{1}{\mu_m}\nu_0([x^*_j, \infty)) - \int_{x^*_{j-1}}^{x^*_j} \overset{\triangle}{V}_j(x)\nu_0(dx)$$

$$= k + 1 - \frac{1}{\log\left(1+\frac{1}{k}\right)} - \frac{1}{2} + \frac{1}{6k} + O\Big(\frac{1}{k^2}\Big) = \frac{1}{4k} + O\Big(\frac{1}{k^2}\Big).$$

Our choice of A_j allows us to compute this integral explicitly:

$$\int_{x^*_{j-1}}^{x^*_j} b_j (x - x^*_{j-1})^2 (x - x^*_j)^2 \frac{dx}{x^2} = b_j \big(\varepsilon_m (m-1)\big)^3 \Big(\frac{2}{3}\frac{1}{k^4} + O\Big(\frac{1}{k^5}\Big)\Big).$$

In particular one gets that asymptotically

$$b_j \sim \frac{1}{(\varepsilon_m(m-1))^3} \frac{3}{2} k^4 \frac{1}{4k} \sim \left(\frac{k}{\varepsilon_m m}\right)^3.$$

This immediately allows us to bound the first order derivative of A_j as asked in point 2: indeed, it is bounded above by $2b_j \ell_{j-1}^3$ where ℓ_{j-1} is again the length of J_{j-1}, namely $\ell_j = \frac{\varepsilon_m(m-1)}{k(k+1)} \sim \frac{\varepsilon_m m}{k^2}$. It follows that for m big enough:

$$\sup_{x\in I} |A_j'(x)| \le \frac{1}{k^3} \ll \frac{1}{\mu_m(x^*_j - x^*_{j-1})} \sim \left(\frac{k}{\varepsilon_m m}\right)^2.$$

The second order derivative of $A_j(x)$ can be easily computed to be bounded by $4b_j\ell_j^2$. Also remark that the conditions that $|f|$ is bounded by M and that f' is Hölder, say $|f'(x) - f'(y)| \le K|x-y|^\gamma$, together give a uniform L_∞ bound of $|f'|$ by $2M + K$. Summing up, we obtain:

$$|\hat f_m''(x^*_j)| \lesssim b_j \ell_m^3 \mu_m \sim \frac{1}{k^3 \varepsilon_m m}$$

(here and in the following we use the symbol $\lesssim$ to stress that we work up to constants and to higher order terms). The leading term of the rest $\hat R_j$ of the Taylor expansion of $\hat f_m$ near x^*_j is

$$\hat f_m''(x^*_j)|x - x^*_j|^2 \sim |f_m''(x^*_j)|\ell_j^2 \sim \frac{\varepsilon_m m}{k^7}.$$

Using Lemmas 5.2 and 5.3 (taking into consideration Remark 5.5) we obtain

$$\begin{aligned}
&\int_{\varepsilon_m}^{\infty} |f(x) - \hat f_m(x)|^2 \nu_0(dx) \\
&\lesssim \sum_{j=2}^{j_{\max}} \int_{J_j} |f(x) - \hat f_m(x)|^2 \nu_0(dx) + \int_{H(m)}^{\infty} |f(x) - \hat f_m(x)|^2 \nu_0(dx) \\
&\lesssim \sum_{k=\frac{\varepsilon_m m}{H(m)}}^{m} \mu_m \left(\frac{(\varepsilon_m m)^{2+2\gamma}}{k^{4+4\gamma}} + \frac{(\varepsilon_m m)^2}{k^{14}} \right) + \frac{1}{H(m)} \sup_{x \ge H(m)} f(x)^2 \\
&\lesssim \left(\frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}} + \frac{H(m)^{13}}{(\varepsilon_m m)^{10}} \right) + \frac{1}{H(m)}. \qquad (21)
\end{aligned}$$

It is easy to see that, since $0 < \gamma \le 1$, as soon as the first term converges, it does so more slowly than the second one. Thus, an optimal choice for $H(m)$ is given by $\sqrt{\varepsilon_m m}$, that gives a rate of

convergence:

$$L_2(f, \hat{f}_m)^2 \lesssim \frac{1}{\sqrt{\varepsilon_m m}}.$$

This directly gives a bound on $H(f, \hat{f}_m)$. Also, the bound on the term $A_m(f)$, which is $L_2(\sqrt{f}, \widehat{\sqrt{f}}_m)^2$, follows as well, since $f \in \mathscr{F}^I_{(\gamma,K,\kappa,M)}$ implies $\sqrt{f} \in \mathscr{F}^I_{(\gamma,\frac{K}{\sqrt{\kappa}},\sqrt{\kappa},\sqrt{M})}$. Finally, the term $B_m^2(f)$ contributes with the same rates as those in (21): using Lemma 5.3,

$$\begin{aligned} B_m^2(f) &\lesssim \sum_{j=2}^{\left\lceil m - \frac{\varepsilon_m(m-1)}{H(m)} \right\rceil} \nu_0(J_j)\|R_j\|_{L_\infty}^2 + \nu_0([H(m), \infty)) \\ &\lesssim \mu_m \sum_{k=\frac{\varepsilon_m(m-1)}{H(m)}}^{m} \left(\frac{\varepsilon_m m}{k^2}\right)^{2+2\gamma} + \frac{1}{H(m)} \\ &\lesssim \frac{H(m)^{3+4\gamma}}{(\varepsilon_m m)^{2+2\gamma}} + \frac{1}{H(m)}. \end{aligned}$$

5.3. Proof of Example 3.1

In this case, since $\varepsilon_m = 0$, the proofs of Theorems 2.5 and 2.6 simplify and give better estimates near zero, namely:

$$\begin{aligned} &\Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\nu_0}) \le C_1\left(\sqrt{T_n}\sup_{f\in\mathscr{F}}\left(A_m(f) + B_m(f) + L_2(f, \hat{f}_m)\right) + \sqrt{\frac{m^2}{T_n}}\right) \\ &\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\nu_0}) \\ &\quad \le C_2\left(\sqrt{n\Delta_n^2} + \frac{m\ln m}{\sqrt{n}} + \sqrt{T_n}\sup_{f\in\mathscr{F}}\left(A_m(f) + B_m(f) + H\big(f, \hat{f}_m\big)\right)\right), \end{aligned} \tag{22}$$

where C_1, C_2 depend only on κ, M and

$$A_m(f) = \sqrt{\int_0^1 \left(\widehat{\sqrt{f}}_m(y) - \sqrt{f(y)}\right)^2 dy}, \qquad B_m(f) = \sum_{j=1}^{m}\left(\sqrt{m}\int_{J_j}\sqrt{f(y)}dy - \sqrt{\theta_j}\right)^2.$$

As a consequence we get:

$$\Delta(\mathscr{P}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\nu_0}) \le O\left(\sqrt{T_n}\left(m^{-\frac{3}{2}} + m^{-1-\gamma}\right) + \sqrt{m^2 T_n^{-1}}\right).$$

To get the bounds in the statement of Example 3.1 the optimal choices are $m_n = T_n^{\frac{1}{2+\gamma}}$ when $\gamma \le \frac{1}{2}$ and $m_n = T_n^{\frac{2}{5}}$ otherwise. Concerning the discrete model, we have:

$$\Delta(\mathscr{Q}^{\mathrm{Leb}}_{n,FV}, \mathscr{W}_n^{\nu_0}) \le O\left(\sqrt{n\Delta_n^2} + \frac{m\ln m}{\sqrt{n}} + \sqrt{n\Delta_n}\left(m^{-\frac{3}{2}} + m^{-1-\gamma}\right)\right).$$

There are four possible scenarios: if $\gamma > \frac{1}{2}$ and $\Delta_n = n^{-\beta}$ with $\frac{1}{2} < \beta < \frac{3}{4}$ (resp. $\beta \geq \frac{3}{4}$), then the optimal choice is $m_n = n^{1-\beta}$ (resp. $m_n = n^{\frac{2-\beta}{5}}$).

If $\gamma \geq \frac{1}{2}$ and $\Delta_n = n^{-\beta}$ with $\frac{1}{2} < \beta < \frac{2+2\gamma}{3+2\gamma}$ (resp. $\beta \geq \frac{2+2\gamma}{3+2\gamma}$), then the optimal choice is $m_n = n^{\frac{2-\beta}{4+2\gamma}}$ (resp. $m_n = n^{1-\beta}$).

5.4. Proof of Example 3.2

As in Examples 2.4, we let $\varepsilon_m = m^{-1-\alpha}$ and consider the standard triangular/trapezoidal V_j's. In particular, $\hat{f}_m$ will be piecewise linear. Condition (C2′) is satisfied and we have $C_m(f) = O(\varepsilon_m)$. This bound, combined with the one obtained in (7), allows us to conclude that an upper bound for the rate of convergence of $\Delta(\mathscr{Q}^{\nu_0}_{n,FV}, \mathscr{W}^{\nu_0}_n)$ is given by:

$$\Delta(\mathscr{Q}^{\nu_0}_{n,FV}, \mathscr{W}^{\nu_0}_n) \leq C\left(\sqrt{\sqrt{n^2\Delta_n}\varepsilon_m} + \sqrt{n\Delta_n}\left(\frac{\ln(\varepsilon_m^{-1})}{m}\right)^2 + \frac{m\ln m}{\sqrt{n}} + \sqrt{n\Delta_n^2}\ln(\varepsilon_m^{-1})\right),$$

where C is a constant only depending on the bound on $\lambda > 0$.

The sequences ε_m and m can be chosen arbitrarily to optimize the rate of convergence. It is clear from the expression above that, if we take $\varepsilon_m = m^{-1-\alpha}$ with $\alpha > 0$, bigger values of α reduce the first term $\sqrt{\sqrt{n^2\Delta_n}\varepsilon_m}$, while changing the other terms only by constants. It can be seen that taking $\alpha \geq 15$ is enough to make the first term negligible with respect to the others. In that case, and under the assumption $\Delta_n = n^{-\beta}$, the optimal choice for m is $m = n^\delta$ with $\delta = \frac{5-4\beta}{14}$. In that case, the global rate of convergence is

$$\Delta(\mathscr{Q}^{\nu_0}_{n,FV}, \mathscr{W}^{\nu_0}_n) = \begin{cases} O\left(n^{\frac{1}{2}-\beta}\ln n\right) & \text{if } \dfrac{1}{2} < \beta \leq \dfrac{9}{10} \\ O\left(n^{-\frac{1+2\beta}{7}}\ln n\right) & \text{if } \dfrac{9}{10} < \beta < 1. \end{cases}$$

In the same way one can find

$$\Delta(\mathscr{P}^{\nu_0}_{n,FV}, \mathscr{W}^{\nu_0}_n) = O\left(\sqrt{n\Delta_n}\left(\frac{\ln m}{m}\right)^2\sqrt{\ln(\varepsilon_m^{-1})} + \sqrt{\frac{m^2}{n\Delta_n\ln(\varepsilon_m)}} + \sqrt{n\Delta_n}\varepsilon_m\right).$$

As above, we can freely choose ε_m and m (in a possibly different way from above). Again, as soon as $\varepsilon_m = m^{-1-\alpha}$ with $\alpha \geq 1$ the third term plays no role, so that we can choose $\varepsilon_m = m^{-2}$. Letting $\Delta_n = n^{-\beta}$, $0 < \beta < 1$, and $m = n^\delta$, an optimal choice is $\delta = \frac{1-\beta}{3}$, giving

$$\Delta(\mathscr{P}^{\nu_0}_{n,FV}, \mathscr{W}^{\nu_0}_n) = O\left(n^{\frac{\beta-1}{6}}(\ln n)^{\frac{5}{2}}\right) = O\left(T_n^{-\frac{1}{6}}(\ln T_n)^{\frac{5}{2}}\right).$$

5.5. Proof of Example 3.3

Using the computations in (21), combined with $\left(f(y) - \hat{f}_m(y)\right)^2 \leq 4\exp(-2\lambda_0 y^3) \leq 4\exp(-2\lambda_0 H(m)^3)$ for all $y \geq H(m)$, we obtain:

$$\int_{\varepsilon_m}^{\infty}\left|f(x) - \hat{f}_m(x)\right|^2\nu_0(dx) \lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \int_{H(m)}^{\infty}\left|f(x) - \hat{f}_m(x)\right|^2\nu_0(dx)$$

$$\lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \frac{e^{-2\lambda_0 H(m)^3}}{H(m)}.$$

As in Examples 2.4, this bounds directly $H^2(f, \hat{f}_m)$ and $A_m^2(f)$. Again, the first part of the integral appearing in $B_m^2(f)$ is asymptotically smaller than the one appearing above:

$$\begin{aligned} B_m^2(f) &= \sum_{j=1}^{m} \left(\frac{1}{\sqrt{\mu_m}} \int_{J_j} \sqrt{f} \nu_0 - \sqrt{\int_{J_j} f(x)\nu_0(dx)} \right)^2 \\ &\lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \sum_{k=1}^{\frac{\varepsilon_m m}{H(m)}} \left(\frac{1}{\sqrt{\mu_m}} \int_{J_{m-k}} \sqrt{f} \nu_0 - \sqrt{\int_{J_{m-k}} f(x)\nu_0(dx)} \right)^2 \\ &\lesssim \frac{H(m)^7}{(\varepsilon_m m)^4} + \frac{e^{-\lambda_0 H(m)^3}}{H(m)}. \end{aligned}$$

As above, for the last inequality we have bounded f in each J_{m-k}, $k \leq \frac{\varepsilon_m m}{H(m)}$, with $\exp(-\lambda_0 H(m)^3)$. Thus the global rate of convergence of $L_2(f, \hat{f}_m)^2 + A_m^2(f) + B_m^2(f)$ is $\frac{H(m)^7}{(\varepsilon_m m)^4} + \frac{e^{-\lambda_0 H(m)^3}}{H(m)}$.

Concerning $C_m(f)$, we have $C_m^2(f) = \int_0^{\varepsilon_m} \frac{(\sqrt{f(x)}-1)^2}{x^2} dx \lesssim \varepsilon_m^5$. To write the global rate of convergence of the Le Cam distance in the discrete setting we make the choice $H(m) = \sqrt[3]{\frac{\eta}{\lambda_0} \ln m}$, for some constant η, and obtain:

$$\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\left(\frac{\sqrt{n}\Delta_n}{\varepsilon_m} + \frac{m \ln m}{\sqrt{n}} + \sqrt{n\Delta_n} \left(\frac{(\ln m)^{\frac{7}{6}}}{(\varepsilon_m m)^2} + \frac{m^{-\frac{\eta}{2}}}{\sqrt[3]{\ln m}} \right) + \sqrt[4]{n^2 \Delta_n \varepsilon_m^5} \right).$$

Letting $\Delta_n = n^{-\beta}$, $\varepsilon_m = n^{-\alpha}$ and $m = n^{\delta}$, optimal choices give $\alpha = \frac{\beta}{3}$ and $\delta = \frac{1}{3} + \frac{\beta}{18}$. We can also take $\eta = 2$ to get a final rate of convergence:

$$\Delta(\mathscr{Q}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = \begin{cases} O\left(n^{\frac{1}{2} - \frac{2}{3}\beta}\right) & \text{if } \frac{3}{4} < \beta < \frac{12}{13} \\ O\left(n^{-\frac{1}{6} + \frac{\beta}{18}} (\ln n)^{\frac{7}{6}}\right) & \text{if } \frac{12}{13} \leq \beta < 1. \end{cases}$$

In the continuous setting, we have

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\left(\sqrt{n\Delta_n} \left(\frac{(\ln m)^{\frac{7}{6}}}{(\varepsilon_m m)^2} + \frac{m^{-\frac{\eta}{2}}}{\sqrt[3]{\ln m}} + \varepsilon_m^{\frac{5}{2}} \right) + \sqrt{\frac{\varepsilon_m m^2}{n\Delta_n}} \right).$$

Using $T_n = n\Delta_n$, $\varepsilon_m = T_n^{-\alpha}$ and $m = T_n^{\delta}$, optimal choices are given by $\alpha = \frac{4}{17}$, $\delta = \frac{9}{17}$; choosing any $\eta \geq 3$ we get the rate of convergence

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{W}_n^{\nu_0}) = O\left(T_n^{-\frac{3}{34}} (\ln T_n)^{\frac{7}{6}} \right).$$

Acknowledgments

I am very grateful to Markus Reiss for several interesting discussions and many insights; this paper would never have existed in the present form without his advice and encouragement. My

deepest thanks go to the anonymous referee, whose insightful comments have greatly improved the exposition of the paper; some gaps in the proofs have been corrected thanks to his/her remarks.

Appendix. Background

A.1. Le Cam theory of statistical experiments

A *statistical model* or *experiment* is a triplet $\mathscr{P}_j = (\mathscr{X}_j, \mathscr{A}_j, \{P_{j,\theta}; \theta \in \Theta\})$ where $\{P_{j,\theta}; \theta \in \Theta\}$ is a family of probability distributions all defined on the same σ-field $\mathscr{A}_j$ over the *sample space* $\mathscr{X}_j$ and Θ is the *parameter space*. The *deficiency* $\delta(\mathscr{P}_1, \mathscr{P}_2)$ of $\mathscr{P}_1$ with respect to $\mathscr{P}_2$ quantifies "how much information we lose" by using $\mathscr{P}_1$ instead of $\mathscr{P}_2$ and it is defined as $\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_K \sup_{\theta\in\Theta} \|K P_{1,\theta} - P_{2,\theta}\|_{TV}$, where TV stands for "total variation" and the infimum is taken over all "transitions" K (see [32, page 18]). The general definition of transition is quite involved but, for our purposes, it is enough to know that Markov kernels are special cases of transitions. By $K P_{1,\theta}$ we mean the image measure of $P_{1,\theta}$ via the Markov kernel K, that is

$$K P_{1,\theta}(A) = \int_{\mathscr{X}_1} K(x, A) P_{1,\theta}(dx), \quad \forall A \in \mathscr{A}_2.$$

The experiment $K\mathscr{P}_1 = (\mathscr{X}_2, \mathscr{A}_2, \{K P_{1,\theta}; \theta \in \Theta\})$ is called a *randomization* of $\mathscr{P}_1$ by the Markov kernel K. When the kernel K is deterministic, that is $K(x, A) = \mathbb{I}_A S(x)$ for some random variable $S : (\mathscr{X}_1, \mathscr{A}_1) \to (\mathscr{X}_2, \mathscr{A}_2)$, the experiment $K\mathscr{P}_1$ is called the *image experiment by the random variable* S. The Le Cam distance is defined as the symmetrization of δ and it defines a pseudometric. When $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$ the two statistical models are said to be *equivalent*. Two sequences of statistical models $(\mathscr{P}_1^n)_{n\in\mathbb{N}}$ and $(\mathscr{P}_2^n)_{n\in\mathbb{N}}$ are called *asymptotically equivalent* if $\Delta(\mathscr{P}_1^n, \mathscr{P}_2^n)$ tends to zero as n goes to infinity. A very interesting feature of the Le Cam distance is that it can be also translated in terms of statistical decision theory. Let $\mathscr{D}$ be any (measurable) decision space and let $L : \Theta \times \mathscr{D} \mapsto [0, \infty)$ denote a loss function. Let $\|L\| = \sup_{(\theta,z)\in\Theta\times\mathscr{D}} L(\theta, z)$. Let π_i denote a (randomized) decision procedure in the ith experiment. Denote by $R_i(\pi_i, L, \theta)$ the risk from using procedure π_i when L is the loss function and θ is the true value of the parameter. Then, an equivalent definition of the deficiency is:

$$\delta(\mathscr{P}_1, \mathscr{P}_2) = \inf_{\pi_1} \sup_{\pi_2} \sup_{\theta\in\Theta} \sup_{L:\|L\|=1} \left|R_1(\pi_1, L, \theta) - R_2(\pi_2, L, \theta)\right|.$$

Thus $\Delta(\mathscr{P}_1, \mathscr{P}_2) < \varepsilon$ means that for every procedure π_i in problem i there is a procedure π_j in problem j, $\{i, j\} = \{1, 2\}$, with risks differing by at most ε, uniformly over all bounded L and $\theta \in \Theta$. In particular, when minimax rates of convergence in a nonparametric estimation problem are obtained in one experiment, the same rates automatically hold in any asymptotically equivalent experiment. There is more: When explicit transformations from one experiment to another are obtained, statistical procedures can be carried over from one experiment to the other one.

There are various techniques to bound the Le Cam distance. We report below only the properties that are useful for our purposes. For the proofs see, e.g., [32,47].

Property A.1. *Let* $\mathscr{P}_j = (\mathscr{X}, \mathscr{A}, \{P_{j,\theta}; \theta \in \Theta\})$, $j = 1, 2$, *be two statistical models having the same sample space and define* $\Delta_0(\mathscr{P}_1, \mathscr{P}_2) := \sup_{\theta\in\Theta} \|P_{1,\theta} - P_{2,\theta}\|_{TV}$. *Then,* $\Delta(\mathscr{P}_1, \mathscr{P}_2) \le \Delta_0(\mathscr{P}_1, \mathscr{P}_2)$.

In particular, Property A.1 allows us to bound the Le Cam distance between statistical models sharing the same sample space by means of classical bounds for the total variation distance. To that aim, we collect below some useful results.

Fact A.2. *Let P_1 and P_2 be two probability measures on $\mathscr{X}$, dominated by a common measure ξ, with densities $g_i = \frac{dP_i}{d\xi}$, $i = 1, 2$. Define*

$$L_1(P_1, P_2) = \int_{\mathscr{X}} |g_1(x) - g_2(x)|\xi(dx),$$
$$H(P_1, P_2) = \left(\int_{\mathscr{X}} \left(\sqrt{g_1(x)} - \sqrt{g_2(x)}\right)^2 \xi(dx)\right)^{1/2}.$$

Then,

$$\|P_1 - P_2\|_{TV} = \frac{1}{2} L_1(P_1, P_2) \leq H(P_1, P_2). \tag{A.1}$$

Fact A.3. *Let P and Q be two product measures defined on the same sample space: $P = \otimes_{i=1}^n P_i$, $Q = \otimes_{i=1}^n Q_i$. Then*

$$H^2(P, Q) \leq \sum_{i=1}^n H^2(P_i, Q_i). \tag{A.2}$$

Fact A.4. *Let P_i, $i = 1, 2$, be the law of a Poisson random variable with mean λ_i. Then*

$$H^2(P_1, P_2) = 1 - \exp\left(-\frac{1}{2}\left(\sqrt{\lambda_1} - \sqrt{\lambda_2}\right)^2\right).$$

Fact A.5. *Let $Q_1 \sim \mathscr{N}(\mu_1, \sigma_1^2)$ and $Q_2 \sim \mathscr{N}(\mu_2, \sigma_2^2)$. Then*

$$\|Q_1 - Q_2\|_{TV} \leq \sqrt{2\left(1 - \frac{\sigma_1^2}{\sigma_2^2}\right)^2 + \frac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}}.$$

Fact A.6. *For $i = 1, 2$, let Q_i, $i = 1, 2$, be the law on $(C, \mathscr{C})$ of two Gaussian processes of the form*

$$X_t^i = \int_0^t h_i(s)ds + \int_0^t \sigma(s)dW_s, \quad t \in [0, T]$$

where $h_i \in L_2(\mathbb{R})$ and $\sigma \in \mathbb{R}_{>0}$. Then:

$$L_1\big(Q_1, Q_2\big) \leq \sqrt{\int_0^T \frac{\big(h_1(y) - h_2(y)\big)^2}{\sigma^2(s)} ds}.$$

Property A.7. *Let $\mathscr{P}_i = (\mathscr{X}_i, \mathscr{A}_i, \{P_{i,\theta}, \theta \in \Theta\})$, $i = 1, 2$, be two statistical models. Let $S : \mathscr{X}_1 \to \mathscr{X}_2$ be a sufficient statistics such that the distribution of S under $P_{1,\theta}$ is equal to $P_{2,\theta}$. Then $\Delta(\mathscr{P}_1, \mathscr{P}_2) = 0$.*

Remark A.8. Let P_i be a probability measure on $(E_i, \mathcal{E}_i)$ and K_i a Markov kernel on $(G_i, \mathcal{G}_i)$. One can then define a Markov kernel K on $(\prod_{i=1}^n E_i, \otimes_{i=1}^n \mathcal{G}_i)$ in the following way:

$$K(x_1, \dots, x_n; A_1 \times \cdots \times A_n) := \prod_{i=1}^n K_i(x_i, A_i), \quad \forall x_i \in E_i,\ \forall A_i \in \mathcal{G}_i.$$

Clearly $K \otimes_{i=1}^n P_i = \otimes_{i=1}^n K_i P_i$.

Finally, we recall the following result that allows us to bound the Le Cam distance between Poisson and Gaussian variables.

Theorem A.9 (*See [4, Theorem 4]*)**.** *Let $\tilde{P}_\lambda$ be the law of a Poisson random variable $\tilde{X}_\lambda$ with mean λ. Furthermore, let P_λ^* be the law of a random variable Z_λ^* with Gaussian distribution $\mathcal{N}(2\sqrt{\lambda}, 1)$, and let $\tilde{U}$ be a uniform variable on $\left[-\frac{1}{2}, \frac{1}{2}\right)$ independent of $\tilde{X}_\lambda$. Define*

$$\tilde{Z}_\lambda = 2sgn\big(\tilde{X}_\lambda + \tilde{U}\big)\sqrt{\big|\tilde{X}_\lambda + \tilde{U}\big|}. \tag{A.3}$$

Then, denoting by P_λ the law of $\tilde{Z}_\lambda$,

$$H^2\big(P_\lambda, P_\lambda^*\big) = O(\lambda^{-1}).$$

Remark A.10. Thanks to Theorem A.9, denoting by Λ a subset of $\mathbb{R}_{>0}$, by $\tilde{\mathscr{P}}$ (resp. $\mathscr{P}^*$) the statistical model associated with the family of probabilities $\{\tilde{P}_\lambda : \lambda \in \Lambda\}$ (resp. $\{P_\lambda^* : \lambda \in \Lambda\}$), we have

$$\Delta\big(\tilde{\mathscr{P}}, \mathscr{P}^*\big) \le \sup_{\lambda \in \Lambda} \frac{C}{\lambda},$$

for some constant C. Indeed, the correspondence associating $\tilde{Z}_\lambda$ to $\tilde{X}_\lambda$ defines a Markov kernel; conversely, associating to $\tilde{Z}_\lambda$ the closest integer to its square, defines a Markov kernel going in the other direction.

A.2. Lévy processes

Definition A.11. A stochastic process $\{X_t : t \ge 0\}$ on $\mathbb{R}$ defined on a probability space $(\Omega, \mathscr{A}, \mathbb{P})$ is called a *Lévy process* if the following conditions are satisfied.

1. $X_0 = 0$ $\mathbb{P}$-a.s.
2. For any choice of $n \ge 1$ and $0 \le t_0 < t_1 < \cdots < t_n$, random variables X_{t_0}, $X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent.
3. The distribution of $X_{s+t} - X_s$ does not depend on s.
4. There is $\Omega_0 \in \mathscr{A}$ with $\mathbb{P}(\Omega_0) = 1$ such that, for every $\omega \in \Omega_0$, $X_t(\omega)$ is right-continuous in $t \ge 0$ and has left limits in $t > 0$.
5. It is stochastically continuous.

Thanks to the *Lévy–Khintchine formula*, the characteristic function of any Lévy process $\{X_t\}$ can be expressed, for all u in $\mathbb{R}$, as:

$$\mathbb{E}\big[e^{iuX_t}\big] = \exp\left(-t\left(iub - \frac{u^2\sigma^2}{2} - \int_{\mathbb{R}} (1 - e^{iuy} + iuy\mathbb{I}_{|y|\le 1})\nu(dy)\right)\right),$$

where $b, \sigma \in \mathbb{R}$ and ν is a measure on $\mathbb{R}$ satisfying

$$\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}} (|y|^2 \wedge 1)\nu(dy) < \infty.$$

In the sequel we shall refer to (b, σ^2, ν) as the characteristic triplet of the process $\{X_t\}$ and ν will be called the *Lévy measure*. This data characterizes uniquely the law of the process $\{X_t\}$.

Let $D = D([0, \infty), \mathbb{R})$ be the space of mappings ω from $[0, \infty)$ into $\mathbb{R}$ that are right-continuous with left limits. Define the *canonical process* $x : D \to D$ by

$$\forall \omega \in D, \quad x_t(\omega) = \omega_t, \ \forall t \geq 0.$$

Let $\mathscr{D}_t$ and $\mathscr{D}$ be the σ-algebras generated by $\{x_s : 0 \leq s \leq t\}$ and $\{x_s : 0 \leq s < \infty\}$, respectively (here, we use the same notations as in [44]).

By the condition (4) above, any Lévy process on $\mathbb{R}$ induces a probability measure P on $(D, \mathscr{D})$. Thus $\{X_t\}$ on the probability space $(D, \mathscr{D}, P)$ is identical in law with the original Lévy process. By saying that $(\{x_t\}, P)$ is a Lévy process, we mean that $\{x_t : t \geq 0\}$ is a Lévy process under the probability measure P on $(D, \mathscr{D})$. For all $t > 0$ we will denote P_t for the restriction of P to $\mathscr{D}_t$. In the case where $\int_{|y|\leq 1} |y|\nu(dy) < \infty$, we set $\gamma^\nu := \int_{|y|\leq 1} y\nu(dy)$. Note that, if ν is a finite Lévy measure, then the process having characteristic triplet $(\gamma^\nu, 0, \nu)$ is a compound Poisson process.

Here and in the sequel we will denote by Δx_r the jump of process $\{x_t\}$ at the time r:

$$\Delta x_r = x_r - \lim_{s \uparrow r} x_s.$$

For the proof of Theorems 2.5 and 2.6 we also need some results on the equivalence of measures for Lévy processes. By the notation $\ll$ we will mean "is absolutely continuous with respect to".

Theorem A.12 (*See [44, Theorems 33.1–33.2], [45, Corollary 3.18, Remark 3.19]*)**.** *Let P^1 (resp. P^2) be the law induced on $(D, \mathscr{D})$ by a Lévy process of characteristic triplet $(\eta, 0, \nu_1)$ (resp. $(0, 0, \nu_2)$), where*

$$\eta = \int_{|y|\leq 1} y(\nu_1 - \nu_2)(dy) \tag{A.4}$$

is supposed to be finite. Then $P_t^1 \ll P_t^2$ for all $t \geq 0$ if and only if $\nu_1 \ll \nu_2$ and the density $\frac{d\nu_1}{d\nu_2}$ satisfies

$$\int \left(\sqrt{\frac{d\nu_1}{d\nu_2}(y)} - 1\right)^2 \nu_2(dy) < \infty. \tag{A.5}$$

Remark that the finiteness in (A.5) implies that in (A.4). When $P_t^1 \ll P_t^2$, the density is

$$\frac{dP_t^1}{dP_t^2}(x) = \exp(U_t(x)),$$

with

$$U_t(x) = \lim_{\varepsilon \to 0} \left(\sum_{r \leq t} \ln \frac{d\nu_1}{d\nu_2}(\Delta x_r)\mathbb{I}_{|\Delta x_r| > \varepsilon} - \int_{|y|>\varepsilon} t\left(\frac{d\nu_1}{d\nu_2}(y) - 1\right)\nu_2(dy)\right),$$

$$P^{(0,0,\nu_2)}\text{-a.s.} \tag{A.6}$$

The convergence in (A.6) is uniform in t on any bounded interval, $P^{(0,0,\nu_2)}$-a.s. Besides, $\{U_t(x)\}$ defined by (A.6) is a Lévy process satisfying $\mathbb{E}_{P^{(0,0,\nu_2)}}[e^{U_t(x)}] = 1$, $\forall t \geq 0$.

Finally, let us consider the following result giving an explicit bound for the L_1 and the Hellinger distances between two Lévy processes of characteristic triplets of the form $(b_i, 0, \nu_i)$, $i = 1, 2$ with $b_1 - b_2 = \int_{|y|\leq 1} y(\nu_1 - \nu_2)(dy)$.

Theorem A.13 (*See [30]*)**.** *For any $0 < T < \infty$, let P_T^i be the probability measure induced on $(D, \mathscr{D}_T)$ by a Lévy process of characteristic triplet $(b_i, 0, \nu_i)$, $i = 1, 2$ and suppose that $\nu_1 \ll \nu_2$.*

If $H^2(\nu_1, \nu_2) := \int \left(\sqrt{\frac{d\nu_1}{d\nu_2}(y)} - 1\right)^2 \nu_2(dy) < \infty$, then

$$H^2(P_T^1, P_T^2) \leq \frac{T}{2} H^2(\nu_1, \nu_2).$$

We conclude Appendix with a technical statement about the Le Cam distance for finite variation models.

Lemma A.14.

$$\Delta(\mathscr{P}_n^{\nu_0}, \mathscr{P}_{n,FV}^{\nu_0}) = 0.$$

Proof. Consider the Markov kernels π_1, π_2 defined as follows

$$\pi_1(x, A) = \mathbb{I}_A(x^d), \qquad \pi_2(x, A) = \mathbb{I}_A(x - \cdot\gamma^{\nu_0}), \quad \forall x \in D, A \in \mathscr{D},$$

where we have denoted by x^d the discontinuous part of the trajectory x, i.e. $\Delta x_r = x_r - \lim_{s\uparrow r} x_s$, $x_t^d = \sum_{r\leq t} \Delta x_r$ and by $x - \cdot\gamma^{\nu_0}$ the trajectory $x_t - t\gamma^{\nu_0}$, $t \in [0, T_n]$. On the one hand we have:

$$\begin{aligned}\pi_1 P^{(\gamma^{\nu-\nu_0},0,\nu)}(A) &= \int_D \pi_1(x, A) P^{(\gamma^{\nu-\nu_0},0,\nu)}(dx) = \int_D \mathbb{I}_A(x^d) P^{(\gamma^{\nu-\nu_0},0,\nu)}(dx) \\ &= P^{(\gamma^{\nu},0,\nu)}(A),\end{aligned}$$

where in the last equality we have used the fact that, under $P^{(\gamma^{\nu-\nu_0},0,\nu)}$, $\{x_t^d\}$ is a Lévy process with characteristic triplet $(\gamma^{\nu}, 0, \nu)$ (see [44, Theorem 19.3]). On the other hand:

$$\begin{aligned}\pi_2 P^{(\gamma^{\nu},0,\nu)}(A) &= \int_D \pi_2(x, A) P^{(\gamma^{\nu_0},0,\nu)}(dx) = \int_D \mathbb{I}_A(x - \cdot\gamma^{\nu_0}) P^{(\gamma^{\nu},0,\nu)}(dx) \\ &= P^{(\gamma^{\nu-\nu_0},0,\nu)}(A),\end{aligned}$$

since, by definition, $\gamma^{\nu} - \gamma^{\nu_0}$ is equal to $\gamma^{\nu-\nu_0}$. The conclusion follows by the definition of the Le Cam distance. □

References

[1] M. Bec, C. Lacour, Adaptive pointwise estimation for pure jump Lévy processes, Stat. Inference Stoch. Process. (2013) 1–28.

[2] D. Belomestny, F. Comte, V. Genon-Catalot, H. Masuda, M. Reiß, Lévy Matters IV: Estimation for Discretely Observed Lévy Processes, in: Lecture Notes in Mathematics, vol. 2128, Springer, 2015.

[3] L.D. Brown, T.T. Cai, M.G. Low, C.-H. Zhang, Asymptotic equivalence theory for nonparametric regression with random design, Ann. Statist. (3) (2002) 688–707, dedicated to the memory of Lucien Le Cam.

[4] L.D. Brown, A.V. Carter, M.G. Low, C.-H. Zhang, Equivalence theory for density estimation, Poisson processes and Gaussian white noise with drift, Ann. Statist (5) (2004) 2074–2097.
[5] L.D. Brown, M.G. Low, Asymptotic equivalence of nonparametric regression and white noise, Ann. Statist. (6) (1996) 2384–2398.
[6] L.D. Brown, C.-H. Zhang, Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2, Ann. Statist. (1) (1998) 279–287.
[7] B. Buchmann, G. Müller, Limit experiments of GARCH, Bernoulli (1) (2012) 64–99.
[8] P. Carr, H. Geman, D.B. Madan, M. Yor, The fine structure of asset returns: An empirical investigation, J. Bus. 75 (2) (2002) 305–333.
[9] A.V. Carter, Asymptotic approximation of nonparametric regression experiments with unknown variances, Ann. Statist. (4) (2007) 1644–1673.
[10] A.V. Carter, Deficiency distance between multinomial and multivariate normal experiments, Ann. Statist. (3) (2002) 708–730, dedicated to the memory of Lucien Le Cam.
[11] A.V. Carter, A continuous Gaussian approximation to a nonparametric regression in two dimensions, Bernoulli 12 (1) (2006) 143–156.
[12] A.V. Carter, Asymptotically sufficient statistics in nonparametric regression experiments with correlated noise, J. Probab. Stat. (2009) 19. Art. ID 275308.
[13] F. Comte, V. Genon-Catalot, Nonparametric adaptive estimation for pure jump Lévy processes, Ann. Inst. H. Poincaré Probab. Statist. 46 (3) (2010) 595–617.
[14] F. Comte, V. Genon-Catalot, Estimation for Lévy processes from high frequency data within a long time interval, Ann. Statist. 39 (2) (2011) 803–837.
[15] R. Cont, P. Tankov, Financial Modelling with Jump Processes, in: Chapman & Hall/CRC Financial Mathematics Series, Chapman & Hall/CRC, Boca Raton, FL, 2004.
[16] A. Dalalyan, M. Reiß, Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case, Probab. Theory Related Fields (1–2) (2007) 25–47.
[17] A. Dalalyan, M. Reiß, Asymptotic statistical equivalence for scalar ergodic diffusions, Probab. Theory Related Fields (2) (2006) 248–282.
[18] S. Delattre, M. Hoffmann, Asymptotic equivalence for a null recurrent diffusion, Bernoulli 8 (2) (2002) 139–174.
[19] C. Duval, Density estimation for compound Poisson processes from discrete data, Stochastic Process. Appl. (11) (2013) 3963–3986.
[20] S. Efromovich, A. Samarov, Asymptotic equivalence of nonparametric regression and white noise model has its limits, Statist. Probab. Lett. (2) (1996) 143–145.
[21] P. Étoré, S. Louhichi, E. Mariucci, Asymptotic equivalence of jumps Lévy processes and their discrete counterpart, ArXiv Preprint arXiv:1305.6725.
[22] J.E. Figueroa-López, Nonparametric estimation of Lévy models based on discrete-sampling, in: Optimality, in: IMS Lecture Notes Monogr. Ser., vol. 57, Inst. Math. Statist., Beachwood, OH, 2009, pp. 117–146.
[23] J.E. Figueroa-López, C. Houdré, Risk bounds for the non-parametric estimation of Lévy processes, in: High Dimensional Probability, in: IMS Lecture Notes Monogr. Ser., vol. 51, Inst. Math. Statist., Beachwood, OH, 2006, pp. 96–116.
[24] V. Genon-Catalot, C. Laredo, Asymptotic equivalence of nonparametric diffusion and Euler scheme experiments, Ann. Statist. 42 (3) (2014) 1145–1165.
[25] V. Genon-Catalot, C. Laredo, M. Nussbaum, Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift, Ann. Statist. (3) (2002) 731–753, dedicated to the memory of Lucien Le Cam.
[26] G.K. Golubev, M. Nussbaum, H.H. Zhou, Asymptotic equivalence of spectral density estimation and Gaussian white noise, Ann. Statist. (1) (2010) 181–214.
[27] I. Grama, M. Nussbaum, Asymptotic equivalence for nonparametric generalized linear models, Probab. Theory Related Fields (2) (1998) 167–214.
[28] I. Grama, M. Nussbaum, Asymptotic equivalence for nonparametric regression, Math. Methods Statist. 11 (1) (2002) 1–36.
[29] I.G. Grama, M.H. Neumann, Asymptotic equivalence of nonparametric autoregression and nonparametric regression, Ann. Statist. (4) (2006) 1701–1732.
[30] J. Jacod, A.N. Shiryaev, Limit theorems for stochastic processes, in: Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], second ed., Springer-Verlag, Berlin, 2003.
[31] M. Jähnisch, M. Nussbaum, Asymptotic equivalence for a model of independent non identically distributed observations, Statist. Decisions (3) (2003) 197–218.
[32] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, in: Springer Series in Statistics, Springer-Verlag, New York, 1986.

[33] L. Le Cam, G.L. Yang, Asymptotics in Statistics. Some Basic Concepts, second ed., in: Springer Series in Statistics, Springer-Verlag, New York, 2000.
[34] E. Mariucci, Asymptotic equivalence for inhomogeneous jump diffusion processes and white noise, ESAIM: Probab. Stat. (2015) in press.
[35] E. Mariucci, Asymptotic equivalence of discretely observed diffusion processes and their Euler scheme: small variance case, Stat. Inference Stoch. Process. (2015) in press.
[36] A. Meister, Asymptotic equivalence of functional linear regression and a white noise inverse problem, Ann. Statist. (3) (2011) 1471–1495.
[37] A. Meister, M. Reiß, Asymptotic equivalence for nonparametric regression with non-regular errors, Probab. Theory Related Fields (1–2) (2013) 201–229.
[38] G. Milstein, M. Nussbaum, Diffusion approximation for nonparametric autoregression, Probab. Theory Related Fields (4) (1998) 535–543.
[39] M. Nussbaum, Asymptotic equivalence of density estimation and Gaussian white noise, Ann. Statist. (6) (1996) 2399–2430.
[40] M. Reiß, Asymptotic equivalence for inference on the volatility from noisy observations, Ann. Statist. (2) (2011) 772–802.
[41] M. Reiß, Asymptotic equivalence for nonparametric regression with multivariate and random design, Ann. Statist. (4) (2008) 1957–1982.
[42] P. Reynaud-Bouret, Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities, Probab. Theory Related Fields 126 (1) (2003) 103–153.
[43] A. Rohde, On the asymptotic equivalence and rate of convergence of nonparametric regression and Gaussian white noise, Statist. Decisions (3) (2004) 235–243.
[44] K.-i. Sato, Lévy Processes and Infinitely Divisible Distributions, in: Cambridge Studies in Advanced Mathematics, vol. 68, Cambridge University Press, Cambridge, 1999, translated from the 1990 Japanese original, Revised by the author.
[45] K.-i. Sato, Density transformation in Lévy processes, 2000. Available online at: http://www.maphysto.dk/cgi-bin/gp.cgi?publ=218.
[46] J. Schmidt-Hieber, Asymptotic equivalence for regression under fractional noise, Ann. Statist. 42 (6) (2014) 2557–2585.
[47] H. Strasser, Mathematical theory of statistics, in: de Gruyter Studies in Mathematics, Walter de Gruyter & Co., Berlin, 1985.
[48] Y. Wang, Asymptotic nonequivalence of Garch models and diffusions, Ann. Statist. (3) (2002) 754–783, dedicated to the memory of Lucien Le Cam.