On Non-parametric Estimation of the Lévy Kernel of Markov processes

Florian A. J. Ueltzhöfer[*†]

We consider a recurrent Markov process which is an Itô semi-martingale. The Lévy kernel describes the law of its jumps. Based on observations $X_0, X_\Delta, \dots, X_{n\Delta}$, we construct an estimator for the Lévy kernel's density. We prove its consistency (as $n\Delta \to \infty$ and $\Delta \to 0$) and a central limit theorem. In the positive recurrent case, our estimator is asymptotically normal; in the null recurrent case, it is asymptotically mixed normal. Our estimator's rate of convergence equals the non-parametric minimax rate of smooth density estimation. The asymptotic bias and variance are analogous to those of the classical Nadaraya–Watson estimator for conditional densities. Asymptotic confidence intervals are provided.

AMS Subject Classification 2010: Primary 62M05; secondary 62G07, 60F05, 60J25
Keywords: Markov process, Itô semi-martingale, Lévy system, Lévy kernel, null recurrence, density estimation, central limit theorem

1. Introduction

Statistical inference for jumps in continuous-time models has received significant attention in recent years. Due to their well-known tractability properties, a vast amount of literature has been devoted to the class of processes with stationary and independent increments, called *Lévy processes*. The law of their jumps is characterised by their Lévy measure. Parametric inference for Lévy measures has a long history. For recent developments in non-parametric settings, we refer, for instance, to Comte and Genon-Catalot (2011); to Figueroa-López (2011); to the special issue Gugushvili, Klaassen, and Spreij (2010), which contains a collection of interesting papers; to Neumann and Reiß (2009);

*With the support of the Technische Universität München—Institute for Advanced Study, funded by the German Excellence Initiative; and support provided by the TUM International Graduate School of Science and Engineering (IGSSE).

†Lehrstuhl für mathematische Statistik and TUM Institute for Advanced Study, Technische Universität München, Boltzmannstraße 3, D–85 748 Garching b. M.; ueltzhoefer@ma.tum.de

and to Ueltzhöfer and Klüppelberg (2011). Ample references to previous literature can be found within the aforementioned.

In this paper, we consider a Harris recurrent Markov process X which is an Itô semi-martingale. Such a process is a solution of some stochastic differential equation

$$\begin{aligned} \mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t &+ \int \delta(X_{t-},y)\mathbb{1}_{\{\|\delta(X_{t-},y)\|>1\}}\mathfrak{p}(\mathrm{d}t,\mathrm{d}y) \\ &+ \int \delta(X_{t-},y)\mathbb{1}_{\{\|\delta(X_{t-},y)\|\leq 1\}}(\mathfrak{p}-\mathfrak{q})(\mathrm{d}t,\mathrm{d}y), \end{aligned} \tag{1.1}$$

with coefficients b, σ and δ; the SDE is driven by some Wiener process W and some Poisson random measure $\mathfrak{p}$ (with intensity measure $\mathfrak{q}(\mathrm{d}t,\mathrm{d}y) = \mathrm{d}t\otimes\lambda(\mathrm{d}y)$); X_{t-} denotes the left-limit. The law of its jumps is more or less described by the kernel F where, for each x, the measure $F(x,\cdot)$ coincides with the image of the measure λ under the map $y\mapsto\delta(x,y)$ restricted to the set $\{y:\delta(x,y)\neq 0\}$. We call F the *(canonical) Lévy kernel* of X. We assume that the measures $F(x,\mathrm{d}y)$ admit a density $y\mapsto f(x,y)$, and we aim for non-parametric estimation of the function $(x,y)\mapsto f(x,y)$.

On an equidistant time grid, we observe a sample $X_0(\omega), X_\Delta(\omega),\ldots,X_{n\Delta}(\omega)$ of the process; the jumps are latent. We study a kernel density estimator for $f(x,y)$. We show its consistency as $n\Delta\to\infty$ and $\Delta\to 0$ under a smoothness hypothesis on the estimated density. In the ergodic case, we obtain asymptotic normality. In the null recurrent case, we impose a condition on the resolvent of the process which goes back to Darling and Kac (1957). Thereunder, we prove asymptotic mixed normality. We also provide a standardised version of our central limit theorem for the construction of asymptotic confidence intervals.

Our results are comparable to those in classical non-parametric density estimation. In particular: Our estimator's asymptotic bias and variance resemble those of the Nadaraya–Watson estimator in classical conditional density estimation. Just as in the classical context, moreover, the bandwidth choice is crucial for our estimator's rate of convergence. We conjecture that, for instance, a cross-validation method applies here analogously; see Fan and Yim (2004) and Hall, Racine, and Li (2004). By an optimal choice, if $\Delta\to 0$ fast enough, the rate is $v(n\Delta)^{\alpha_1\alpha_2/[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]}$, where $\alpha_1>0$ (resp., $\alpha_2>0$) stands for the smoothness of f as a function in x (resp., in y), and the function v plays the role of an information rate. In the ergodic case, $v(t)=t$; in the null recurrent case with Darling–Kac's condition imposed, $v(t)=t^\delta\ell(t)$ for some $0<\delta\leq 1$ and some slowly varying function ℓ. We remark that, in the case $\alpha_1=\alpha_2$, our achieved rate $v(n\Delta)^{\alpha_1/(2\alpha_1+2d)}$ equals the non-parametric minimax rate of smooth density estimation, related to the smoothness of f as a $2d$-dimensional function and with respect to $v(n\Delta)$.

At the core of our statistical problem, we essentially have to study the case first, where the process is observed continuously in time and, in particular, all jumps are discerned. In this case, we can consider a more general class of quasi-left-continuous, strong Markov processes with càdlàg sample paths than just Itô semi-martingales. For these, the law of their jumps is again described by their Lévy kernel. We present a version of our estimator which utilises that the sojourn time of certain sets and the jumps are observed. Under slightly weaker assumptions, we prove the estimator's consistency and asymptotic (mixed) normality. As these results are valid for a quite general class of processes, we believe that they are of independent interest, not only as a benchmark for all possible estimators which are based on some discrete observation scheme.

For discrete-time Markov chains, a related result is presented in Karlsen and Tjøstheim (2001). We are aware that our final steps of proof appear to be similar. We emphasise that the main difficulties in our context, however, come in two respects: on the one hand, from establishing an appropriate auxiliary framework where related methods apply; on the other hand, from the discrete observation scheme where our primary objects of interest – the jumps – are latent.

For continuous-time Markov processes, apart from the Lévy process case and as far as known to us, estimation of their Lévy kernel has been confined to the special case of Markov step processes. For these, there is a one-to-one correspondence between the Lévy kernel and the infinitesimal generator. Efficient non-parametric estimation of Markov step process models has been studied by Greenwood and Wefelmeyer (1994). They assume the mean holding times to be bounded, and the transition kernel to be uniformly ergodic. This excludes the null recurrent case. The work on parametric estimation of Markov step processes is more exhaustive. The null recurrent case has been studied, for instance, by Höpfner (1993). There, the process is observed up to a random stopping time such that a deterministic amount of information (or more) has been discerned. Local asymptotic normality is shown in various situations. With a slightly different aim, in contrast, Höpfner, Jacod, and Ladelli (1990) considers Markov step processes observed up to a deterministic time. Accordingly, the observed amount of information is random. Local asymptotic mixed normality (of statistical experiments) is shown under Darling–Kac's condition. Here, we utilise some of their results and methods. We improve upon the restrictions within the aforementioned literature: First and foremost, we do not restrict ourselves to Markov step processes. Second, we consider processes, null recurrent in the sense of Harris, in a non-parametric setting. Third, we address the influence of observations on a discrete time grid.

We briefly outline our paper. In section 2 we study the estimation of the Lévy kernel based on discrete observations. Split into three subsections, we present the statistical problem with our standing assumptions; we give our estimator along with a bias correction; and state our main results – the estimator's consistency and the central limit theorem. In section 3, we study the case where continuous-time observations are available. This section is organised analogously to section 2. The corresponding proofs are in section 4. The proofs for our main results of section 2 are in section 5. Each proofs section comes with its own short outline at its beginning. Since we bring together potential theoretic aspects of Markov processes with functional and martingale limit theory, we put some of our technical considerations off to appendix A.

2. Density estimation of the Lévy kernel from high-frequency observations

2.1. Preliminaries and assumptions

On the filtered probability space(s) $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, (\mathbb{P}^x)_{x\in E})$, let $X = (X_t)_{t\geq 0}$ be a Markovian Itô semi-martingale with values in Euclidean space $E = (\mathbb{R}^d, \mathscr{B}^d)$, or a subset thereof, such that $\mathbb{P}^x(X_0 = x) = 1$ for all x. For $n \in \mathbb{N}$ and $\Delta > 0$, we observe $X_0(\omega)$ and the increments

$$\Delta_k^n X(\omega) := X_{k\Delta}(\omega) - X_{(k-1)\Delta}(\omega) \quad k = 1, \dots, n. \tag{2.1}$$

We emphasise that the jumps of the process are latent.

Throughout this paper, we use the following notation: We abbreviate $E^* := E \setminus \{0\}$. We denote the Dirac measure at x by ϵ_x. For π an (initial) probability on E, we denote the expectation w. r. t. the law $\mathbb{P}^\pi := \int \pi(\mathrm{d}x)\, \mathbb{P}^x$ by $\mathbb{E}^\pi$. For $\alpha \geq 0$ and $A \subseteq E$, in addition, $\mathcal{C}^\alpha_{\mathrm{loc}}(A)$ denotes the class of all continuous functions on A which are $\lfloor\alpha\rfloor$-times continuously differentiable such that every $x \in A$ has a neighbourhood on which the function's (partial) $\lfloor\alpha\rfloor$-derivatives are uniformly Hölder of order $\alpha - \lfloor\alpha\rfloor$.

The characteristics $(B, C, \mathfrak{n})$ of X are absolutely continuous with respect to Lebesgue measure; there are mappings $b : E \to E$ and $c : E \to E \otimes E$, and a kernel F on E with $F(x, \{0\}) = 0$ such that

$$B_t = \int_0^t b(X_s)\mathrm{d}s, \quad C_t = \int_0^t c(X_s)\mathrm{d}s, \quad \text{and} \quad \mathfrak{n}(\mathrm{d}t, \mathrm{d}y) = \mathrm{d}t \otimes F(X_t, \mathrm{d}y). \tag{2.2}$$

The integer-valued random measure $\sum_{\{s:\Delta X_s \neq 0\}} \epsilon_{(s,\Delta X_s)}(\mathrm{d}t, \mathrm{d}y)$ on $\mathbb{R}_+ \times \mathbb{R}^d$ is called the process's *jump measure*. The random measure $\mathfrak{n}$ is its predictable compensator: For every Borel function $g : E \times E \to \mathbb{R}_+$, (inital) probability π, and $t > 0$, we have

$$\mathbb{E}^\pi \sum_{0<s\leq t} g(X_{s-}, \Delta X_s)\mathbb{1}_{\{X_{s-}\neq X_s\}} = \mathbb{E}^\pi \int_0^t \mathrm{d}s \int_E F(X_s, \mathrm{d}y) g(X_s, y). \tag{2.3}$$

We call F the *Lévy kernel.* It is unique outside a set of potential zero. We assume it admits a density $(x, y) \mapsto f(x, y)$ which we want to estimate.

Throughout, we work under the following technical hypothesis on the characteristics:

2.1 Assumption.

(i) The process X satisfies the following (linear) growth condition: There exists a constant $\zeta < \infty$ and a Lévy measure $\bar{F}$ on E such that

$$\|b(x)\| \leq \zeta(1 + \|x\|), \quad \|c(x)\| \leq \zeta(1 + \|x\|^2), \quad \text{and} \quad F(x, A) \leq (1 + \|x\|)\bar{F}(A)$$

holds for all $x \in E$ and every Borel set $A \subseteq E$. We denote by $\beta \in [0, 2]$ some constant such that $\int \bar{F}(\mathrm{d}w)(\|w\|^\beta \wedge 1)) < \infty$.

(ii) The Lévy measure $\bar{F}$ admits a density $\bar{f}$ which is continuous on E^*.

(iii) There exists a constant $\zeta < \infty$ such that $\sup_{\|z\|>1} \|z\| \bar{f}(z) \leq \zeta$. $\diamond$

Remark. Apart from the growth condition, there is no assumption on b and c. Whether X is a weak or a strong solution of (1.1) is irrelevant to us.

We impose assumptions on the recurrence of X and on the smoothness of f. To obtain consistency for our estimator below, we impose:

2.2 Assumption. The process X is Harris recurrent: On E, there exists a σ-finite, invariant measure μ for X such that, for every Borel set $A \subseteq E$, we have

$$\mu(A) > 0 \implies \mathbb{P}^x \left(\int_0^\infty \mathbb{1}_A(X_s) \mathrm{d}s = \infty \right) = 1 \quad \forall x \in E.$$

$\diamond$

2.3 Assumption. For some $\alpha > 0$, the Lévy kernel admits a density $f \in \mathcal{C}^\alpha_{\mathrm{loc}}(E \times E^*)$; and the invariant measure from Assumption 2.2 admits a continuous density μ'. $\diamond$

To obtain a central limit theorem, we also impose:

2.4 Assumption. The process X satisfies the following Darling–Kac condition: For some $0 < \delta \leq 1$, there exists a function $v : \mathbb{R}_+ \to \mathbb{R}_+$ – at infinity, regularly varying of index δ – such that, for every μ-integrable g,

$$\frac{1}{v(1/\lambda)} \int_0^\infty \mathrm{e}^{-\lambda t}\, \mathbb{E}^x[g(X_t)] \mathrm{d}t \to \mu(g) \quad \mu\text{-a. e. as } \lambda \downarrow 0. \tag{2.4}$$

$\diamond$

Remark. In the positive recurrent case (that is, when μ is finite), Assumption 2.4 indeed is satisfied for $\delta = 1$ and with $v(t) = t/\mu(E)$. We refer the interested reader to Touati (1987) and to Höpfner and Löcherbach (2003).

2.5 Assumption. For some $\alpha_1, \alpha_2 \geq 2$, the Lévy kernel admits a density f which is twice continuously differentiable on $E \times E^*$ such that $x \mapsto f(x, y) \in \mathcal{C}^{\alpha_1}_{\mathrm{loc}}(E)$ for all $y \in E^*$, and $y \mapsto f(x, y) \in \mathcal{C}^{\alpha_2}_{\mathrm{loc}}(E^*)$ for all $x \in E$; and the invariant measure from Assumption 2.2 admits a continuous density μ' which is $(\lceil \alpha_1 \rceil - 1)$-times continuously differentiable. $\diamond$

Example. Suppose that f is bounded and vanishes outside $\{\|x\| \leq 1, \|y\| \leq 1\}$; that is, there are neither jumps with left-limit outside the unit ball nor jumps of size bigger than one. Then our process's recurrence (or transience) is completely determined by drift and volatility. For instance:

(i) If the volatility σ vanishes everywhere and the drift satisfies $b(x) = -x$, then X is positive recurrent.

(ii) If the drift b vanishes everywhere, and the volatility satisfies $\sigma(x) = 1$, then X is not positive. In fact, X has the recurrence (or transience) of Brownian motion: In the univariate case, X is null recurrent and Darling–Kac's condition holds with $\delta = 1/2$; in the bivariate case, X is null recurrent and Darling–Kac's condition fails; and in all other multivariate cases, X is transient.

2.2. Kernel density estimator

In principle, we are free to choose our favourite estimation method, e. g., the method of sieves with projection estimators. Here, however, we introduce a kernel density estimator as it allows for a more comprehensible presentation of the proofs. Also, the method is well-understood in the context of classical (conditional) density estimation.

An outline: First, we choose smooth kernels g_1 and g_2 with support $B_1(0)$ (the unit ball centred at zero) which are, at least, of order α_1 and α_2, respectively; that is, for every multi-index $m = (m_1, \dots, m_d) \in \mathbb{N}^d \setminus \{0\}$ and each $i \in \{1, 2\}$, we have

$$|m| := m_1 + \dots + m_d < \alpha_i \implies \kappa_m(g_i) := \int x_1^{m_1} \cdot \dots \cdot x_d^{m_d} g_i(x) \mathrm{d}x = 0. \tag{2.5}$$

Second, we choose a bandwidth vector $\eta = (\eta_1, \eta_2) > 0$. Last, we construct an estimator for $f(x, y)$ using the kernels $g_i^{\eta,x}(z) := \eta_i^{-d} g_i((z - x)/\eta_i)$. If the bandwidth is chosen appropriately, we achieve a consistent estimator which follows a central limit theorem.

2.7 Definition. For $\eta = (\eta_1, \eta_2) > 0$, we call $\hat{f}_n^{\Delta,\eta}$ defined by

$$\hat{f}_n^{\Delta,\eta}(x, y) := \begin{cases} \frac{\sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta}) g_2^{\eta,y}(\Delta_k^n X)}{\Delta \sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta})} & \text{if } \sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta}) > 0, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

the *kernel density estimator* of f (w. r. t. bandwidth η based on $X_0, X_\Delta, \dots, X_{n\Delta}$). ⋄

In analogy to classical conditional density estimation, we also introduce a bias correction for our estimator.

2.8 Definition. For $\eta = (\eta_1, \eta_2) > 0$, we call $\hat{\gamma}_n^{\Delta,\eta}$ defined by

$$\hat{\gamma}_n^{\Delta,\eta}(x, y) := \begin{cases} \eta_1^{\alpha_1} \sum_{\substack{|m_1+m_2|=\alpha_1 \\ |m_2|\neq 0}} \frac{\kappa_{m_1+m_2}(g_1)}{m_1! m_2!} \frac{\sum_{k=1}^n \frac{\partial^{m_1}}{\partial x^{m_1}} g_1^{\eta,x}(X_{(k-1)\Delta})}{\sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta})} \frac{\partial^{m_2}}{\partial x^{m_2}} \hat{f}_n^{\Delta,\eta}(x, y) & \\ \quad + \eta_2^{\alpha_2} \sum_{|m|=\alpha_2} \frac{\kappa_m(g_2)}{m!} \frac{\partial^m}{\partial y^m} \hat{f}_n^{\Delta,\eta}(x, y), & \text{if } \begin{matrix} \sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta}) > 0 \\ \alpha_1, \alpha_2 \in \mathbb{N}^* \end{matrix}, \\ 0, & \text{otherwise,} \end{cases}$$

the *bias correction* for $\hat{f}_n^{\Delta,\eta}$. (The sums in the previous equation are over all multi-indices of appropriate length.) ⋄

2.3. Consistency and central limit theorem

Here, we present our main results. We agree to the following conventions: Under Assumptions 2.2 and 2.4, v denotes the regularly varying function given in (2.4). Under Assumption 2.2 only, v denotes an arbitrary deterministic equivalent (see Definition 4.1 below) of the Markov process X. For typographical reasons, we may write v_t for $v(t)$ or $X(t)$ for X_t etc. as convenient.

We utilise the following conditions as $n\Delta \to \infty$ and $\Delta \to 0$, where $0 \le \zeta_1, \zeta_2 < \infty$:

$$v_{n\Delta}\eta_{1,n}^d\eta_{2,n}^d \to \infty, \quad \text{and} \quad \eta_{1,n} \to 0, \quad \eta_{2,n} \to 0; \tag{2.7}$$

$$v_{n\Delta}\eta_{1,n}^{d+2\alpha_1}\eta_{2,n}^d \to \zeta_1^2, \quad \text{and} \quad v_{n\Delta}\eta_{1,n}^d\eta_{2,n}^{d+2\alpha_2} \to \zeta_2^2; \tag{2.8}$$

In addition, we also utilise the following conditions due to discretisation, where $\zeta < \infty$ is independent of n:

$$\Delta\eta_{1,n}^{-2-d[(1-2/(\beta+d))\vee 0]} \to 0, \quad \text{and} \quad \Delta\eta_{2,n}^{-2\vee(\beta+d)} \to 0; \tag{2.9a}$$

$$n\Delta^2\eta_{1,n}^d\eta_{2,n}^d \le \zeta, \quad v_{n\Delta}\Delta^2\eta_{1,n}^{d-4-2d[(1-2/(\beta+d)\wedge 0]}\eta_{2,n}^d \to 0, \tag{2.9b}$$

$$\text{and} \quad v_{n\Delta}\Delta^2\eta_{1,n}^d\eta_{2,n}^{d-4\vee 2(\beta+d)} \to 0. \tag{2.9c}$$

Remark. If $\Delta \to 0$ fast enough, then (2.7) and (2.8) are the crucial conditions.

2.9 Theorem. *Grant Assumptions 2.1 to 2.3. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.7) and (2.9a) hold. Moreover, let $(x, y) \in E \times E^*$ be such that $\mu'(x) > 0$ and $F(x, E) > 0$.*

(i) If $n\Delta^2 \to 0$, then, under any law $\mathbb{P}^\pi$, we have the following convergence in probability:

$$\hat{f}_n^{\Delta,\eta_n}(x, y) \xrightarrow[n\to\infty]{\mathbb{P}^\pi} f(x, y). \tag{2.10}$$

(ii) Grant Assumption 2.4 in addition. If $(n\Delta)^{1-\delta}\Delta \to 0$, then, under any law $\mathbb{P}^\pi$, (2.10) holds as well.

Remark. By this theorem, our estimator is consistent for every x and $y \neq 0$ if $n\Delta \to \infty$ and $\Delta \to 0$. In practice, however, both n and Δ are given! Then, for instance, if a continuous martingale component is present, our estimator is unreliable for all y close to the origin. To illustrate this important point, suppose that X is a univariate process with constant volatility $\sigma^2 > 0$. Increments with absolute value less than $\zeta\sigma\Delta^{1/2}$, where ζ is quite a large constant (e. g., $\zeta = 5$), are predominantly due to the continuous martingale and not due to jumps. On the set $\{y : |y| \le \zeta\sigma\Delta^{1/2}\}$, therefore, our estimator $\hat{f}_n^{\Delta,\eta}(x,\cdot)$ is unreliable regardless of the chosen bandwidth η.

For the next theorem, we establish additional notation. For $0 < \alpha < 1$, let K denote the α-stable Lévy subordinator with Laplace transform $\mathbb{E}\, \mathrm{e}^{-\xi K_t} = \mathrm{e}^{-t\xi^\alpha}$ for $\xi, t \geq 0$. Its right inverse $L_t := \inf\{s > 0 : K_s > t\}$ is called the *Mittag-Leffler process of order* α. By abuse of notation, we call $L_t = t$ the *Mittag-Leffler process of order* 1. On an extension

$$(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{\mathbb{P}}) := (\Omega \times \Omega', \mathscr{F} \otimes \mathscr{F}', \mathbb{P}^\pi \otimes \mathbb{P}') \tag{2.11}$$

of the probability space, let $V = (V(x,y))_{x \in E, y \in E^*}$ be a standard Gaussian white noise random field (that is, the finite dimensional marginals of V are i. i. d. standard normal) and let $L = (L_t)_{t \geq 0}$ be the Mittag-Leffler process of order δ (from Assumption 2.4) such that V, L and $\mathscr{F}$ are independent. In the theorem below, convergence holds *stably in law*; that is, pre-limiting and limiting random variables are defined on the extended space (2.11) and we have joint convergence in law of our pre-limiting random variables with any bounded, $\mathscr{F}$-measurable random variable. This notion, labelled $\mathscr{L}$−st, is due to Renyi (1963).

2.10 Theorem. *Grant Assumptions 2.1 to 2.5. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.7) and (2.9) hold, and let $(x_i, y_i)_{i \in I}$ be a finite family of pairwise distinct points in $E \times E^*$ such that $\mu'(x_i) > 0$ and $F(x_i, E) > 0$ for each $i \in I$. If $(n\Delta)^{1-\delta}\Delta \to 0$, then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:*

$$\left(\sqrt{v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d}\Big(\hat{f}_n^{\Delta,\eta_n}(x_i,y_i) - \frac{\mu(g_1^{\eta_n,x_i} F g_2^{\eta_n,y_i})}{\mu(g_1^{\eta_n,x_i})}\Big)\right)_{i\in I} \xrightarrow[n\to\infty]{\mathscr{L}-\mathrm{st}} \left(\frac{\sigma(x_i,y_i)}{\sqrt{L_1}} V(x_i,y_i)\right)_{i\in I},$$

where the asymptotic variance is given by

$$\sigma(x,y)^2 := \frac{f(x,y)}{\mu'(x)} \int g_1(w)^2 \mathrm{d}w \int g_2(z)^2 \mathrm{d}z. \tag{2.12}$$

In addition, let η_n be such that (2.8) holds as well. Suppose either that $\alpha_1, \alpha_2 \in \mathbb{N}^*$ or that $\zeta_1 = \zeta_2 = 0$ in (2.8). Then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:

$$\left(\sqrt{v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d}\Big(\hat{f}_n^{\Delta,\eta_n}(x_i,y_i) - f(x_i,y_i)\Big)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\mathrm{st}} \left(\gamma(x_i,y_i) + \frac{\sigma(x_i,y_i)}{\sqrt{L_1}} V(x_i,y_i)\right)_{i\in I},$$

where – in the former case – the asymptotic bias $\gamma(x,y)$ is given by

$$\begin{aligned}\gamma(x,y) = \frac{\zeta_1}{\mu'(x)} \sum_{\substack{|m_1+m_2|=\alpha_1 \\ |m_2|\neq 0}} \frac{\kappa_{m_1+m_2}(g_1)}{m_1! m_2!} \frac{\partial^{m_1}}{\partial x^{m_1}} \mu'(x) \frac{\partial^{m_2}}{\partial x^{m_2}} f(x,y) \\ + \zeta_2 \sum_{|m|=\alpha_2} \frac{\kappa_m(g_2)}{m!} \frac{\partial^m}{\partial y^m} f(x,y),\end{aligned} \tag{2.13}$$

and – in the latter case – $\gamma(x,y) = 0$.

Remark. The asymptotic bias and variance of our estimator are analogous to those of the Nadaraya–Watson estimator in classical conditional density estimation: $\kappa_m(g_i)$ and $\int g_i(z)^2 \mathrm{d}z$ are the relevant *moment* and the *roughness* of the kernel g_i, respectively; and f (resp., μ') plays the role of the conditional (resp., marginal) density.

We recall that v from (2.4) satisfies $v_t = t$ in the ergodic case, and $v_t = t^\delta \ell(t)$ for some slowly varying function ℓ in the null recurrent case. If we choose $\eta_{i,n} = v_{n\Delta}^{-\xi_i}$ with $\xi_1 = \alpha_2/[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]$ and $\xi_2 = \alpha_1/[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]$, then (2.7) and (2.8) hold with $\zeta_1 = \zeta_2 = 1$. If $\Delta \to 0$ fast enough such that $n\Delta^{1+[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]/\zeta} \to 0$ in addition, where ζ denotes the maximum of $(1-\delta)d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2$, $\delta\alpha_1(\alpha_2+2+d)$ and $\delta\alpha_2(\alpha_1+2+d^2/(2+d))$, then our choice of η_n also satisfies (2.9) for every $\beta \leq 2$. Consequently, our estimator's rate of convergence is

$$v_{n\Delta}^{\alpha_1\alpha_2/[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]}. \tag{2.14}$$

In the case $\alpha_1 = \alpha_2$, the achieved rate $v_{n\Delta}^{\alpha/(2\alpha+2d)}$ equals the non-parametric minimax rate of smooth density estimation, related to the smoothness of f as a $2d$-dimensional function and w. r. t. $v_{n\Delta}$.

Remark. Bandwidth selection has always been a crucial issue in these kind of studies. Although orders of magnitude are crucial from an asymptotic point of view and $\eta_{i,n} = (n\Delta)^{-\xi_i}$ for some $\xi_i > 0$ may be a good choice , we note that, in practice, $\eta_{i,n} = \zeta(n\Delta)^{-\xi_i}$ with leading constant $\zeta \neq 1$ could be a better one. A detailed analysis would go beyond the scope of this paper. We briefly comment on two problems: How to choose the bandwidths manually such that conditions (2.7–2.9) are satisfied for the unknown $v_{n\Delta}$, α_1, α_2 and β? What needs to be considered when employing data-driven methods for selecting optimal bandwidths?

(i) Let $\alpha_0 \geq 2$ and $0 < \delta_0 \leq 1$ such that $\delta_0 > d/(d+\alpha_0)$. If we choose $\eta_{i,n} = (n\Delta)^{-1/(2d+2\alpha_0)}$, then (2.7) and (2.8) hold for all processes X such that Assumptions 2.4 and 2.5 hold for some $\alpha_1, \alpha_2 \geq \alpha_0$ and $\delta_0 < \delta \leq 1$. If $\Delta \to 0$ fast enough such that $n\Delta^{1+2[\alpha_0+d]/[\alpha_0+(2+d)\vee\alpha_0]} \to 0$ in addition, then our chosen bandwidth also satisfies (2.9).

(ii) The asymptotic bias and variance are proportional to the value of f and its derivatives at the point of interest. The optimal bandwidth choice in terms of the asymptotic mean squared error, therefore, may depend heavily on x and y. Especially for processes with infinite activity – where $y \mapsto f(x,y)$ has a pole at zero – this is an important issue in practice. In a future study on data-driven bandwidth selection methods like cross-validation, this distinction from estimating a bounded probability density has to be addressed carefully.

Theorem 2.10 does not allow for a direct construction of confidence intervals. For this purpose, we also obtain the following standardised version.

2.11 Corollary. *Grant Assumptions 2.1 to 2.5. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.7–2.9) hold. Suppose either that $\alpha_1, \alpha_2 \in \mathbb{N}^*$ or that $\zeta_1 = \zeta_2 = 0$ in (2.8). Then under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:*

$$\left(\sqrt{\frac{\eta_{1,n}^d \eta_{2,n}^d \Delta \sum_{k=1}^n g_1^{\eta_n, x_i}(X_{(k-1)\Delta})}{\xi_g^2 \hat{f}_n^{\Delta,\eta_n}(x_i, y_i)}} \left([\hat{f}_n^{\Delta,\eta_n} - \hat{\gamma}_n^{\eta_n} - f](x_i, y_i)\right)\right)_{i\in I} \xrightarrow[n\to\infty]{\mathscr{L}-\mathrm{st}} \left(V(x_i, y_i)\right)_{i\in I},$$

where $\xi_g^2 = \int g_1(w)^2 \mathrm{d}w \int g_2(z)^2 \mathrm{d}z$.

Remark. In principle, the results of this section are extendible to more general Markov models with Lévy kernel F such that (2.3) holds. In view of our proofs, the assumption that X is an Itô semi-martingale is crucial for the analysis of the influence of discretisation (see section 5.1). Suppose that an explicit upper bound for the small-time asymptotic "error"

$$\left|\frac{1}{\Delta}\mathbb{E}^x\left[g_2^{\eta,y}(\Delta_1^n X)\right] - \int F(x, \mathrm{d}w) g_2^{\eta,y}(w)\right|$$

and an explicit sufficient condition which ensures

$$\sup_{s\le 1} \frac{\xi_n}{v_{n\Delta}\eta_{1,n}^d} \left|\Delta \sum_{k=1}^{\lfloor sn\rfloor} h_n(X_{(k-1)\Delta}) - \int_0^{\lfloor sn\rfloor\Delta} h_n(X_r)\mathrm{d}r\right| \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0$$

for $\xi_n = 1$ or $\xi_n^2 = v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d$ are available for some Markov process X. Then it is straightforward (see Lemma 5.7 and (5.32) — Lemmata 5.6, 5.9 and 5.10, respectively) to come up with sufficient conditions for Theorems 2.9 and 2.10, which replace (2.9).

3. Density estimation of the Lévy kernel from continuous-time observations — A benchmark

The Lévy kernel of a Markov process is related with jumps. In fact, our estimator (2.6) uses $X_{(k-1)\Delta}$ and $\Delta_k^n X$ as proxies for the pre-jump value X_{t-} and the jump size ΔX_t if, at a time $t \in [(k-1)\Delta, k\Delta]$, there is a jump from a neighbourhood of x and of size close to y. Eventually, such time intervals contain either zero or one such jump; never more. Certainly, the statistical analysis simplifies if we observed the whole path of X; introducing proxies would be useless. So, despite observing the whole path of X is somewhat unrealistic, it is theoretically important to study what happens in this case. This section is devoted to this question and can be viewed as a benchmark for what properties are achievable with a more realistic, discrete observation scheme.

3.1. Preliminaries and assumptions

On the filtered probability space(s) $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\ge 0}, (\mathbb{P}^x)_{x\in E})$, let $X = (X_t)_{t\ge 0}$ be a strong Markov process with values in Euclidean space $E = (\mathbb{R}^d, \mathscr{B}^d)$, or a subset thereof. Its sample paths are supposed to be càdlàg. We observe – continuously in time – one sample path $\{X_s(\omega) : s \in [0,t]\}$ for $t > 0$; in particular, we discern all jumps.

In addition to the notation introduced before, we use some classical notation from Getoor (1975): We denote the shift semi-group on Ω by $(\theta_t)_{t\geq 0}$ so that $X_{t+s} = X_t \circ \theta_s$ for all $s, t \geq 0$. We denote the transition semi-group of X on E by $(P_t)_{t\geq 0}$.

A *(perfect homogeneous) additive functional* H of X is an $\mathscr{F}_t$-adapted process such that $H_{t+s} = H_t \circ \theta_s + H_s$ for all $s, t \geq 0$. A *Lévy system* (F, H) of X (in a wide sense) is a kernel F on E with $F(x, \{0\}) = 0$ and a non-decreasing additive functional H of X such that, for every Borel function $g : E \times E \to \mathbb{R}_+$, probability π on E, and $t > 0$,

$$\mathbb{E}^\pi \sum_{0<s\leq t} g(X_{s-}, \Delta X_s) \mathbb{1}_{\{X_{s-}\neq X_s\}} = \mathbb{E}^\pi \int_0^t \mathrm{d}H_s \int_E F(X_s, \mathrm{d}y) g(X_s, y). \tag{3.1}$$

The disintegration into F and H is by no means unique. For an appropriate reference function g_0 with $Fg_0(x) > 0$, nevertheless, ratios of the form $Fg(x)/Fg_0(x)$ are unique outside a set of potential zero. In the cases where X is quasi-left-continuous (that is, when all jump times are totally inaccessible) Benveniste and Jacod (1973) proved the existence of a Lévy system (F, H) where H is continuous. Such a process – càdlàg, strong Markov, quasi-left-continuous – is called a *Hunt process*.

Remark. The continuity of the additive functional was included as a part of the original definition of Lévy systems due to Watanabe (1964).

Throughout this section, we work under the following hypothesis:

3.1 Assumption. There exists a Lévy system (F, H) of X where $H_t = t$. ⋄

Recalling (2.3), we observe that all Markovian Itô semi-martingales satisfy Assumption 3.1. In analogy to the semi-martingale case, we call this F in Assumption 3.1 the *(canonical) Lévy kernel* of X. It is unique outside a set of potential zero. Again, we assume it admits a density $(x, y) \mapsto f(x, y)$ which we want to estimate.

Compared to section 2, we slightly weaken the assumptions imposed on the smoothness of f. To obtain consistency for our estimator below, we impose Assumption 2.2 and:

3.2 Assumption. The canonical Lévy kernel admits a density f, continuous on $E \times E^*$; and the invariant measure from Assumption 2.2 admits a continuous density μ'. ⋄

To obtain a central limit theorem, we also impose Assumption 2.4 and:

3.3 Assumption. For some $\alpha_1, \alpha_2 > 0$, the canonical Lévy kernel admits a density f such that $x \mapsto f(x, y) \in \mathcal{C}^{\alpha_1}_{\text{loc}}(E)$ for all $y \in E^*$, and $y \mapsto f(x, y) \in \mathcal{C}^{\alpha_2}_{\text{loc}}(E^*)$ for all $x \in E$; and the invariant measure from Assumption 2.2 admits a continuous density μ' which is $(\lceil\alpha_1\rceil - 1)$-times continuously differentiable. ⋄

3.2. Kernel density estimator

In section 2.2, we introduced a kernel density estimator and its bias correction based on discrete observations. Here, we present corresponding versions which utilise the continuous-time observation scheme. We recall that g_1 and g_2 are kernels with support $B_1(0)$ which are, at least, of order α_1 and α_2, respectively. Given some bandwidth vector $\eta = (\eta_1, \eta_2) > 0$, we utilise the kernels $g_i^{\eta,x}(z) = \eta_i^{-d} g_i((z-x)/\eta_i)$.

3.4 Definition. For $\eta = (\eta_1, \eta_2) > 0$, we call $\hat{f}_t^\eta$ defined by

$$\hat{f}_t^\eta(x,y) := \begin{cases} \frac{\sum_{0<s\le t} g_1^{\eta,x}(X_{s-}) g_2^{\eta,y}(\Delta X_s) \mathbb{1}_{\{X_{s-}\neq X_s\}}}{\int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s} & \text{if } \int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s > 0, \\ 0 & \text{otherwise,} \end{cases}$$

the *kernel density estimator* of f (w. r. t. bandwidth η up to time t). ⋄

Our estimator in Definition 2.7 is the discretised analogue from the one presented here: In the numerator of the former, the jumps ΔX_t and the pre-jump left-limits X_{t-} are replaced by the increments $\Delta_k^n X$ and the pre-increment values $X_{(k-1)\Delta}$, respectively. In the denominator, the sojourn time $\int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s$ is replaced by its Riemann sum approximation $\Delta \sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta})$. In analogy to Definition 2.8, we also introduce a bias correction for our estimator:

3.5 Definition. For $\eta = (\eta_1, \eta_2) > 0$, we call $\hat{\gamma}_t^\eta$ defined by

$$\hat{\gamma}_t^\eta(x,y) := \begin{cases} \eta_1^{\alpha_1} \sum_{\substack{|m_1+m_2|=\alpha_1 \\ |m_2|\neq 0}} \frac{\kappa_{m_1+m_2}(g_1)}{m_1! m_2!} \frac{\int_0^t \frac{\partial^{m_1}}{\partial x^{m_1}} g_1^{\eta,x}(X_s)\mathrm{d}s}{\int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s} \frac{\partial^{m_2}}{\partial x^{m_2}} \hat{f}_t^\eta(x,y) & \\ \quad + \eta_2^{\alpha_2} \sum_{|m|=\alpha_2} \frac{\kappa_m(g_2)}{m!} \frac{\partial^m}{\partial y^m} \hat{f}_t^\eta(x,y), & \text{if } \begin{array}{c} \int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s > 0 \\ \alpha_1, \alpha_2 \in \mathbb{N}^* \end{array}, \\ 0, & \text{otherwise,} \end{cases}$$

the *bias correction* for $\hat{f}_t^\eta$. ⋄

3.3. Consistency and central limit theorem

Here, we present our results of this section. We continue to use the notation and conventions from section 2.3.

We utilise the following conditions as $t \to \infty$, where $0 \le \zeta_1, \zeta_2 < \infty$:

$$v_t \eta_{1,t}^d \eta_{2,t}^d \to \infty, \quad \text{and} \quad \eta_{1,t} \to 0, \eta_{2,t} \to 0; \tag{3.2}$$

$$v_t \eta_{1,t}^{d+2\alpha_1} \eta_{2,t}^d \to \zeta_1^2, \quad \text{and} \quad v_t \eta_{1,t}^d \eta_{2,t}^{d+2\alpha_2} \to \zeta_2^2. \tag{3.3}$$

3.6 Theorem. *Grant Assumptions 2.2, 3.1 and 3.2. Let $\eta_t = (\eta_{1,t}, \eta_{2,t})$ be such that (3.2) holds. Moreover, let $(x, y) \in E \times E^*$ be such that $\mu'(x) > 0$ and $F(x, E) > 0$. Then, under any law $\mathbb{P}^\pi$, we have the following convergence in probability:*

$$\hat{f}_t^{\eta_t}(x, y) \xrightarrow[t\to\infty]{\mathbb{P}^\pi} f(x, y).$$

3.7 Theorem. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let $\eta_t = (\eta_{1,t}, \eta_{2,t})$ be such that (3.2) holds. Moreover, let $(x_i, y_i)_{i\in I}$ be a finite family of pairwise distinct points in $E \times E^*$ such that $\mu'(x_i) > 0$ and $F(x_i, E) > 0$ for each $i \in I$. Then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:*

$$\left(\sqrt{v_t \eta_{1,t}^d \eta_{2,t}^d}\left(\hat{f}_t^{\eta_t}(x_i, y_i) - \frac{\mu(g_1^{\eta_t, x_i} F g_2^{\eta_t, y_i})}{\mu(g_1^{\eta_t, x_i})}\right)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\mathrm{st}} \left(\frac{\sigma(x_i, y_i)}{\sqrt{L_1}} V(x_i, y_i)\right)_{i\in I},$$

where the asymptotic variance is given by

$$\sigma(x, y)^2 := \frac{f(x, y)}{\mu'(x)} \int g_1(w)^2 \mathrm{d}w \int g_2(z)^2 \mathrm{d}z. \tag{3.4}$$

In addition, grant Assumption 3.3 and let η_t be such that (3.3) holds as well. Suppose either that $\alpha_1, \alpha_2 \in \mathbb{N}^*$ or that $\zeta_1 = \zeta_2 = 0$ in (3.3). Then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:

$$\left(\sqrt{v_t \eta_{1,t}^d \eta_{2,t}^d}\big(\hat{f}_t^{\eta_t}(x_i, y_i) - f(x_i, y_i)\big)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\mathrm{st}} \left(\gamma(x_i, y_i) + \frac{\sigma(x_i, y_i)}{\sqrt{L_1}} V(x_i, y_i)\right)_{i\in I}, \tag{3.5}$$

where – in the former case – the asymptotic bias $\gamma(x, y)$ is given by

$$\gamma(x, y) = \frac{\zeta_1}{\mu'(x)} \sum_{\substack{|m_1+m_2|=\alpha_1 \\ |m_2|\neq 0}} \frac{\kappa_{m_1+m_2}(g_1)}{m_1! m_2!} \frac{\partial^{m_1}}{\partial x^{m_1}} \mu'(x) \frac{\partial^{m_2}}{\partial x^{m_2}} f(x, y) + \zeta_2 \sum_{|m|=\alpha_2} \frac{\kappa_m(g_2)}{m!} \frac{\partial^m}{\partial y^m} f(x, y), \tag{3.6}$$

and – in the latter case – $\gamma(x, y) = 0$.

We compare Theorems 2.10 and 3.7. First, we remark that the asymptotic bias and variance of $\hat{f}_n^{\Delta,\eta}$ are equal to those of our benchmark estimator $\hat{f}_t^\eta$. Second, if we choose $\eta_{i,t} = v_t^{-\xi_i}$ with $\xi_1 = \alpha_2/[d(\alpha_1 + \alpha_2) + 2\alpha_1\alpha_2]$ and $\xi_2 = \alpha_1/[d(\alpha_1 + \alpha_2) + 2\alpha_1\alpha_2]$ again, then (3.2) and (3.3) hold with $\zeta_1 = \zeta_2 = 1$. The rate of convergence in Theorem 3.7 is

$$v_t^{\alpha_1\alpha_2/[d(\alpha_1+\alpha_2)+2\alpha_1\alpha_2]}; \tag{3.7}$$

the rates (2.14) and (3.7) are equivalent. Third, we observe that our remark on the issue of bandwidth selection holds analogously. Last, we note that Theorem 3.7 does not allow for a direct construction of confidence intervals just as Theorem 2.10. In analogy to Corollary 2.11, we also obtain the following standardised version.

3.8 Corollary. *Grant Assumptions 2.2, 2.4 and 3.1 to 3.3. Let $\eta_t = (\eta_{1,t}, \eta_{2,t})$ be such that (3.2) and (3.3) hold. Suppose either that $\alpha_1, \alpha_2 \in \mathbb{N}^*$ or that $\zeta_1 = \zeta_2 = 0$ in (3.3). Then under any law $\mathbb{P}^\pi$, we have the following stable convergence in law:*

$$\left(\sqrt{\frac{\eta_{1,t}^d \eta_{2,t}^d \int_0^t g_1^{\eta_t,x_i}(X_s)\mathrm{d}s}{\xi_g^2 \hat{f}_t^{\eta_t}(x_i,y_i)}}\Big([\hat{f}_t^{\eta_t} - \hat{\gamma}_t^{\eta_t}](x_i,y_i) - f(x_i,y_i)\Big)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\mathrm{st}} \Big(V(x_i,y_i)\Big)_{i\in I},$$

where $\xi_g^2 = \int g_1(w)^2\mathrm{d}w \int g_2(z)^2\mathrm{d}z$.

4. Proofs for results of section 3

The notion of a deterministic equivalent of a Markov process plays a crucial role in the limit theory for our estimator.

4.1 Definition. A non-decreasing function $v : \mathbb{R}_+ \to \mathbb{R}_+$ is called a *deterministic equivalent* of the Markov process X if the families

$$\Big\{\mathscr{L}(v(t)^{-1}H_t \mid \mathbb{P}^\pi) : t > 0\Big\} \quad \text{and} \quad \Big\{\mathscr{L}(v(t)H_t^{-1} \mid \mathbb{P}^\pi) : t > 0\Big\}$$

are tight for every probability π on E and every non-decreasing additive functional H of X with $0 < \mathbb{E}^\mu H_1 < \infty$. ⋄

We emphasise the following consequence of Théorème 3 of Touati (1987): Under Darling–Kac's condition, the function v in (2.4) is a deterministic equivalent of X. For every H as in Definition 4.1, furthermore, we have that $(v(t)^{-1}H_{st})_{s\geq 0}$ converges in law to a non-trivial process as $t \to \infty$. For Markov processes violating Darling–Kac's condition, the latter convergence may not hold. Nevertheless, Löcherbach and Loukianova (2008) showed that some deterministic equivalent already exists when X is Harris recurrent.

Throughout the proofs, we denote convergence of processes by double arrow ("$\Rightarrow$") and understand it as convergence on the relevant Skorokhod space. For instance, we denote by $\mathcal{D}(\mathbb{R}^d) := \mathcal{D}(\mathbb{R}_+;\mathbb{R}^d)$ the space of all càdlàg functions from $\mathbb{R}_+$ to $\mathbb{R}^d$ equipped with Skorokhod's topology. For a kernel F, a measurable function g, and a σ-finite measure ν, the function Fg, the measure νF, and the number $\nu(g)$ are given by

$$Fg(x) := \int F(x,\mathrm{d}y)g(y), \quad \nu F(A) := \int \nu(\mathrm{d}x)F(x,A), \quad \nu(g) := \int \nu(\mathrm{d}x)g(x).$$

A kernel F is called *strong Feller* if Fg is in the class of continuous functions for every bounded g.

This section is organised as follows: First, in section 4.1 we prove a triangular array extension of Birkhoff's theorem for additive functionals. Second, in section 4.2 we introduce auxiliary Markov chains Z and Z' derived from our Markov process X. We show that our result from section 4.1 applies to these chains. Some technicalities are put off to appendix A. Third, in section 4.3 we demonstrate a preliminary version of Theorem 3.6 which depends only on Z and Z'; we conclude with the final steps in the proof of consistency. Last, in section 4.4 we demonstrate a preliminary central limit theorem which depends only on Z and Z'; we conclude with the final steps in the proof of Theorem 3.7 and Corollary 3.8.

4.1. An extension of Birkhoff's theorem

The theorem presented in this subsection is the underlying key result for our proofs. It is a triangular array extension of Birkhoff's theorem for additive functionals (cf. Théorème II.2 of Azéma, Kaplan-Duflo, and Revuz, 1967). We prove a rather general version.

4.2 Theorem. *Let $Z = (Z_k)_{k\in\mathbb{N}^*}$ be a Markov chain with values in some state space D, with invariant probability ψ, and with transition kernel Ψ. Assume that the state space is* petite*, that is, there exist a probability ρ on $\mathbb{N}^*$ and a non-trivial measure ν_ρ on D such that, for every Borel set $A \subseteq D$,*

$$\inf_{x\in D} \sum_{k=1}^{\infty} \rho(k)\Psi^k(x, A) \geq \nu_\rho(A).$$

Let $(h_n)_{n\in\mathbb{N}^*}$ be a sequence of functions such that $(\Psi h_n)_{n\in\mathbb{N}^*}$ is uniformly bounded. Let $\xi_n > 0$ be such that

$$n\xi_n \to \infty, \quad \xi_n^{-1}\psi(h_n) \to c < \infty, \quad (n\xi_n^2)^{-1}\psi(|h_n|) \to 0 \quad \textit{and} \quad (n\xi_n^2)^{-1}\psi(h_n^2) \to 0$$

as $n \to \infty$. Then, under every law $\mathbb{P}^\pi$ for some probability π on D, the following convergence holds uniformly on compacts in probability:

$$G_s^n \xrightarrow[n\to\infty]{\text{ucp}} cs, \qquad \textit{where} \quad G_s^n := \frac{1}{n\xi_n}\sum_{k=1}^{\lfloor sn\rfloor} h_n(Z_k). \tag{4.1}$$

Remark. If $(h_n)_{n\in\mathbb{N}^*}$ is non-negative (resp., uniformly bounded), then $n\xi_n \to \infty$ and $\xi_n^{-1}\psi(h_n) \to c < \infty$ already imply $(n\xi_n^2)^{-1}\psi(|h_n|) \to 0$ (resp., $(n\xi_n^2)^{-1}\psi(h_n^2) \to 0$).

Proof (of Theorem 4.2). Convergence in probability is equivalent to the property that – given any subsequence – there exists a further subsequence which converges almost surely. By Proposition 17.1.6 of Meyn and Tweedie (1993), therefore, it is sufficient to prove this theorem under the law $\mathbb{P}^\psi$ only.

For each $s \geq 0$ and $n \in \mathbb{N}^*$, we observe $G_s^n = H_s^n + H_s'^n$, where

$$H_s^n = \frac{\lfloor sn\rfloor\psi(h_n)}{n\xi_n} \quad \text{and} \quad H_s'^n = \frac{1}{n\xi_n}\sum_{k=1}^{\lfloor sn\rfloor}\Big(h_n(Z_k) - \psi(h_n)\Big).$$

By assumption, we have $H_s^n \to sc$ uniformly in s as $n \to \infty$. It remains to show that $H_s'^n$ converges to zero uniformly on compacts in probability.

We note $\mathbb{E}^\psi[h_n(Z_k)] = \psi(h_n)$ for every $k, n \in \mathbb{N}^*$; thus, $\mathbb{E}^\psi[H_s'^n] = 0$ for all $s \geq 0$. Moreover, its second moment satisfies $\mathbb{E}^\psi[(H_s'^n)^2] = K_s^n + K_s'^n$, where

$$K_s^n = \frac{1}{n^2\xi_n^2} \sum_{k=1}^{\lfloor sn \rfloor} \left(\psi(h_n^2) - \psi(h_n)^2\right)$$

and

$$K_s'^n = \frac{2}{n^2\xi_n^2} \sum_{k=1}^{\lfloor sn \rfloor - 1} \int \psi(\mathrm{d}z) h_n(z) \sum_{l=k+1}^{\lfloor sn \rfloor} \left(\Psi^{l-k} h_n(z) - \psi(h_n)\right).$$

First, we note

$$|K_s^n| \leq \frac{\lfloor sn \rfloor}{n} \left| \frac{\psi(h_n^2)}{n\xi_n^2} - \frac{\psi(h_n)^2}{n\xi_n^2} \right| \xrightarrow[n\to\infty]{} 0. \tag{4.2}$$

Second, let $m \in \mathbb{N}^*$ denote the period of Z. As the state space is petite w. r. t. Ψ, by Theorem 16.2.2 of Meyn and Tweedie (1993), the sampled chain with transition probability Ψ^m is aperiodic and uniformly ergodic. By Theorem 5.4.4 of Meyn and Tweedie (1993), there exists a partition $D_0, \ldots, D_{m-1}$ of the state space such that each D_i is a recurrence class and such that the measures $m\psi(\,\cdot\, \cap D_i)$ are invariant w. r. t. Ψ^m. For every $i \in \{1, \ldots, m\}$ and $z \in D_i$, we denote $j(l, z) := (i + l) \bmod m$, where 'mod' stands for the modulo operator. For every $n_0 \in \mathbb{N}^*$, we observe

$$\begin{aligned}\sum_{l=1}^{n_0} \left(\Psi^l h_n(z) - \psi(h_n)\right) = &\sum_{k=0}^{\lfloor \frac{n_0}{m} \rfloor} \sum_{l=1}^{m} \left(\Psi^{km+l} h_{n|D_{j(l,z)}}(z) - m\psi(h_{n|D_{j(l,z)}})\right) \\ &+ \sum_{l=1}^{n_0 \bmod m} \left(\Psi^{\lfloor \frac{n_0}{m} \rfloor m+l} h_{n|D_{j(l,z)}}(z) - \psi(h_n)\right).\end{aligned} \tag{4.3}$$

Hence,

$$\left|\sum_{l=1}^{n_0} \left(\Psi^l h_n(z) - \psi(h_n)\right)\right| \leq \sum_{k=0}^{\infty} \sum_{l=1}^{m} \left|\Psi^{km+l} h_{n|D_{j(l,z)}}(z) - m\psi(h_{n|D_{j(l,z)}})\right| + m|\psi(h_n)|.$$

By Theorem 16.2.1 of Meyn and Tweedie (1993), there exists a $\zeta < 1$ such that, for every $l = 1, \ldots, m$ and each $k \in \mathbb{N}$,

$$\sup_{z \in D} \left|\Psi^{km+l} h_{n|D_{j(l,z)}}(z) - m\psi(h_{n|D_{j(l,z)}})\right| \leq \zeta^k. \tag{4.4}$$

Consequently,

$$|K_s'^n| \leq \frac{2\lfloor sn \rfloor m}{n} \left(\frac{\zeta\psi(|h_n|)}{(1-\zeta)n\xi_n^2} + \frac{\psi(|h_n|)|\psi(h_n)|}{n\xi_n^2}\right) \xrightarrow[n\to\infty]{} 0. \tag{4.5}$$

By (4.2) and (4.5), $\mathbb{E}^{\psi}[(H_s'^n)^2] \to 0$, hence $H_s'^n \to 0$ in probability as $n \to \infty$. It remains to show the local uniformity in s of this convergence.

By (4.3) and (4.4), we have that $h_n - \psi(h_n)$ is in the range of $(\mathrm{I} - \Psi)$. Let $\hat{h}_n$ denote its pre-image under $(\mathrm{I} - \Psi)$ (that is, its *potential*), and define the process M^n by

$$M_s^n := \frac{1}{n\xi_n} \sum_{k=1}^{\lfloor sn \rfloor} \Big(\hat{h}_n(Z_k) - \Psi \hat{h}_n(Z_{k-1})\Big).$$

We note that M^n is a $\mathscr{G}_s^n$-martingale where $\mathscr{G}_s^n := \sigma(Z_k : k \leq \lfloor sn \rfloor)$. Since $(\Psi h_n)_{n\in\mathbb{N}^*}$ is uniformly bounded by assumption, so is $(\Psi\hat{h}_n)_{n\in\mathbb{N}^*}$. As $n \to \infty$, therefore, we have $|H_s'^n - M_s^n| = (n\xi_n)^{-1}|\Psi\hat{h}_n(Z_0) - \Psi\hat{h}_n(Z_{\lfloor sn \rfloor})| \to 0$. Likewise, $\mathbb{E}^{\psi}[(M_s^n)^2] \leq 2\,\mathbb{E}^{\psi}(H_s'^n)^2 + 2\,\mathbb{E}^{\psi}\,|H_s'^n - M_s^n|^2 \to 0$. By Doob's inequality, therefore, $M^n \Rightarrow 0$ in ucp. Hence, also $H'^n \Rightarrow 0$ uniformly on compacts in probability as $n \to \infty$. □

4.2. The auxiliary Markov chains

In this subsection, we construct auxiliary Markov chains Z and Z' to which Theorem 4.2 applies. Once and for all, we fix our points of interest, i. e., $\{(x_i, y_i) : i \in I\}$ of Theorem 3.7 such that $\mu'(x_i) > 0$ and $F(x_i, E) > 0$ for each i. Moreover, we choose a compact set $C \supset \{x_i : i \in I\}$ and constants $0 < \varepsilon, \varepsilon' < \infty$ such that $\varepsilon < \|y_i\| < \varepsilon'$ for all $i \in I$ and such that

$$\inf_{x\in C} F\Big(x, \{y : \varepsilon < \|y\| < \varepsilon'\}\Big) > 0. \tag{4.6}$$

Remark. Under Assumptions 2.2 and 3.2, such a set C always exists by the choice of the points x_i and the continuity of f on $E \times E^*$.

Let $T_1, T_2, \ldots$ denote the successive times of jumps of size between ε and ε' starting from C; that is,

$$T_1 := \inf\Big\{t > 0 : \varepsilon < \|\Delta X_t\| < \varepsilon', X_{t-} \in C\Big\} \quad \text{and} \quad T_{n+1} := T_1 \circ \theta_{T_n} + T_n.$$

The conditional expectation w. r. t. the strict past of the stopping times T_n plays a key role. We set

$$q(x) := F\Big(x, \{y : \varepsilon < \|y\| < \varepsilon'\}\Big)\mathbb{1}_C(x),$$
$$p(x,y) := \begin{cases} q^{-1}(x) f(x,y), & \text{if } x \in C \text{ and } \varepsilon < \|y\| < \varepsilon', \\ 0, & \text{else.} \end{cases}$$

It is well-known that $T_1 < \infty$ a. s. if, and only if, $\mu(q) > 0$. In our case, this holds by (4.6). Therefore, $T_n < \infty$ a. s. for all n as well. For convenience, we abbreviate the kernel with density p by Π; its shifted version with density $(x, y) \mapsto p(x, y - x)$ we denote by $\bar{\Pi}$. By Weil (1971), Π (resp., $\bar{\Pi}$) is the conditional transition probability kernel of the jumps at the time(s) T_n in the following sense: On the set $\{T_n < \infty\}$, for every random variable Y, measurable function g, and all x, we have

$$\mathbb{E}^x[g(\Delta X_{T_n}) \mid \mathscr{F}_{T_n-}] = \Pi g(X_{T_n-}), \tag{4.7}$$

$$\mathbb{E}^x[Y \circ \theta_{T_n} \mid \mathscr{F}_{T_n-}] = \bar{\Pi}\, \mathbb{E}^{\cdot}[Y](X_{T_n-}). \tag{4.8}$$

We note $\bar{\Pi}\, \mathbb{E}^{\cdot}[Y](x) = \int p(x,y)\, \mathbb{E}^{x+y}[Y]\mathrm{d}y$.

Let $\mathbf{D} := \mathcal{D}([0,1[; E) \times \mathbb{R}_+ \times C$. For every $k \in \mathbb{N}^*$, we define the $\mathbf{D}$-valued and C-valued random variables

$$Z_k := \Big(s \mapsto X_{(1-s)T_{k-1}+sT_k}, T_k - T_{k-1}, X_{T_k-}\Big) \quad \text{and} \quad Z'_k := X_{T_k-}.$$

The corresponding filtration $(\mathscr{G}_k)_{k\in\mathbb{N}^*}$ is given by $\mathscr{G}_k := \mathscr{F}_{T_k-}$. We emphasise that we exclude time $k = 0$. From (4.8) and $T_1 < \infty$ a.s., we deduce that $Z = (Z_k)_{k\in\mathbb{N}^*}$ and $Z' = (Z'_k)_{k\in\mathbb{N}^*}$ are $\mathscr{G}_k$-Markov chains. We denote their transition probabilities by Ψ and Φ, respectively. We refer to appendix A for technical results on these auxiliary Markov chains.

4.3 Lemma. *Let $(g,t,x) \in \mathbf{D}$, let $A \subseteq C$ and $\mathbf{A} \subseteq \mathbf{D}$ be measurable, and let $k \in \mathbb{N}^*$. Then*

$$\Phi(x, A) = \bar{\Pi}\, \mathbb{P}^{\cdot}(Z'_1 \in A)(x), \tag{4.9}$$

$$\Psi^{k+1}((g,t,x), \mathbf{A}) = \Phi^k \Psi(x, \mathbf{A}). \tag{4.10}$$

Proof. We deduce (4.9) and (4.10) directly from (4.8) and the Markov property of X, respectively. □

By Lemma 4.3, Theorem 4.2 applies to Z' and, also, to Z.

4.4 Lemma. *Grant Assumptions 2.2 and 3.2. Then the Markov chain Z' is strong Feller. Its state space C is* petite *with respect to Φ.*

Proof. Let f be a bounded Borel function and $x_0 \in C$. Under Assumption 3.2, we deduce from Lebesgue's dominated convergence theorem that q is continuous. By (4.6), we have that $x \mapsto p(x,y)$ is also continuous for every y and $\sup\{p(x,y) : x \in C, y \in E\} < \infty$. Again by Lebesgue's dominated convergence theorem, we conclude that

$$\lim_{x\to x_0} \bar{\Pi} g(x) = \lim_{x\to x_0} \int p(x,y) g(x+y)\mathrm{d}y = \int p(x_0,y) g(x+y)\mathrm{d}y = \bar{\Pi} g(x_0).$$

By (4.9), consequently, $\Phi = \bar{\Pi}\, \mathbb{P}^{\cdot}(Z'_1 \in \cdot)$ is strong Feller on C.

By the same argument as for the equivalence of $T_1 < \infty$ a.s. and $\mu(q) > 0$, we have that the measure with μ-density q is an irreducibility measure of Z'. Under Assumption 2.2, it is absolutely continuous. Thus, its support has non-empty interior. By Theorem 6.2.5 (ii) of Meyn and Tweedie (1993), therefore, every compact set – hence the state space C of Z' – is petite with respect to Φ. □

4.5 Corollary. *Grant Assumptions 2.2 and 3.2. Then the state space* $\mathbf{D}$ *of* Z *is petite w. r. t.* Ψ*.*

Proof. By Lemma 4.4, there exists a probability ρ on $\mathbb{N}^*$ and a non-trivial measure ν_ρ on C such that, for every Borel set $A \subseteq C$,

$$\inf_{x\in C} \sum_{k=1}^{\infty} \rho(k)\Phi^k(x,A) \geq \nu_b(A).$$

Let $(g,t,x) \in \mathbf{D}$, $\mathbf{A} \subseteq \mathbf{D}$ be measurable, and $\tilde{\rho}$ be the probability on $\mathbb{N}^*$ given by $\tilde{\rho}(1) = 0$ and $\tilde{\rho}(k) = \rho(k-1)$ for $k > 1$. By (4.10), then

$$\sum_{k=1}^{\infty} \tilde{\rho}(k)\Psi^k((g,t,x),A) = \sum_{k=1}^{\infty} \rho(k)\Phi^k\Psi(x,A) \geq \nu_\rho\Psi(A) =: \tilde{\nu}_{\tilde{\rho}}(A).$$

Since ν_ρ is non-trival, so is $\tilde{\nu}_{\tilde{\rho}}$. □

4.3. Proof of Theorem 3.6

Throughout the remainder of section 4, we work under the law $\mathbb{P}^\pi$ for some initial probability π on E and, for presentational purposes, we suppose w. l. o. g. that $\mu(q) = 1$.

We consider the processes $G^{n,\eta}$, $J^{n,\eta}$ and $S^{n,\eta}$ given by

$$G^{n,\eta}_s(x,y) := \frac{1}{n}\sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta,x}(X_{T_k-})g_2^{\eta,y}(\Delta X_{T_k}), \tag{4.11}$$

$$J^{n,\eta}_s(x) := \frac{1}{n}\sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta,x}(X_{T_k-}) \quad \text{and} \quad S^{n,\eta}_s(x) := \frac{1}{n}\int_0^{T_{\lfloor sn\rfloor}} g_1^{\eta,x}(X_r)\mathrm{d}r. \tag{4.12}$$

We emphasise that these processes are of the form $\sum_{k=1}^{\lfloor sn\rfloor} h_n(Z_k)$ where Z is the auxiliary Markov chain defined in section 4.2. We utilise the following preliminary condition as $n \to \infty$ (cf., (3.2)):

$$n\eta_{1,n}^d\eta_{2,n}^d \to \infty, \quad \text{and} \quad \eta_{1,n} \to 0, \eta_{2,n} \to 0. \tag{4.13}$$

4.6 Lemma. *Grant Assumptions 2.2, 3.1 and 3.2. Let* $\eta_n = \eta_{1,n}$ *be such that (4.13) holds. Then the following convergences hold uniformly on compacts in probability:*

$$J^{n,\eta_n}_s(x) \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} sq(x)\mu'(x) \quad \textit{and} \quad S^{n,\eta_n}_s(x) \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} s\mu'(x).$$

Proof. Let ψ and φ denote the invariant probabilities of Z and Z', respectively. We apply Theorem 4.2:

(i) We note that $J^{n,\eta_n}(x)$ is of the form (4.1) with $\xi_n = \eta_n^d$ and $h_n : C \to \mathbb{R}$ given by $h_n(z) = g_1((z-x)/\eta_n)$; $(h_n)_{n\in\mathbb{N}^*}$ is uniformly bounded. By Corollary A.6 where $\mu(q) = 1$,

q is the μ-density of φ. Also q and μ' are continuous. By Lebesgue's differentiation theorem, thus,

$$\eta_n^{-d}\varphi(h_n) = \eta_n^{-d}\int \mu(\mathrm{d}z)q(z)g_1((z-x)/\eta_n) \xrightarrow[n\to\infty]{} q(x)\mu'(x).$$

Since $n\eta_n^d \to \infty$, likewise, $(n\eta_n^{2d})^{-1}\varphi(|h_n|) \to 0$ as $n \to \infty$. $\diamond$

(ii) We note that $S^{n,\eta_n}(x)$ is of form (4.1) with $\xi_n = \eta_n^d$ and $h_n : \mathbf{D} \to \mathbb{R}$ given by $h_n(g,t,z) = t\int_0^1 g_1((g(s)-x)/\eta_n)\mathrm{d}s$. By Corollary A.6, $\psi = \varphi\Psi$. By Lemmata A.2 and A.5, thus,

$$\eta_n^{-d}\varphi(h_n) = \eta_n^{-d}\int \mu(\mathrm{d}z)g_1((z-x)/\eta_n) \xrightarrow[n\to\infty]{} \mu'(x).$$

Likewise, $(n\eta_n^{2d})^{-1}\varphi(|h_n|) \le (n\eta_n^{2d})^{-1}\int \mu(\mathrm{d}z)|g_1((z-x)/\eta_n)| \to 0$. By Corollary A.4, in addition, we observe

$$\frac{\psi(h_n^2)}{n\eta_n^{2d}} \le \frac{2\|g_1\|_\infty}{\inf_{z\in C} q(z)} \frac{\int \mu(\mathrm{d}z)|g_1((z-x)/\eta_n)|}{n\eta_n^{2d}} \xrightarrow[n\to\infty]{} 0. \qquad \square$$

4.7 Lemma. *Grant Assumptions 2.2, 3.1 and 3.2. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (4.13) holds. Then the following convergence holds uniformly on compacts in probability:*

$$G_s^{n,\eta_n}(x,y) \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} sf(x,y)\mu'(x).$$

Proof. Let $(\mathscr{H}_s^n)_{s\ge0}$ be the filtration given by $\mathscr{H}_s^n := \mathscr{F}_{T_{\lfloor sn\rfloor+1}-}$. By (4.7), we have $\mathbb{E}[\Delta G_s^{n,\eta_n} \mid \mathscr{H}_{s-}^n] = g_1^{\eta_n,x}(Z_k')\Pi g_2^{\eta_n,y}(Z_k')$ for $s = k/n$. Thus, the compensator of G^{n,η_n} w. r. t. $(\mathscr{H}_s^n)_{s\ge0}$ is given by $H_s^{n,\eta_n} := n^{-1}\sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta_n,x}(Z_k')\Pi g_2^{\eta_n,y}(Z_k')$.

Fix $s \ge 0$. In analogy to the proof of Lemma 4.4, $\Pi g_2^{\eta_n,y}$ is continuous under Assumption 3.2. In analogy to Lemma 4.6, $n^{-1}\sum_{k=1}^{\lfloor sn\rfloor}|g_1^{\eta_n,x}(Z_k')|$ converges in ucp to a non-trivial process as $n \to \infty$. Therefore,

$$|H_s^{n,\eta_n} - \Pi g_2^{\eta_n,y}(x)J_s^{n,\eta_n}(x)| \le \sup_{z\in B_{\eta_n}(x)}|\Pi g_2^{\eta_n,y}(z) - \Pi g_2^{\eta_n,y}(x)| \cdot \frac{1}{n}\sum_{k=1}^{\lfloor sn\rfloor}|g_1^{\eta_n,x}(Z_k')| \xrightarrow[n\to\infty]{} 0.$$

Since p is continuous under Assumption 3.2, $\lim_{n\to\infty}\Pi g_2^{\vartheta_n,y}(x) = p(x,y)$ by Lebesgue's differentiation theorem. We recall $f(x,y) = q(x)p(x,y)$. By Lemma 4.6, hence,

$$H_s^n \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} sf(x,y)\mu'(x).$$

It remains to prove $M_s^n := G_s^n - H_s^n \Rightarrow 0$ uniformly on compacts in probability. By (4.13), we have $\sup_s\|\Delta M_s^n\|_\infty \le (n\eta_n^d\vartheta_n^d)^{-1}\|g_1\|_\infty\|g_2\|_\infty \to 0$. By Theorem VIII.3.33 of

Jacod and Shiryaev (2003), thus, it is sufficient to show that the predictable quadratic variation $\langle M^n, M^n\rangle_s$ of M^n converges in probability to zero for all s. We observe

$$\Big\langle M^n, M^n\Big\rangle_s = \frac{1}{n^2}\sum_{k=1}^{\lfloor sn\rfloor} \mathbb{E}^\pi\left[g_1^{\eta_n,x}(Z'_k)^2\Big(g_2^{\eta_n,y}(\Delta X_{T_k}) - \Pi g_2^{\eta_n,y}(Z'_k)\Big)^2 \,\Big|\, \mathscr{H}^n_{k/n}\right]$$

$$\leq \frac{1}{n\eta_{1,n}^d\eta_{2,n}^d}\cdot\frac{1}{n}\sum_{k=1}^{\lfloor sn\rfloor}\eta_{1,n}^d g_1^{\eta_n,x}(Z'_k)^2\int_{B_1(0)} p(Z'_k, y+\eta_{2,n}z)g_2(z)^2\mathrm{d}z.$$

In analogy to Lemma 4.6 again, $n^{-1}\sum_{k=1}^{\lfloor sn\rfloor}\eta_{1,n}^d g_1^{\eta_n,x}(Z'_k)^2$ converges in ucp to a non-trivial process as $n\to\infty$. As in the proof of Lemma 4.4, moreover, p is bounded on $C\times E$. Consequently, $\langle M^n, M^n\rangle_s \to 0$ in probability as $n\to\infty$. □

Next, we carry Lemmata 4.6 and 4.7 over to the time-scale of X. Let J be the process given by

$$J_t := \sum_{k=1}^{\infty}\mathbb{1}_{[0,t]}(T_k). \tag{4.14}$$

We note that J is a non-decreasing additive functional of X. It is the random clock of Z (and Z') in terms of X. By (3.1) – where $H_t = t$ –, and by $\mu(q) = 1$, we have $\mathbb{E}^\mu J_t = t$ for all $t > 0$.

4.8 Lemma. *Grant Assumptions 2.2, 3.1 and 3.2. Let $v : \mathbb{R}_+ \to \mathbb{R}_+$ denote a deterministic equivalent of X, and let η_t and $(x,y)\in E\times E^*$ be as in Theorem 3.6. Then*

$$\text{the family } \left\{\mathscr{L}\left(G^{v_t,\eta_t}_{J_t/v_t}(x,y), S^{v_t,\eta_t}_{J_t/v_t}(x) \mid \mathbb{P}^\pi\right) : t>0\right\} \text{ is tight.} \tag{4.15}$$

Moreover, each limit point of the family in (4.15) is the law $\mathscr{L}(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L})$ for some positive random variable $\tilde{L}$.

Proof. As J is a non-decreasing additive functional of X, by Löcherbach and Loukianova (2008), the families $\{\mathscr{L}(J_t/v_t \mid \mathbb{P}^\pi) : t>0\}$ and $\{\mathscr{L}(v_t/J_t \mid \mathbb{P}^\pi) : t>0\}$ are tight. By Corollary VI.3.33 of Jacod and Shiryaev (2003) and Lemma 4.7, thus,

$$\text{the family } \{\mathscr{L}(G^{v_t,\eta_t}(x,y), S^{v_t,\eta_t}(x), J_t/v_t, v_t/J_t \mid \mathbb{P}^\pi) : t>0\} \text{ is tight.} \tag{4.16}$$

Let $\mathbb{Q}$ denote a limit point of the family in (4.16), and let $(t_n)_{n\in\mathbb{N}}$ a sequence such that

$$\mathscr{L}(G^{v_{t_n},\eta_{t_n}}(x,y), S^{v_{t_n},\eta_{t_n}}(x), J_{t_n}/v_{t_n}, v_{t_n}/J_{t_n} \mid \mathbb{P}^\pi) \xrightarrow[n\to\infty]{\mathrm{w}} \mathbb{Q}.$$

On some extension of the probability space, w. l. o. g., there exists a random variable $\tilde{L} > 0$ such that $\mathbb{Q} = \mathscr{L}(s\mapsto sf(x,y)\mu'(x), s\mapsto s\mu'(x), \tilde{L}, 1/\tilde{L})$. Since its first and second marginal are the laws of continuous processes, we have

$$\mathscr{L}\Big(G^{v_{t_n},\eta_{t_n}}_{J_{t_n}/v_{t_n}}(x,y), S^{v_{t_n},\eta_{t_n}}_{J_{t_n}/v_{t_n}}(x) \mid \mathbb{P}^\pi\Big) \xrightarrow[n\to\infty]{\mathrm{w}} \mathscr{L}\left(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L}\right).$$

□

Proof (of Theorem 3.6). For every $t \geq 0$ and each x and y, we have

$$\hat{f}_t^{\eta_t}(x,y) = \frac{G^{v_t,\eta_t}_{J_t/v_t}(x,y)}{S^{v_t,\eta_t}_{J_t/v_t}(x) + v_t^{-1}\int_{T_{J_t}}^t g_1^{\eta_t,x}(X_s)\mathrm{d}s}.$$

Let $h_n : \mathbf{D} \to \mathbb{R}$ be given by $h_n(g,t,z) := t\int_0^1 |g_1^{\eta_n,x}(g(s))|\mathrm{d}s$. By Lemma A.2 and Corollaries A.4 and A.6, we have $\psi(h_n^2) \leq 2\|g_1\|_\infty \eta_{1,n}^{-d}(\inf_{z\in C} q(z))^{-1}\mu(|g_1^{\eta_n,x}|)$. By Markov's inequality, since $v_t^2\eta_{1,t}^d \to \infty$, therefore,

$$v_t^{-1}\int_{T_{J_t}}^t g_1^{\eta_t,x}(X_s)\mathrm{d}s \leq v_t^{-1}h_{v_t}(Z_{J_t+1}) \xrightarrow[t\to\infty]{\mathbb{P}^\psi} 0. \tag{4.17}$$

By Proposition 17.1.6 of Meyn and Tweedie (1993), in analogy to the proof of Theorem 4.2, this convergence in probability holds under every law $\mathbb{P}^\pi$.

We recall the results from Lemma 4.8. Let $\tilde{L} > 0$ be a random variable such that the law $\mathscr{L}(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L})$ is a limit point of the family in (4.15). Moreover, let $(t_n)_{n\in\mathbb{N}^*}$ be a sequence such that

$$\Big(G^{v_{t_n},\eta_{t_n}}_{J_{t_n}/v_{t_n}}(x,y), S^{v_{t_n},\eta_{t_n}}_{J_{t_n}/v_{t_n}}(x)\Big) \xrightarrow[n\to\infty]{\mathscr{L}} \Big(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L}\Big).$$

We recall $\mu'(x) > 0$. Consequently, $\hat{f}_{t_n}^{\eta_{t_n}}(x,y) \to f(x,y)$ in law as $n \to \infty$ by the continuous mapping theorem. As this limit is unique and independent of the particular limit point of the family in (4.15), we have that $\hat{f}_t^{\eta_t}(x,y)$ converges to $f(x,y)$ in law, hence, in probability. □

4.4. Proofs of Theorem 3.7 and Corollary 3.8

In this subsection, we work on the extended space (2.11), L denotes the Mittag-Leffler process of order $0 < \delta \leq 1$, and $W = (W^i)_{i\in I}$ denotes an I-dimensional standard Wiener process such that L, W and $\mathscr{F}$ are independent.

In addition to the processes $G^{n,\eta}$, $J^{n,\eta}$ and $S^{n,\eta}$ given in (4.11) and (4.12), we consider the process $U^{n,\eta}$ given by

$$U_s^{n,\eta}(x,y) := \sqrt{n\eta_1^d\eta_2^d}\left(G_s^{n,\eta}(x,y) - \frac{\mu(g_1^{\eta,x}Fg_2^{\eta,y})}{\mu(g_1^{\eta,x})}S_s^{n,\eta}(x)\right). \tag{4.18}$$

We emphasise again that these processes are of the form $\sum_{k=1}^{\lfloor sn\rfloor} h_n(Z_k)$ where Z is the auxiliary Markov chain defined in section 4.2.

4.9 Lemma. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (4.13) holds. Then we have the following convergence in law in $\mathcal{D}(\mathbb{R}^I)$:*

$$\Big(U_s^{n,\eta_n}(x_i,y_i)\Big)_{i\in I} \xrightarrow[n\to\infty]{\mathscr{L}} \Big(\mu'(x_i)\sigma(x_i,y_i)W_s^i\Big)_{i\in I},$$

where $\sigma(x,y)^2$ is given by (3.4).

Proof. For $n \in \mathbb{N}^*$, let $M^{n,\eta}$ be the process given by

$$M_s^{n,\eta}(x,y) := \frac{\sqrt{\eta_1^d \eta_2^d}}{\sqrt{n}} \sum_{k=1}^{\lfloor sn \rfloor} \left(g_1^{\eta,x}(Z_k') g_2^{\eta,y}(\Delta X_{T_k}) - \int_{T_{k-1}}^{T_k} g_1^{\eta,x}(X_s) F g_2^{\eta,y}(X_s) \mathrm{d}s \right),$$

and let $(\mathscr{H}_s^n)_{s\geq 0}$ be given by $\mathscr{H}_s^n := \mathscr{F}_{T_{\lfloor sn \rfloor}}$. By Theorem VIII.3.33 of Jacod and Shiryaev (2003), it is sufficient to prove (i)–(iv) as follows:

(i) We have $U_s^{n,\eta_n}(x,y) - M_s^{n,\eta_n}(x,y) \Rightarrow 0$ in ucp as $n \to \infty$.

(ii) The process $M^{n,\eta}$ is an $\mathscr{H}_s^n$-martingale for each n.

(iii) For all $i, j \in I$, we have

$$\left\langle M^{n,\eta_n}(x_i,y_i), M^{n,\eta_n}(x_j,y_j) \right\rangle_s \xrightarrow[n\to\infty]{\mathbb{P}^\pi} s[\sigma(x_i,y_i)\mu'(x)]^2 \delta_{ij}.$$

(iv) We have the "conditional Lyapunov condition"

$$K_s^{n,\eta_n}(x,y) := \sum_{k=1}^{\lfloor sn \rfloor} \mathbb{E}^\pi \left[\left(\Delta M_{k/n}^{n,\eta_n}(x,y) \right)^4 \middle| \mathscr{H}_{k/n-}^n \right] \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0.$$

(i) We note that $U^{n,\eta}(x,y) - M^{n,\eta}(x,y)$ is of form (4.1) with $h_n : \mathbf{D} \to \mathbb{R}$ given by

$$h_n(g,t,z) = t \int_0^1 g_1 \left(\frac{g(s) - x}{\eta_{1,n}} \right) \left(F g_2^{\eta_n,y}(g(s)) - \frac{\mu(g_1^{\eta_n,x} F g_2^{\eta_n,y})}{\mu(g_1^{\eta_n,x})} \right) \mathrm{d}s,$$

and $\xi_n = \eta_{1,n}^{d/2} \eta_{2,n}^{-d/2} n^{-1/2}$. By Lemmata A.2 and A.5 and Corollary A.6, we have

$$\xi_n^{-1} \psi(h_n) = \sqrt{n \eta_{1,n}^d \eta_{2,n}^d} \int \mu(\mathrm{d}z) g_1^{\eta,x}(z) \left(F g_2^{\eta_n,y}(z) - \frac{\mu(g_1^{\eta_n,x} F g_2^{\eta_n,y})}{\mu(g_1^{\eta_n,x})} \right) \equiv 0.$$

Since $\eta_{2,n} \to 0$, we also observe

$$\frac{\psi(|h_n|)}{n\xi_n^2} \leq \eta_{2,n}^d \left(\mu(|g_1^{\eta_n,x} F g_2^{\eta_n,y}|) + \mu(|g_1^{\eta_n,x}|) \cdot \left| \frac{\mu(g_1^{\eta_n,x} F g_2^{\eta_n,y})}{\mu(g_1^{\eta_n,x})} \right| \right) \xrightarrow[n\to\infty]{} 0.$$

By Corollary A.4, likewise,

$$\frac{\psi(h_n^2)}{n\xi_n^2} \leq \frac{2\eta_{2,n}^d \|g_1\|_\infty \|F g_2^{\eta_n,y}\|_\infty}{\inf_{z\in C} q(z)} \left(\mu(|g_1^{\eta_n,x} F g_2^{\eta_n,y}|) + \mu(|g_1^{\eta_n,x}|) \left| \frac{\mu(g_1^{\eta_n,x} F g_2^{\eta_n,y})}{\mu(g_1^{\eta_n,x})} \right| \right) \xrightarrow[n\to\infty]{} 0.$$

Since $n\xi_n \to \infty$, we deduce from Theorem 4.2 that (i) holds. ⋄

(ii) By construction, $M^{n,\eta}$ is integrable and adapted to $(\mathscr{H}^n_s)_{s\geq 0}$. For $s = k/n$, we note $\mathscr{H}^n_{s-} = \mathscr{F}_{T_{k-1}}$. By (3.1) – where $H_t = t$ – the compensator of our process's jump measure is given by $\mathrm{d}t \otimes F(X_t, \mathrm{d}y)$. By Doob's optional sampling theorem, thus,

$$\mathbb{E}^\pi \left[g_1^{\eta,x}(Z'_k) g_2^{\eta,y}(\Delta X_{T_k}) - \int_{T_{k-1}}^{T_k} g_1^{\eta,x}(X_s) F g_2^{\eta,y}(X_s) \mathrm{d}s \,\middle|\, \mathscr{F}_{T_{k-1}} \right] = 0$$

for all $k \in \mathbb{N}^*$. Therefore, $M^{n,\eta}(x,y)$ is an $\mathscr{H}^n_s$-martingale. $\diamond$

(iii) Let $i, j \in I$. In analogy to step (ii), we deduce

$$\left\langle M^{n,\eta_n}(x_i,y_i), M^{n,\eta_n}(x_j,y_j) \right\rangle_s = \frac{\eta_{1,n}^d \eta_{2,n}^d}{n} \sum_{k=1}^{\lfloor sn \rfloor} \mathbb{E}^\pi \left[g_1^{\eta_n,x_i} g_1^{\eta_n,x_j}(Z'_k) g_2^{\eta_n,y_i} g_2^{\eta_n,y_j}(\Delta X_{T_k}) \,\middle|\, \mathscr{F}_{T_{k-1}} \right].$$

For all n large enough, we have $g_1^{\eta_n,x_i} g_1^{\eta_n,x_j} = 0$ whenever $x_i \neq x_j$, and $g_2^{\eta_n,y_i} g_2^{\eta_n,y_j} = 0$ whenever $y_i \neq y_j$. For all ω, if $i \neq j$, thus, $\langle M^{n,\eta_n}(x_i,y_i), M^{n,\eta_n}(x_j,y_j)\rangle_s \to 0$.

Moreover, let $J'^{n,\eta_n}_s(x) := n^{-1}\eta_{1,n}^d \sum_{k=1}^{\lfloor sn \rfloor} \mathbb{E}^{X_{T_{k-1}}}[g_1^{\eta,x}(Z'_1)^2]$. We note that J'^{n,η_n} is of form (4.1) with $\xi_n = \eta_{1,n}^d$ and $h_n : \mathbf{D} \to \mathbb{R}$ given by $h_n(g,t,z) = \mathbb{E}^{g(0)}[g_1((Z'_1 - x)/\eta_{1,n})^2]$. By Lemma A.5 and Corollary A.6 and under Assumption 3.2, we observe

$$\eta_{1,n}^{-d}\psi(h_n) = \int \mu'(x + \eta_{1,n}z) q(x + \eta_{1,n}z) g_1(z)^2 \mathrm{d}z \xrightarrow[n\to\infty]{} \mu'(x)q(x) \int g_1(z)^2 \mathrm{d}z.$$

By Theorem 4.2, since h_n is non-negative and uniformly bounded, thus,

$$J'^{n,\eta_n}_s(x) \underset{n\to\infty}{\overset{\text{ucp}}{\Longrightarrow}} sq(x)\mu'(x) \int g_1(z)^2 \mathrm{d}z. \tag{4.19}$$

Hence, we observe

$$\begin{aligned} &\left| \left\langle M^{n,\eta_n}(x,y), M^{n,\eta_n}(x,y) \right\rangle_s - J'^{n,\eta_n}_s(x) p(x,y) \int g_2(w)^2 \mathrm{d}w \right| \\ &\quad \leq J'^{n,\eta_n}_s(x) \int g_2(w)^2 \mathrm{d}w \sup_{z,w \in B_1(0)} \left| p(x + \eta_{1,n}z, y + \eta_{2,n}w) - p(x,y) \right| \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0. \end{aligned}$$

Since $f(x,y) = q(x)p(x,y)$, consequently,

$$\left\langle M^{n,\eta_n}(x,y), M^{n,\eta_n}(x,y) \right\rangle_s \xrightarrow[n\to\infty]{\mathbb{P}^\pi} sf(x,y)\mu'(x) \int g_1(w)^2 \mathrm{d}w \int g_2(z)^2 \mathrm{d}z;$$

that is, (iii) holds. $\diamond$

(iv) We observe $|K^{n,\eta_n}_s(x,y)| \leq K'^{n,\eta_n}_s + K''^{n,\eta_n}_s$, where

$$K'^{n,\eta_n}_s := \frac{4\eta_{1,n}^{2d}\eta_{2,n}^{2d}}{n^2} \sum_{k=1}^{\lfloor sn \rfloor} \mathbb{E}^{X_{T_{k-1}}} \left[\left(g_1^{\eta,x}(Z'_1) g_2^{\eta,y}(\Delta X_{T_1}) \right)^4 \right],$$

and

$$K_s''^{n,\eta_n} := \frac{4\eta_{1,n}^{2d}\eta_{2,n}^{2d}}{n^2}\sum_{k=1}^{\lfloor sn\rfloor}\mathbb{E}^{X_{T_{k-1}}}\left[\left(\int_0^{T_1} g_1^{\eta,x}Fg_2^{\eta,y}(X_s)\mathrm{d}s\right)^4\right].$$

We note that K'^{n,η_n} and K''^{n,η_n} are of form (4.1) with $\xi_n = n\eta_{1,n}^{2d}\eta_{1,n}^{2d}/4$ and, respectively,

$$h_n(g,t,z) = \mathbb{E}^{g(0)}\left[g_1((Z_1'-x)/\eta_{1,n})^4 g_2((\Delta X_{T_1}-y)/\eta_{2,n})^4\right],$$

and

$$h_n(g,t,z) = \mathbb{E}^{g(0)}\left[\left(\int_0^{T_1} g_1\left(\frac{X_s-x}{\eta_{1,n}}\right)\int F(X_s,\mathrm{d}w)g_2\left(\frac{w-y}{\eta_{2,n}}\right)\right)^4\right].$$

By Lemma A.5 and Corollary A.6, for K'^n, we have

$$\frac{\psi(h_n)}{\xi_n} = \frac{4}{n\eta_{1,n}^d\eta_{2,n}^d}\iint \mu'(x+\eta_{1,n}z)g_1(z)^4 f(x+\eta_{1,n}z, y+\eta_{2,n}w)g_2(w)^4\mathrm{d}w\mathrm{d}z \xrightarrow[n\to\infty]{} 0.$$

By Corollary A.4 and Lemma A.5, for K''^n moreover, there exists a $\zeta < \infty$ such that

$$\frac{\psi(h_n)}{\xi_n} \le \frac{4\zeta}{n\eta_{1,n}^d\eta_{2,n}^d}\iint \mu'(x+\eta_{1,n}z)|g_1(z)|f(x+\eta_{1,n}z, y+\eta_{2,n}w)|g_2(w)|\mathrm{d}w\mathrm{d}z \xrightarrow[n\to\infty]{} 0.$$

Since, in both cases, h_n is non-negative and uniformly bounded, we deduce from Theorem 4.2 that $\left|K_s^{n,\eta_n}(x,y)\right| \le K_s'^{n,\eta_n} + K_s''^{n,\eta_n} \Rightarrow 0$ in ucp as $n\to\infty$. □

Next, we carry Lemma 4.9 over to the time-scale of X. We recall that the additive functional J of X, given in (4.14), is the random clock of Z (and Z') in terms of X. In addition, let L^t denote the process given by $L_s^t := v_t^{-1}J_{st}$.

Under Darling–Kac's condition, we have the important Théorème 3 of Touati (1987) at hand; see also p. 119 of Höpfner et al. (1990) and Theorem 3.15 of Höpfner and Löcherbach (2003). For reference, we include it as the following proposition.

4.10 Proposition. *Grant Assumptions 2.2 and 2.4. Let $H = (H^1,\dots,H^l)$ be a μ-integrable additive functional of X with (component-wise) non-decreasing paths. Then, under every law $\mathbb{P}^\pi$, we have the following convergence in law in $\mathcal{D}(\mathbb{R}^l)$:*

$$(v_t^{-1}H_{st})_{s\ge0} \xrightarrow[t\to\infty]{\mathscr{L}} \left(\mathbb{E}^\mu[H_1^1]L,\cdots,\mathbb{E}^\mu[H_1^l]L\right). \tag{4.20}$$

Recalling Lemma 4.6, by eq. (3.4) of Höpfner et al. (1990), we obtain the following corollary to Proposition 4.10.

4.11 Corollary. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let $\eta_t = \eta_{1,t}$ be such that (3.2) holds. Then we have the following convergence in law in $\mathcal{D}(\mathbb{R}^{1+I})$:*

$$\left(L^t, \left(S_{L^t}^{v_t,\eta_t}(x_i)\right)_{i\in I}\right) \xrightarrow[t\to\infty]{\mathscr{L}} \left(L, \left(\mu'(x_i)L\right)_{i\in I}\right).$$

□

4.12 Lemma. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let* $\eta_t = (\eta_{1,t}, \eta_{2,t})$ *be such that (3.2) holds. Then we have the following convergence in law in* $\mathcal{D}(\mathbb{R}^{1+I})$*:*

$$\left(L^t, (U^{v_t,\eta_t}(x_i,y_i))_{i\in I}\right) \underset{t\to\infty}{\overset{\mathscr{L}}{\Longrightarrow}} \left(L, \left(\mu'(x_i)\sigma(x_i,y_i)W^i\right)_{i\in I}\right),$$

where $\sigma(x,y)^2$ *is given by (3.4).*

Proof. From Corollary 4.11 and Lemma 4.9, we infer

$$L^t \underset{t\to\infty}{\overset{\mathscr{L}}{\Longrightarrow}} L \quad \text{and} \quad (U^{v_t,\eta_t}(x_i,y_i))_{i\in I} \underset{t\to\infty}{\overset{\mathscr{L}}{\Longrightarrow}} \left(\mu'(x_i)\sigma(x_i,y_i)W^i\right)_{i\in I}. \tag{4.21}$$

Thus, the families

$$\left\{\mathscr{L}(L^t \mid \mathbb{P}^\pi) : t \geq 0\right\} \quad \text{and} \quad \left\{\mathscr{L}\left((U^{v_t,\eta_t}(x_i,y_i))_{i\in I} \mid \mathbb{P}^\pi\right) : t \geq 0\right\}$$

are C-tight. By Corollary VI.3.33 of Jacod and Shiryaev (2003), we conclude that

$$\text{the family} \quad \left\{\mathscr{L}\left(L^t, (U^{v_t,\eta_t}(x_i,y_i))_{i\in I} \mid \mathbb{P}^\pi\right) : t \geq 0\right\} \quad \text{is C-tight.} \tag{4.22}$$

In the remainder of this proof, we abbreviate $\boldsymbol{U}^{v_t} := (U^{v_t,\eta_t}(x_i,y_i))_{i\in I}$.

Let $(\bar{\Omega}, \bar{\mathscr{F}}) := (\mathcal{D}(\mathbb{R}\times\mathbb{R}^I), \mathscr{D}(\mathbb{R}\times\mathbb{R}^I))$ denote the canonical space, and let $(L, \boldsymbol{W})$ be the canonical process. Moreover, let $\bar{\mathbb{P}}$ be an arbitrary limit point of the family in (4.22). We deduce from (4.21) that its marginals are given by the Mittag-Leffler law of order δ and the I-dimensional (scaled) Wiener law, respectively. For convenience, we abbreviate $\mathbb{Q}_1 := \mathscr{L}(L \mid \bar{\mathbb{P}})$ and $\mathbb{Q}_2 := \mathscr{L}(\boldsymbol{W} \mid \bar{\mathbb{P}})$. Suppose that L and $\boldsymbol{W}$ are independent processes under $\bar{\mathbb{P}}$. Then $\bar{\mathbb{P}} = \mathbb{Q}_1 \otimes \mathbb{Q}_2$ holds. As $\bar{\mathbb{P}}$ is an arbitrary limit point of the family in (4.22), then it has to be unique. Hence, $(\mathscr{L}((L^t, \boldsymbol{U}^{v_t}) \mid \mathbb{P}^\pi) \to \mathbb{Q}_1 \otimes \mathbb{Q}_2$ weakly as $t \to \infty$. ⋄

Let K denote the right-inverse of L, i. e., $K_t := \inf\{s : L_s > t\}$, and let $(\mathscr{H}_t)_{t\geq 0}$ be the filtration on $\bar{\Omega}$ which is generated by the process $(K, \boldsymbol{W})$. Suppose that – under $\bar{\mathbb{P}}$ – K and $\boldsymbol{W}$ are processes with independent increments relative to $(\mathscr{H}_t)_{t\geq 0}$. (That is, $K_{t+s} - K_t$ and $\mathscr{H}_t$ are independent for all $s, t > 0$, and $\boldsymbol{W}_{t+s} - \boldsymbol{W}_t$ and $\mathscr{H}_t$ are independent for all $s, t > 0$.) Then, in analogy to Step 6 on p. 122 of Höpfner et al. (1990), we deduce that – under $\bar{\mathbb{P}}$ – the pair $(K, \boldsymbol{W})$ itself is a process with independent increments relative to $(\mathscr{H}_t)_{t\geq 0}$. We recall that K is a δ-stable subordinator, thus, purely discontinuous (resp., deterministic if $\delta = 1$). Since $\boldsymbol{W}$ is continuous, hence, K and $\boldsymbol{W}$ are independent processes – under $\bar{\mathbb{P}}$. Consequently, $\bar{\mathbb{P}} = \mathbb{Q}_1 \otimes \mathbb{Q}_2$. ⋄

It remains to show that – under $\bar{\mathbb{P}}$ – K and $\boldsymbol{W}$ are processes with independent increments relative to $(\mathscr{H}_t)_{t\geq 0}$. This, however, follows in analogy to Step 7 on pp. 123f of Höpfner et al. (1990) with obvious notation. □

Next, we demonstrate that the convergence in Lemma 4.12 holds stably in law.

4.13 Lemma. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let η_t be as in Lemma 4.12. Then, we have the following stable convergence in law in $\mathcal{D}(\mathbb{R}^{1+I})$:*

$$\left(L^t, (U^{v_t,\eta_t}_{L^t}(x_i, y_i))_{i\in I}\right) \xrightarrow[t\to\infty]{\mathscr{L}-\text{st}} \left(L, \left(\mu'(x_i)\sigma(x_i, y_i)W^i_L\right)_{i\in I}\right),$$

where $\sigma(x, y)^2$ is given by (3.4).

Proof. Let h be a bounded, Lipschitz continuous function on $\mathcal{D}(\mathbb{R}^{1+I})$ and Y be a bounded $\mathscr{F}$-measurable random variable. With $\sigma(x, y)^2$ given by (3.4), we abbreviate

$$\boldsymbol{U}^{v_t} := \left(U^{v_t,\eta_t}(x_i, y_i)\right)_{i\in I} \quad \text{and} \quad \boldsymbol{W} := \left(\mu'(x_i)\sigma(x_i, y_i)W^i\right)_{i\in I}.$$

We have to demonstrate

$$\mathbb{E}^\pi \left[h(L^t, \boldsymbol{U}^{v_t}_{L^t})Y\right] \xrightarrow[t\to\infty]{} \tilde{\mathbb{E}}\left[h\left(L, \boldsymbol{W}_L\right)\right] \mathbb{E}^\pi Y. \tag{4.23}$$

First, we suppose that Y is $\mathscr{F}_u$-measurable for some $u \geq 0$. Let a^t be given by $a^t_s = (s - ut^{-1})^+$. Then a^t converges to $a_s = s$ as $t \to \infty$. By Lemma 4.12, since a^t is non-random, $\mathscr{L}(a^t, L^t, \boldsymbol{U}^{v_t} \mid \mathbb{P}^\pi) \to \mathscr{L}(a, L, \boldsymbol{W} \mid \tilde{\mathbb{P}})$ weakly as $t \to \infty$. The paths of the limit process are a. s. continuous. By eq. (3.4) of Höpfner et al. (1990), therefore,

$$\mathscr{L}(a^t, L^t_{a^t}, \boldsymbol{U}^{v_t} \circ L^t_{a^t} \mid \mathbb{P}^\pi) \xrightarrow[t\to\infty]{\text{w}} \mathscr{L}(a, L, \boldsymbol{W}_L \mid \tilde{\mathbb{P}}).$$

Since $\mathbb{E}^\pi[h(L^t_{a^t} \circ \theta_u, (\boldsymbol{U}^{v_t} \circ L^t_{a^t}) \circ \theta_u)Y] = \mathbb{E}^\pi[\mathbb{E}^{X_u}[h(L^t_{a^t}, \boldsymbol{U}^{v_t} \circ L^t_{a^t})]Y]$ by the Markov property, and since $\mathbb{E}^\pi[\tilde{\mathbb{E}}[h(L, \boldsymbol{W}_L)]Y] = \tilde{\mathbb{E}}[h(L, \boldsymbol{W}_L)]\,\mathbb{E}^\pi Y$, consequently,

$$\mathbb{E}^\pi[h(L^t_{a^t} \circ \theta_u, (\boldsymbol{U}^{v_t} \circ L^t_{a^t}) \circ \theta_u)Y] \xrightarrow[t\to\infty]{} \tilde{\mathbb{E}}[h(L, \boldsymbol{W}_L)]\,\mathbb{E}^\pi Y.$$

For every $r > 0$, we note

$$\sup_{s\leq r} \left|L^t_s - L^t_{a^t_s} \circ \theta_u\right| = \sup_{s\leq r} \left|v_t^{-1} J_{st\wedge u}\right| \leq v_t^{-1} J_u \xrightarrow[t\to\infty]{\text{a.s.}} 0,$$

and

$$\sup_{s\leq r} \left\|(\boldsymbol{U}^{v_t} \circ L^t_{a^t_s}) \circ \theta_u - \boldsymbol{U}^{v_t} \circ L^t_s\right\|_\infty \leq \frac{\|g_1\|_\infty(\|g_2\|_\infty J_u + \eta^d_{2,t}\|F g_2^{\eta,y}\|_\infty u)}{\sqrt{v_t \eta^d_{1,t}\eta^d_{2,t}}} \xrightarrow[t\to\infty]{\text{a.s.}} 0.$$

Since h is Lipschitz, therefore,

$$\left|h(L^t, \boldsymbol{U}^{v_t} \circ L^t) - h(L^t_{a^t} \circ \theta_u, (\boldsymbol{U}^{v_t} \circ L^t_{a^t}) \circ \theta_u)\right| \xrightarrow[t\to\infty]{\text{a.s.}} 0.$$

Since h and Y are bounded, we deduce from Lebesgue's dominated convergence theorem that (4.23) holds for all bounded $\mathscr{F}_u$-measurable random variables Y.

Second, for arbitrary bounded $\mathscr{F}$-measurable Y, we have $\mathbb{E}^\pi[Y|\mathscr{F}_u] \to Y$ in $\mathcal{L}^1$ as $u \to \infty$. Consequently, again by Lebesgue's dominated convergence theorem,

$$\lim_{u\to\infty} \sup_{t>0} \left|\mathbb{E}^\pi \left[h(L^t, \boldsymbol{U}^{v_t} \circ L^t, \bar{\boldsymbol{U}}^{v_t} \circ L^t)(\mathbb{E}^\pi[Y|\mathscr{F}_u] - Y)\right]\right| = 0.$$

Thus, (4.23) holds in general. □

By Corollary 4.11 and by eq. (3.5) of Höpfner et al. (1990), we obtain the following corollary to Lemma 4.13.

4.14 Corollary. *Grant Assumptions 2.2, 2.4, 3.1 and 3.2. Let η_t be as in Lemma 4.12. Then we have the following stable convergence in law in $\mathcal{D}(\mathbb{R}^{2I})$:*

$$\left(S_{L^t}^{v_t,\eta_t}(x_i), U_{L^t}^{v_t,\eta_t}(x_i,y_i)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\text{st}} \left(\mu'(x_i)L, \mu'(x_i)\sigma(x_i,y_i)W_L^i\right)_{i\in I},$$

where $\sigma(x,y)^2$ is given by (3.4). □

Proof (of Theorem 3.7). For every $t \geq 0$ and each x and y, we have

$$\sqrt{v_t\eta_{1,t}^d\eta_{2,t}^d}\left(\hat{f}_t^{\eta_t}(x,y) - \bar{f}^{\eta_t}(x,y)\right) = \frac{U_{J_t/v_t}^{v_t,\eta_t}(x,y) - \bar{f}^{\eta_t}(x,y)\sqrt{\eta_{1,t}^d\eta_{2,t}^d/v_t}\int_{T_{J_t}}^t g_1^{\eta_t,x}(X_s)\mathrm{d}s}{S_{J_t/v_t}^{v_t,\eta_t}(x) + v_t^{-1}\int_{T_{J_t}}^t g_1^{\eta_t,x}(X_s)\mathrm{d}s},$$

where $\bar{f}^\eta(x,y) := \mu(g_1^{\eta,x}Fg_2^{\eta,y})/\mu(g_1^{\eta,x})$. Let $h_n : \mathbf{D} \to \mathbb{R}$ be as in the proof of Theorem 3.6. We recall $\psi(h_n^2) \leq \zeta\eta_{1,n}^{-d}$ for some $\zeta < \infty$. We also note $v_t\eta_{2,t}^{-d} \to \infty$. In analogy to (4.17), thus,

$$\sqrt{\eta_{1,t}^d\eta_{2,t}^d/v_t}\int_{T_{J_t}}^t g_1^{\eta_t,x}(X_s)\mathrm{d}s \leq \sqrt{\eta_{1,t}^d\eta_{2,t}^d/v_t}h_{v_t}(Z_{J_t+1}) \xrightarrow[t\to\infty]{\mathbb{P}^\pi} 0.$$

Since L and W are independent, $V(x_i,y_i) := L_1^{-1/2}W_{L_1}^i$ defines an I-dimensional standard Gaussian random vector such that L, V and $\mathscr{F}$ are independent. By the continuous mapping theorem and Corollary 4.14, consequently,

$$\left(\sqrt{v_t\eta_{1,t}^d\eta_{2,t}^d}\left(\hat{f}_t^{\eta_t}(x_i,y_i) - \bar{f}^{\eta_t}(x_i,y_i)\right)\right)_{i\in I} \xrightarrow[t\to\infty]{\mathscr{L}-\text{st}} \left(\sigma(x_i,y_i)V(x_i,y_i)L_1^{-1/2}\right)_{i\in I},$$

where $\sigma(x,y)^2$ is given by (3.4). ⋄

In addition, grant Assumption 3.3 and let $\eta_t = (\eta_{1,t}, \eta_{2,t})$ be such that (3.3) holds as well. We abbreviate $\bar{\gamma}^\eta(x,y) = \bar{f}^\eta(x,y) - f(x,y)$ and note

$$\mu(g_1^{\eta,x})\bar{\gamma}^\eta(x,y) = \iint \mu'(x+\eta_1 z)\Big(f(x+\eta_1 z, y+\eta_2 w) - f(x,y)\Big)g_1(z)g_2(w)\mathrm{d}w\mathrm{d}z.$$

We apply Taylor's theorem to μ' and f: In x, we expand up to the order $\lceil\alpha_1\rceil - 1$ and, in y, we expand up to the order $\lceil\alpha_2\rceil - 1$. We recall from (2.5) that g_1 and g_2 are, at least, of order α_1 and α_2, respectively. By a classical approximation argument, therefore, there exists a constant $\zeta < \infty$ such that $|\mu(g_1^{\eta,x})\bar{\gamma}^{\eta_t}(x,y)| \leq \zeta(\eta_{1,t}^{\alpha_1} + \eta_{2,t}^{\alpha_2})$. If $\zeta_1 = \zeta_2 = 0$ in (3.3), then it is immediate that $(v_t\eta_{1,t}^d\eta_{2,t}^d)^{1/2}\bar{\gamma}^{\eta_t}(x,y) \to 0$. If $\alpha_1, \alpha_2 \in \mathbb{N}^*$, more explicitly,

$$\mu(g_1^{\eta,x})\bar{\gamma}^\eta(x,y) = \eta_{1,t}^{\alpha_1}\sum_{\substack{|m_1+m_2|=\alpha_1\\|m_2|\neq 0}}\frac{\kappa_{m_1+m_2}(g_1)}{m_1!m_2!}\frac{\partial^{m_1}}{\partial x^{m_1}}\mu'(x)\frac{\partial^{m_2}}{\partial x^{m_2}}f(x,y)$$

$$+ \eta_{2,t}^{\alpha_2} \sum_{|m|=\alpha_2} \frac{\kappa_m(g_2)}{m!} \mu'(x) \frac{\partial^m}{\partial y^m} f(x,y) + o(\eta_{1,t}^{\alpha_1} + \eta_{2,t}^{\alpha_2}).$$

Since $\mu(g_1^{\eta,x}) \to \mu'(x)$, we have $(v_t \eta_{1,t}^d \eta_{2,t}^d)^{1/2} \bar{\gamma}^{\eta_t}(x,y) \to \gamma(x,y)$ given by (3.6). □

Proof (of Corollary 3.8). In analogy to the proof of Theorem 3.7, by Corollary 4.14 it remains to show that $(v_t \eta_{1,t}^d \eta_{2,t}^d)^{1/2} \hat{\gamma}_t^{\eta_t}(x,y)$ is a consistent estimator for $\gamma(x,y)$.

We recall that in classical (conditional) density estimation, the (partial) derivatives of a consistent density estimator – provided they exist – are consistent for the (partial) derivatives of the estimated density. In analogy to Lemma 4.8, we observe that this is also true in our context. In particular,

$$\frac{\partial^{m_1+m_2}}{\partial x^{m_1} \partial y^{m_2}} \hat{f}_t^{\eta_t}(x,y) \xrightarrow[t\to\infty]{\mathbb{P}^\pi} \frac{\partial^{m_1+m_2}}{\partial x^{m_1} \partial y^{m_2}} f(x,y) \quad \text{and} \quad \frac{\int_0^t \frac{\partial^m}{\partial x^m} g_1^{\eta_t,x}(X_s) \mathrm{d}s}{\int_0^t g_1^{\eta_t,x}(X_s) \mathrm{d}s} \xrightarrow[t\to\infty]{\mathbb{P}^\pi} \frac{\frac{\partial^m}{\partial x^m} \mu'(x)}{\mu'(x)}.$$

If either $\alpha_1, \alpha_2 \in \mathbb{N}^*$ or $\zeta_1 = \zeta_2 = 0$ in (3.3), consequently, $(v_t \eta_{1,t}^d \eta_{2,t}^d)^{1/2} \hat{\gamma}_t^{\eta_t}(x,y) \to \gamma(x,y)$ in probability as $t \to \infty$. □

5. Proofs for results of section 2

Throughout this section, $\zeta < \infty$ denotes some generic constant which may depend on the variables specified at the beginning of each proof. It may change from line to line.

This section is organised as follows: First, in section 5.1 we study the influence of discretisation on our estimator. We prove results for the small-time asymptotic of Itô semi-martingales and for the sojourn time discretisation error. Second, in section 5.2 we prove an auxiliary, non-standard martingale limit theorem. Third, in section 5.3 we prove the consistency of our estimator (Theorem 2.9) utilising our results from sections 4.3 and 5.1. Last, in section 5.4 we apply Theorem 5.5 from section 5.2 to our case and conclude with the final steps in the proof of the central limit theorem (Theorem 2.10 and Corollary 2.11) utilising our results from sections 4.4 and 5.1.

5.1. Small-time asymptotic and sojourn time discretisation error

In this subsection, we study the influence of discretisation.

We compare our estimators in Definitions 2.7 and 3.4: In the numerator of the former, the jumps ΔX_t and the pre-jump left-limits X_{t-} are replaced by the increments $\Delta_k^n X$ and the pre-increment values $X_{(k-1)\Delta}$, respectively. Our Itô semi-martingale meets the following small-time asymptotic:

5.1 Proposition. *Let A be a compact subset of $E \times E^*$, $\eta_0 < \min\{\|y\| : (x,y) \in A\}$, and let g be a twice continuously differentiable kernel with compact support. Grant*

Assumptions 2.1 and 2.3. Then, for every $m \in \mathbb{N}^*$, there exists $\zeta < \infty$ such that

$$\begin{aligned} &\left| \frac{1}{\Delta} \mathbb{E}^x \left[g^{\eta,y}(\Delta_1^n X) \right] - \int F(x, \mathrm{d}w) g^{\eta,y}(w) \right| \\ &\qquad \leq \zeta \left[\Delta^{(\alpha \wedge 1)/2} + \frac{\Delta}{\eta^{2\vee(\beta+d)}} \left(1 + \sum_{k=1}^{m} \frac{\Delta^k}{\eta^{2k}} \right) + \frac{\Delta^m}{\eta^{2(m+1)+d}} \right] \end{aligned} \tag{5.1}$$

holds for every $(x, y) \in A$, $\eta < \eta_0$ and $\Delta \leq 1$, where $g^{\eta,y}(w) = \eta^{-d} g((w - y)/\eta)$.

Remark. For presentational purposes, we have left a small gap in the finite activity case. For instance, if f is locally bounded on $E \times E$, then we can improve the bound in (5.1) replacing $\eta^{2\vee(\beta+d)}$ by η^2 independently of the dimension d.

In the former estimator's denominator, the sojourn time $\int_0^t g_1^{\eta,x}(X_s)\mathrm{d}s$ is replaced by its Riemann sum approximation $\Delta \sum_{k=1}^n g_1^{\eta,x}(X_{(k-1)\Delta})$.

5.2 Proposition. *Let $x \in E$, $v : \mathbb{R}_+ \to \mathbb{R}_+$ be a non-decreasing function, $\xi_n > 0$, $\eta_n \to 0$, and $(h_n)_{n\in\mathbb{N}^*}$ be a uniformly bounded family of twice continuously differentiable functions supported on $B_{\eta_n}(x)$ such that $(\eta_n^{|m|}\partial^m h_n)_{n\in\mathbb{N}^*}$ is uniformly bounded for every multi-index m with $|m| \in \{1, 2\}$. As $n\Delta \to \infty$ and $\Delta \to 0$, we suppose $v(n\Delta)\eta_n^d \to \infty$ and $\xi_n \Delta \eta_n^{-2-d[(1-2/(\beta+d))\vee 0]} \to 0$.*

(i) Grant Assumptions 2.1 to 2.3. If $n\Delta^2\xi_n \to 0$ and $v(s) = \bar{v}(st)$ for some deterministic equivalent $\bar{v}$ of X and some $t > 0$, then, under any law $\mathbb{P}^\pi$, we have the following convergence in probability:

$$\sup_{s\leq t} \frac{\xi_n}{v(n\Delta)\eta_n^d} \left| \Delta \sum_{k=1}^{\lfloor sn \rfloor} h_n(X_{(k-1)\Delta}) - \int_0^{\lfloor sn \rfloor \Delta} h_n(X_r)\mathrm{d}r \right| \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0. \tag{5.2}$$

(ii) Grant Assumptions 2.1 to 2.4. If $(n\Delta)^{1-\delta}\Delta\xi_n \to 0$ and v is the regularly varying function from (2.4), then, under any law $\mathbb{P}^\pi$, (5.2) holds for all $t > 0$.

Before we turn to the proofs of Propositions 5.1 and 5.2, we present two auxiliary upper bounds for the small-time asymptotic of Itô semi-martingales. Below, we heavily utilise results and notation from the books Jacod and Shiryaev (2003) (esp., Chapter II) and Jacod and Protter (2012) (esp., Section 2.1).

We recall that our underlying process X is an Itô semi-martingale with absolutely continuous characteristics $(B, C, \mathfrak{n})$ satisfying (2.2), and that its jump measure $\mathfrak{m}$ is the random measure on $\mathbb{R}_+ \times E$ given by $\mathfrak{m}(\mathrm{d}t, \mathrm{d}x) := \sum_{\{s:\Delta X_s \neq 0\}} \epsilon_{(s,\Delta X_s)}(\mathrm{d}t, \mathrm{d}x)$. For a function g on $\Omega \times \mathbb{R}_+ \times E$, we define the stochastic integrals

$$g \star \mathfrak{m}_t := \int_{[0,t]\times E} g(\omega, s, w)\mathfrak{m}(\omega; \mathrm{d}s, \mathrm{d}w) \quad \text{and} \quad g \star \mathfrak{n}_t := \int_{[0,t]\times E} g(\omega, s, w)\mathfrak{n}(\omega; \mathrm{d}s, \mathrm{d}w),$$

and also the purely discontinuous martingale $g \star (\mathfrak{m}-\mathfrak{n})_t$, as soon as these integrals are well-defined. By Lévy–Itô and Grigelionis decomposition, we can assume w. l. o. g. that there exists a d-dimensional Wiener process W, defined on $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, (\mathbb{P}^x)_{x\in E})$, and an $E\otimes E$-valued function σ with $c = \sigma\sigma^\intercal$ such that

$$X_t = X_0 + \int_0^t b(X_s)\mathrm{d}t + \int_0^t \sigma(X_s)\mathrm{d}W_s + (w\mathbb{1}_{\|w\|\leq 1}) \star (\mathfrak{m}-\mathfrak{n})_t + (w\mathbb{1}_{\|w\|>1}) \star \mathfrak{m}_t.$$

Itô's formula plays a crucial role in the sequel. By a version derived from (2.1.20) of Jacod and Protter (2012), if $g : E \to \mathbb{R}$ is twice continuously differentiable, then

$$\begin{aligned} g(X_t) = g(X_0) &+ \int_0^t b(X_s)^\intercal \nabla g(X_s)\mathrm{d}s + \frac{1}{2}\int_0^t \operatorname{tr}\Big(c(X_s)\nabla^2 g(X_s)\Big)\mathrm{d}s \\ &+ \Big(g(X_- + w) - g(X_-) - w^\intercal \nabla g(X_-)\Big)\,\mathbb{1}_{\|w\|\leq 1} \star \mathfrak{n}_t \\ &+ \int_0^t \sigma(X_s)\mathrm{d}W_s + \Big(g(X_- + w) - g(X_-)\Big)\mathbb{1}_{\|w\|\leq 1} \star (\mathfrak{m}-\mathfrak{n})_t \\ &+ \Big(g(X_- + w) - g(X_-)\Big)\mathbb{1}_{\|w\|>1} \star \mathfrak{m}_t, \end{aligned} \tag{5.3}$$

where $\operatorname{tr}(\cdot)$ denotes the trace operator on $E\otimes E$ and $\nabla^2 g$ denotes the Hessian of g.

For $\xi > 0$, we denote by $T^\xi := \inf\{t > 0 : \|\Delta X_t\| > \xi\}$ the first time of a jump greater than ξ . Also, we introduce the following decomposition of our semi-martingale X:

$$X_t = X_0 + X_t^\xi + X_t'^\xi, \quad \text{where } X_t'^\xi := (w\mathbb{1}_{\|w\|>\xi}) \star \mathfrak{m}_t = \sum_{s\leq t} \Delta X_s \mathbb{1}_{\|\Delta X_s\|>\xi}.$$

We note that X^ξ and X'^ξ are again Itô semi-martingales; we denote their characteristics by $(B^\xi, C, \mathfrak{n}^\xi)$ and $(B'^\xi, 0, \mathfrak{n}'^\xi)$, respectively. Furthermore, we decompose X^ξ into drift B^ξ, continuous martingale part M^c, and purely discontinuous martingale part M^ξ. These are given by

$$B_t^\xi = \int_0^t b^\xi(X_s)\mathrm{d}s, \quad M_t^c = \int_0^t \sigma(X_s)\mathrm{d}W_s \quad \text{and} \quad M_t^\xi = (w1_{\|w\|\leq\xi}) \star (\mathfrak{m}-\mathfrak{n})_t,$$

where $b^\xi(x) = b(x) - \int_{\xi<\|w\|\leq 1} F(x,\mathrm{d}w)w$ if $\xi < 1$, and $b^\xi(x) = b(x) + \int_{1<\|w\|\leq\xi} F(x,\mathrm{d}w)w$ if $\xi \geq 1$. Under Assumption 2.1, we derive the following two lemmata.

5.3 Lemma. *Let $\xi_0 > 0$ and $p \geq 2$. Grant Assumption 2.1. Then, there exists a constant $\zeta < \infty$ such that, for every $0 < \xi \leq \xi_0$, $x \in E$, and $t \leq 1$, we have*

$$\mathbb{E}^x \sup_{s\leq t} \|X^\xi_{s\wedge T^\xi}\|^p \leq \zeta(1 + \|x\|^p)t.$$

Proof. In this proof, $\zeta < \infty$ may depend on ξ_0 and p but neither on t, x, ξ nor ζ'.

First, let $1 \leq \xi \leq \xi_0$. We emphasise that, in this case,

$$\|b^\xi(x)\| \leq \|b(x)\| + \xi_0^{d+1} F(x, \{1 < \|w\| \leq \xi_0\}). \tag{5.4}$$

By (2.2), we have $\mathfrak{n}^\xi(\mathrm{d}t, A) = \mathrm{d}t F^\xi(X_t, A) := \mathrm{d}t F(X_t, A \cap B_\xi(0))$ for every Borel set A. By construction, $X_t'^\xi = 0$ on $\{t < T^\xi\}$. By (2.1.43) of Jacod and Protter (2012), thus,

$$\begin{aligned}\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi}\|^p \le \zeta\,\mathbb{E}^x\Big[t^{p-1}\int_0^t \|b^\xi(X_0 + X^\xi_{s\wedge T^\xi})\|^p \mathrm{d}s + t^{p/2-1}\int_0^t \|c(X_0+X^\xi_{s\wedge T^\xi})\|^{p/2}\mathrm{d}s\Big]\\ + \zeta\,\mathbb{E}^x \int_0^t \mathrm{d}s \int F^{\xi_0}(X_0 + X^\xi_s, \mathrm{d}w)\|w\|^p\\ + \zeta\,\mathbb{E}^x\, t^{p/2-1}\int_0^t \mathrm{d}s \Big(\int F^{\xi_0}(X_0+X^\xi_s, \mathrm{d}w)\|w\|^2\Big)^{p/2}.\end{aligned}$$

Under Assumption 2.1, for all $t \le 1$, we observe

$$\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi}\|^p \le \zeta \int_0^t (1 + \mathbb{E}^x\|X_0 + X^\xi_{s\wedge T^\xi}\|^p)\mathrm{d}s.$$

For $\zeta' > 0$, let $S^{\zeta'} := \inf\{s > 0 : \|X^\xi_s\| > \zeta'\}$. Then

$$\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi\wedge S^{\zeta'}}\|^p \le \zeta \int_0^t (1 + \mathbb{E}^x\|X_0 + X^\xi_{s\wedge T^\xi\wedge S^{\zeta'}}\|^p)\mathrm{d}s,$$

where we note $\sup_{s\le t}\|X^\xi_{s\wedge T^\xi\wedge S^{\zeta'}}\| \le \zeta' + \xi$. By the Grönwall–Bellmann inequality, thus,

$$\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi\wedge S^{\zeta'}}\|^p \le \zeta(1+\|x\|^p)\Big(t + \int_0^t \zeta \mathrm{e}^{\zeta(t-s)}\mathrm{d}s\Big) = \zeta(1+\|x\|^p)(\mathrm{e}^{\zeta t} - 1).$$

Since $S^{\zeta'} \wedge T^\xi \to T^\xi$ as $\zeta' \to \infty$, consequently, $\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi}\|^p \le \zeta(1 + \|x\|^p)t$.

Second, let $0 < \xi < 1$. We note that $X^\xi_t \mathbb{1}_{t<T^\xi} = (X_t - X_0)\mathbb{1}_{t<T^\xi}$ holds, and that X^ξ is continuous at T^ξ outside the null set $\{\|\Delta X_{T^\xi}\| = \xi\}$. As $T^\xi \le T^1$ for all ω, thus,

$$\sup_{s\le t}\|X^\xi_{s\wedge T^\xi}\| = \sup_{s\le t}\|(X_s - X_0)\mathbb{1}_{s<T^\xi}\| \le \sup_{s\le t}\|(X_s - X_0)\mathbb{1}_{s<T^1}\| = \sup_{s\le t}\|X^1_{s\wedge T^1}\|$$

almost surely. By case $\xi \ge 1$, consequently, $\mathbb{E}^x \sup_{s\le t}\|X^\xi_{s\wedge T^\xi}\|^p \le \zeta(1 + \|x\|^p)t$. □

5.4 Lemma. *Let $y \neq 0$ and $\eta_0 < \|y\|$. Grant Assumption 2.1. Then, for every $m \in \mathbb{N}^*$, there exists a constant $\zeta < \infty$ – non-increasing in $\|y\|$ – such that, for every $x \in E$, $\eta < \eta_0$, and $t \le 1$,*

$$\begin{aligned}&\mathbb{P}^x(X_t \in B_\eta(X_0 + y))\\ &\quad \le \zeta\Big(1 + \|x\|^{2(m+1)} + \|y\|^{2(m+1)}\Big)\Big[t\eta^d\Big(1 + \sum_{k=1}^m t^k \eta^{-2\vee(\beta+d)-2(k-1)}\Big) + \frac{t^m}{\eta^{2m}}\Big].\end{aligned} \tag{5.5}$$

Proof. Let $1 < \zeta' < (\|y\|/\eta_0)^{1/(m+1)}$, $\varepsilon := (\zeta'^{m+1}\eta_0 - \zeta'^m\eta_0)/6 > 0$ and $\xi < \varepsilon/2$. In addition, let g be a $\mathcal{C}^2$-kernel such that $\mathbb{1}_{B_1(0)} \le g \le \mathbb{1}_{B_{(\zeta'+1)/2}(0)}$. We set $g_\eta(z) = g((z - x - y)/\eta)$ and abbreviate

$$h(t, \eta) := \mathbb{P}^x(X_t \in B_\eta(x + y)) \le \mathbb{E}^x\, g_\eta(X_t).$$

In this proof, $\zeta < \infty$ may depend on η_0, ζ', β and m, but neither on x, t nor η.

By Itô's formula (5.3), we have $h(t,\eta) \le |H_t^\eta| + |H_t'^\eta| + |H_t''^\eta|$, where

$$H_t^\eta := \mathbb{E}^x \int_0^t b(X_s)^\intercal \nabla g_\eta(X_s) ds + \frac{1}{2}\mathbb{E}^x \int_0^t \operatorname{tr}\left(c(X_s)\nabla^2 g_\eta(X_s)\right)\mathrm{d}s,$$
$$H_t'^\eta := \mathbb{E}^x \int_0^t \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)}(X_s) \int F(X_s,\mathrm{d}w)\{g_\eta(X_s+w) - g_\eta(X_s) - w^\intercal \nabla g_\eta(X_s)\mathbb{1}_{\|w\|\le 1}\},$$
$$H_t''^\eta := \mathbb{E}^x \int_0^t \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_s) \int F(X_s,\mathrm{d}w) g_\eta(X_s+w).$$

Under Assumption 2.1, $b(z)$ and $c(z)$ are bounded in norm by $\zeta(1+\|z\|^2)$. Moreover, the gradient and Hessian of g_η vanish outside $B_{(\zeta'+1)\eta/2}(x+y)$ and satisfy $\|\partial_i g_\eta\| \le \zeta\eta^{-1}$ and $\|\partial_{ij} g_\eta\| \le \zeta\eta^{-2}$. Hence,

$$|H_t^\eta| \le \zeta(1+\|x\|^2+\|y\|^2)\eta^{-2}\,\mathbb{E}^x \int_0^t \mathbb{1}_{B_{(\zeta'+1)\eta/2}(x+y)}(X_s)\mathrm{d}s.$$

For $z \in B_{\zeta'\eta}(x+y)$, furthermore,

$$\int F(z,\mathrm{d}w)\{g_\eta(z+w) - g_\eta(z) - w^\intercal \nabla g_\eta(z)\mathbb{1}_{\|w\|\le 1}\} \le \frac{\zeta(1+\|z\|)}{\eta^2}\int \bar{F}(\mathrm{d}w)(1\wedge \|w\|^2).$$

Therefore,

$$|H_t^\eta| + \left|H_t'^\eta\right| \le \frac{\zeta(1+\|x\|^2+\|y\|^2)}{\eta^2}\int_0^t h(s,\zeta'\eta)\mathrm{d}s. \tag{5.6}$$

Suppose that $|H_t''^\eta| \le \zeta(1+\|x\|^3+\|y\|^3)(t\eta^d + t^2\eta^{-\beta})$ holds. Then,

$$h(t,\eta) \le \zeta(1+\|x\|^3+\|y\|^3)t\eta^d(1+t\eta^{-(\beta+d)}) + \frac{\zeta(1+\|x\|^2+\|y\|^2)}{\eta^2}\int_0^t h(s,\zeta'\eta)\mathrm{d}s.$$

By iteration, we obtain (5.5) after m steps. ⋄

It remains to prove $|H_t''^\eta| \le \zeta(1+\|x\|^3+\|y\|^3)(t\eta^d+t^2\eta^{-\beta})$. Under Assumption 2.1 (iii), on the one hand, we have

$$\int F(z,\mathrm{d}w)g_\eta(z+w) \le \zeta(1+\|z\|)\eta^d \int \bar{f}(y+x-z+\eta w)g(w)\mathrm{d}w$$
$$\le \begin{cases} \zeta(1+\|x\|)\eta^d, & \text{if } z\in B_{3\varepsilon}(x), \\ \zeta(1+\|x+y\|)\eta^d & \text{if } z \in B_{1+\zeta'\eta}(x+y)^{\mathrm{c}}. \end{cases}$$

For $z \in B_{1+\zeta'\eta}(x+y)\setminus B_{\zeta'\eta}(x+y)$, on the other hand, we have

$$\int F(z,\mathrm{d}w)g_\eta(z+w) \le \frac{\zeta(1+\|z\|)}{((\zeta'-1)\eta/2)^\beta}\int \mathrm{d}w g\left(\frac{w+z-x-y}{\eta}\right)\bar{f}(w)\|w\|^\beta.$$

Since $\eta^d \le \eta^{-\beta}$ and $\int \bar{F}(\mathrm{d}w)(\|w\|^\beta \wedge 1) < \infty$ by assumption, thus,

$$\int F(z, \mathrm{d}w) g_\eta(z+w) \le \begin{cases} \zeta(1+\|x+y\|)\eta^{-\beta}, & \text{if } z \in B_{\zeta'\eta}(x+y)^{\mathrm{c}}, \\ \zeta(1+\|x\|)\eta^d, & \text{if } z \in B_{3\varepsilon}(x). \end{cases} \tag{5.7}$$

Let $S^{\varepsilon,\xi} := \inf\{t > 0 : \|X_t^\xi\| > 3\varepsilon\}$, and $\Omega_t^{\varepsilon,\xi} := \{S^{\varepsilon,\xi} \le T^\xi \wedge t\}$. We split the set $\Omega \times [0,t]$ into $A_1 := \Omega \times [\![0, t\wedge T^\xi \wedge S^{\varepsilon,\xi}[\![$, $A_2 := (\Omega_t^{\varepsilon,\xi})^{\mathrm{c}} \times [\![T^\xi \wedge t, t]\!]$ and $A_3 := \Omega_t^{\varepsilon,\xi} \times [\![S^{\eta,\xi}, t]\!]$. Then we obtain the following:

First: Since $\sup_{s\le t}\|X^\xi_{s\wedge T^\xi\wedge S^{\varepsilon,\xi}} - X_0\| \le 3\varepsilon$, by (5.7), we obtain

$$\iint_{A_1} \mathrm{d}\mathbb{P}^x\, \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_s) \int F(X_s, \mathrm{d}w) g_\eta(X_s+w) \le \zeta(1+\|x\|)t\eta^d.$$

Second: Under Assumption 2.1, we have

$$\mathbb{P}^x(T^\xi \le t \wedge S^{\varepsilon,\xi}) \le \mathbb{E}^x \int_0^t \mathrm{d}s \mathbb{1}_{B_{3\varepsilon}(x)}(X_s) F(X_s, \|w\| > \xi) \le \zeta(1+\|x\|)t.$$

By the Markov property and (5.7), therefore,

$$\begin{aligned} &\iint_{A_2} \mathrm{d}\mathbb{P}^x\, \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_s) \int F(X_s, \mathrm{d}w) g_\eta(X_s+w) \\ &\qquad \le \mathbb{E}^x\, \mathbb{1}_{\{T^\xi \le t\wedge S^{\varepsilon,\xi}\}}\, \mathbb{E}^{X_{T^\xi}} \int_0^t \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_s) \int F(X_s, \mathrm{d}w) g_\eta(X_s+w) \\ &\qquad \le \zeta(1+\|x+y\|)t\eta^{-\beta}\, \mathbb{P}^x(T^\xi \le t\wedge S^{\varepsilon,\xi}) \\ &\qquad \le \zeta(1+\|x\|^2+\|y\|^2)t^2\eta^{-\beta}. \end{aligned} \tag{5.8}$$

Third: By Lemma 5.3, we have $\mathbb{P}^x(\Omega_t^{\varepsilon,\xi}) \le \zeta(1+\|x\|^2)t$. By the Markov property and (5.7), therefore,

$$\begin{aligned} \iint_{A_3} \mathrm{d}\mathbb{P}^x\, \mathrm{d}s \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_s) &\int F(X_s, \mathrm{d}w) g_\eta(X_s+w) \\ &\le \zeta(1+\|x+y\|)t\eta^{-\beta}\, \mathbb{P}^x(\Omega_t^{\varepsilon,\xi}) \\ &\le \zeta(1+\|x\|^3+\|y\|^3)t^2\eta^{-\beta}. \end{aligned} \tag{5.9}$$

□

We turn to the proofs of Propositions 5.1 and 5.2.

Proof (of Proposition 5.1). Let $1 < \zeta' < (\min\{\|y\| : (x,y) \in A\}/\eta_0)^{1/(m+2)}$, and $\varepsilon, \xi > 0$ be given as in the proof of Lemma 5.4. In this proof, $\zeta < \infty$ may depend on η_0, ζ', β, m and the set A, but neither on x, y, Δ nor η.

Let $\eta \le \eta_0$, and $(x,y) \in A$. W.l.o.g., we assume that g is supported on $B_1(0)$. To avoid cumbersome notation, we abbreviate $h_\eta = g^{\eta,x+y}$. From (2.2) and Itô's formula (5.3), we obtain $\mathbb{E}^x\, h_\eta(X_\Delta) = H_\Delta^\eta + H_\Delta'^\eta + H_\Delta''^\eta$, where

$$H_\Delta^\eta = \mathbb{E}^x \int_0^\Delta b(X_t)^\intercal \nabla h_\eta(X_t)\mathrm{d}t + \frac{1}{2}\,\mathbb{E}^x \int_0^\Delta \operatorname{tr}\left(c(X_t)\nabla^2 h_\eta(X_t)\right)\mathrm{d}t,$$

$$H'^{\eta}_{\Delta} = \mathbb{E}^x \int_0^\Delta \mathrm{d}t \mathbb{1}_{B_{\zeta'\eta}(x+y)}(X_t) \int F(X_t, \mathrm{d}w)\{h_\eta(X_t + w) - h_\eta(X_t) - w^\intercal \nabla h_\eta(X_t)\mathbb{1}_{\|w\|\le 1}\},$$
$$H''^{\eta}_{\Delta} = \mathbb{E}^x \int_0^\Delta \mathrm{d}t \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_t) \int F(X_t, \mathrm{d}w) h_\eta(X_t + w).$$

By (5.6), we observe

$$|H^{\eta}_{\Delta}| + \left|H'^{\eta}_{\Delta}\right| \le \frac{\zeta}{\eta^{d+2}} \int_0^\Delta \mathbb{P}^x(X_t \in B_{\zeta'\eta}(x+y))\mathrm{d}t.$$

By the choice of ζ', Lemma 5.4 implies

$$|H^{\eta}_{\Delta}| + \left|H'^{\eta}_{\Delta}\right| \le \zeta \left[\frac{\Delta^2}{\eta^2}\left(1 + \sum_{k=1}^m \frac{\Delta^k}{\eta^{2\vee(\beta+d)+2(k-1)}}\right) + \frac{\Delta^{m+1}}{\eta^{2(m+1)+d}}\right]. \tag{5.10}$$

Suppose

$$\left|H''^{\eta}_{\Delta} - \int F(x, \mathrm{d}w) h_\eta(x+w) \int_0^\Delta \mathbb{P}^x(X_t \notin B_{\zeta'\eta}(x+y))\right| \le \zeta(\Delta^{1+(\alpha\wedge1)/2} + \Delta^2\eta^{-(\beta+d)}). \tag{5.11}$$

Combining (5.10) and (5.11), we obtain (5.1). ⋄

It remains to prove (5.11). By (5.7), we observe

$$\int F(z, \mathrm{d}w) h_\eta(z+w) \le \begin{cases} \zeta\eta^{-(\beta+d)}, & \text{if } z \in B_{\zeta'\eta}(x+y)^{\mathrm{c}}, \\ \zeta, & \text{if } z \in B_{3\varepsilon}(x). \end{cases} \tag{5.12}$$

Let the stopping time $S^{\varepsilon,\xi}$, and the event $\Omega^{\varepsilon,\xi}_\Delta$ be given as in the proof of Lemma 5.4. We split the set $\Omega \times [0, \Delta]$ into $A_1 := \Omega \times [\![0, \Delta \wedge T^\xi \wedge S^{\varepsilon,\xi}[\![$, $A_2 := (\Omega^{\varepsilon,\xi}_\Delta)^{\mathrm{c}} \times [\![T^\xi \wedge \Delta, \Delta]\!]$ and $A_3 := \Omega^{\varepsilon,\xi}_\Delta \times [\![S^{\eta,\xi}, \Delta]\!]$. For convenience, we also abbreviate

$$\tilde{f}^{\eta}_{x,y}(z, w) := f(z, y + x - z + \eta w) - f(x, y + \eta w).$$

Then we obtain, first: By the choice of ε, we have that the convex hull of the set

$$\{(z, y + (x - z) + \eta w) : (x, y) \in A, \|z - x\| \le 3\varepsilon, \|w\| \le 1\}$$

is a compact subset of $E \times E^*$. By Assumption 2.3 and for all $(z, w) \in B_{3\varepsilon}(x) \times B_1(0)$, we have $|\tilde{f}^{\eta}_{x,y}(z, w)| \le \zeta\|z - x\|^{\alpha\wedge1}$. By Lemma 5.3, therefore,

$$\iint_{A_1} \mathrm{d}\mathbb{P}^x \,\mathrm{d}t \int \mathrm{d}w g(w) \tilde{f}^{\eta}_{x,y}(X_t, w) \le \zeta\Delta\, \mathbb{E}^x \sup_{t\le\Delta} \|X^\xi_{t\wedge T^\xi \wedge S^{\varepsilon,\xi}}\| \le \zeta\Delta^{1+(\alpha\wedge1)/2}.$$

Second and third: We compare (5.7) and (5.12). In analogy to (5.8) and (5.9), respectively, by the Markov property and (5.12), therefore,

$$\iint_{A_i} \mathrm{d}\mathbb{P}^x \,\mathrm{d}t \mathbb{1}_{B_{\zeta'\eta}(x+y)^{\mathrm{c}}}(X_t) \int \mathrm{d}w g(w) \tilde{f}^{\eta}_{x,y}(X_t, w) \le \zeta\Delta^2\eta^{-(\beta+d)},$$

for $i \in \{2, 3\}$. In summary, we proved (5.11). □

Proof (of Proposition 5.2). W. l. o. g., we assume $\eta < 1/4$. In this proof, $\zeta < \infty$ may neither depend on n, Δ nor η.

By Itô's formula (5.3), we observe

$$\frac{\xi_n}{v_{n\Delta}\eta_n^d}\left|\int_0^{\lfloor sn\rfloor\Delta} h_n(X_r)\mathrm{d}r - \Delta\sum_{k=1}^{\lfloor sn\rfloor} h_n(X_{(k-1)\Delta})\right| \leq |H_s^n| + |H_s'^n| + |H_s''^n| + |M_s^n|,$$

where

$$H_s^n := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\sum_{k=1}^{\lfloor sn\rfloor}\int_{(k-1)\Delta}^{k\Delta}\mathrm{d}t\int_{(k-1)\Delta}^{t}\left(b(X_r)^\intercal\nabla h_n(X_r) + \frac{1}{2}\operatorname{tr}\left(c(X_r)\nabla^2 h_n(X_r)\right)\right)\mathrm{d}r,$$

$$H_s'^n := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\sum_{k=1}^{\lfloor sn\rfloor}\int_{(k-1)\Delta}^{k\Delta}\mathrm{d}t\int_{(k-1)\Delta}^{t}\mathrm{d}r$$
$$\int_{\|w\|\leq 1} F(X_r,\mathrm{d}w)\{h_n(X_r+w) - h_n(X_r) - w^\intercal\nabla h_n(X_r)\},$$

$$H_s''^n := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\sum_{k=1}^{\lfloor sn\rfloor}\int_{(k-1)\Delta}^{k\Delta}\mathrm{d}t\sum_{(k-1)\Delta<r\leq t}\mathbb{1}_{\|\Delta X_r\|>1}\{h_n(X_{r-}+\Delta X_r) - h_n(X_{r-})\},$$

and

$$M_s^n := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\sum_{k=1}^{\lfloor sn\rfloor}\int_{(k-1)\Delta}^{k\Delta}\mathrm{d}t\Bigg(\int_{(k-1)\Delta}^{t}\nabla h_n(X_r)^\intercal\sigma(X_r)\mathrm{d}W_r$$
$$+\int_{(k-1)\Delta}^{t}\int_{\|w\|\leq 1}\{h_n(X_{r-}+w) - h_n(X_{r-})\}(\mathfrak{m}-\mathfrak{n})(\mathrm{d}r,\mathrm{d}w)\Bigg).$$

It remains to show:

(i) Under Assumptions 2.1 to 2.3, if $v(s) = \bar{v}(st)$ for some deterministic equivalent $\bar{v}$ of X and some $t > 0$, and if $n\Delta^2\xi_n \to 0$, then H_s^n, $H_s'^n$, $H_s''^n$ and M_s^n converge to zero uniformly on $\{0 \leq s \leq t\}$ in probability.

(ii) Under Assumptions 2.1 to 2.4, if v is the regularly varying function from (2.4), and if $(n\Delta)^{1-\delta}\Delta\xi_n \to 0$, then H_s^n, $H_s'^n$, $H_s''^n$ and M_s^n converge to zero uniformly for $\{0 \leq s \leq t\}$ in probability for all $t > 0$.

(a) Under Assumption 2.1, $b(z)$ and $c(z)$ are bounded in norm by $\zeta(1+\|z\|^2)$. Moreover, the gradient and Hessian of h_n vanish outside $B_{\eta_n}(x)$ and satisfy $\|\partial_i h_n\| \leq \zeta\eta_n^{-1}$ and $\|\partial_{ij}h_n\| \leq \zeta\eta^{-2}$, by assumption. Thus,

$$\left|b(z)^\intercal\nabla h_n(z) + \frac{1}{2}\operatorname{tr}\left(c(z)\nabla^2 h_n(z)\right)\right| \leq \zeta(1+\|z\|)\eta^{-2}\mathbb{1}_{B_{\eta_n}(x)}(z).$$

By Fubini's theorem, therefore,

$$\sup_{r\leq s}|H_r^n| \leq \zeta(1+\|x\|^2)\frac{\Delta\xi_n}{\eta^2}S_s'^{n,\Delta,\eta_n},\quad \text{where } S_s'^{n,\Delta,\eta} = \frac{1}{v_{n\Delta}\eta^d}\int_0^{\lfloor sn\rfloor\Delta}\mathbb{1}_{B_\eta(x)}(X_r)\mathrm{d}r.$$

In case (i), we deduce from Lemma 4.8 that the family $\{\mathscr{L}(S_t^{\prime n,\Delta,\eta_n} \mid \mathbb{P}^x) : n \in \mathbb{N}^*\}$ is tight under Assumptions 2.2 and 2.3. As $\Delta\xi_n\eta_n^{-2} \to 0$, $\sup_{s\le t}|H_s^n| \to 0$ in probability. In case (ii), we obtain from Corollary 4.11 that $S^{\prime n,\Delta,\eta_n}$ converges stably in law to a non-trivial process. As $\Delta\xi_n\eta_n^{-2} \to 0$, $\sup_{s\le t}|H_s^n| \to 0$ in probability for all $t > 0$. ⋄

(b) Let $\zeta' > 1$ and $\kappa = 1 \wedge 2/(\beta + d)$. Under Assumption 2.1, we have

$$\left|\int_{\|w\|\le 1} F(z, \mathrm{d}w)\{h_n(z+w) - h_n(z) - w^\intercal \nabla h_n(z)\}\right| \le \begin{cases} \zeta(1+\|z\|)\eta_n^{-2}\int_{\|w\|\le 1}\bar{F}(\mathrm{d}w)\|w\|^2, & \text{for } \|z-x\| \le \zeta'\eta_n^\kappa,\\ \zeta(1+\|z\|)\eta_n^{-\kappa\beta}\int_{\|w\|\le 1}\bar{F}(\mathrm{d}w)\|w\|^\beta, & \text{for } \zeta'\eta_n^\kappa < \|z-x\| \le 1+\eta_n,\\ 0, & \text{else.}\end{cases} \tag{5.13}$$

Again by Fubini's theorem, therefore,

$$\sup_{t\le s}|H_t^{\prime n}| \le \zeta(1+\|x\|)\left(\frac{\Delta\xi_n\eta_n^{\kappa d}}{\eta_n^{d+2}} S_s^{\prime n,\Delta,\zeta'\eta_n^\kappa} + \frac{\Delta\xi_n}{\eta^{d+\kappa\beta}} S_s^{\prime n,\Delta,1+\eta_n}\right).$$

In analogy to step (a), since $\Delta\xi_n\eta_n^{-2-d(1-\kappa)} \to 0$, $H_s^{\prime n} \to 0$ uniformly on $\{0 \le s \le t\}$ in probability in case (i); and for all $t > 0$ in case (ii). ⋄

(c) In analogy to steps (a) and (b), we note

$$\begin{aligned}|H_s^{\prime\prime n}| &\le \xi_n\Delta(v_{n\Delta}\eta_n^d)^{-1}(|h_n(X_- + w)| + |h_n(X_-)|)\mathbb{1}_{\|w\|>1} \star \mathfrak{m}_{\lfloor sn\rfloor\Delta}\\ &\le |K_s^n| + |N_{\lfloor sn\rfloor/n}^n| + |K_s^{\prime n}| + |N_{\lfloor sn\rfloor/n}^{\prime n}|,\end{aligned}$$

where

$$\begin{aligned}K_s^n &:= \xi_n\Delta(v_{n\Delta}\eta_n^d)^{-1}|h_n(X_- + w)|\mathbb{1}_{\|w\|>1} \star \mathfrak{n}_{\lfloor sn\rfloor\Delta},\\ K_s^{\prime n} &:= \xi_n\Delta(v_{n\Delta}\eta_n^d)^{-1}|h_n(X_-)|\mathbb{1}_{\|w\|>1} \star \mathfrak{n}_{\lfloor sn\rfloor\Delta},\\ N_s^n &:= \xi_n\Delta(v_{n\Delta}\eta_n^d)^{-1}|h_n(X_- + w)|\mathbb{1}_{\|w\|>1} \star (\mathfrak{m} - \mathfrak{n})_{sn\Delta},\\ N_s^{\prime n} &:= \xi_n\Delta(v_{n\Delta}\eta_n^d)^{-1}|h_n(X_-)|\mathbb{1}_{\|w\|>1} \star (\mathfrak{m} - \mathfrak{n})_{sn\Delta}.\end{aligned}$$

Under Assumption 2.1, since $\int_{\|w\|>1} F(z, \mathrm{d}w)|h_n(z+w)| = 0$ for $z \in B_{1-2\eta_n}(x)$, we have

$$\int_{\|w\|>1} F(z, \mathrm{d}w)|h_n(z+w)| \le \zeta(1+\|x\|).$$

In both cases (i) and (ii), therefore,

$$\sup_{s\le t}|K_s^n| \le \zeta(1+\|x\|)\frac{tn\Delta^2\xi_n}{v_{n\Delta}} \xrightarrow[n\to\infty]{} 0,$$

for all $t > 0$. Furthermore, we observe that N^n is a martingale w. r. t. the filtration $(\mathscr{F}_{sn\Delta})_{s\ge 0}$. Its predictable quadratic variation satisfies

$$\langle N^n, N^n\rangle_s = \frac{\Delta^2\xi_n^2}{v_{n\Delta}^2}|h_n(X_- + w)|^2\mathbb{1}_{\|w\|>1} \star \mathfrak{n}_{sn\Delta} \le \zeta(1+\|x\|)\frac{sn\Delta^3\xi_n^2}{v_{n\Delta}^2\eta_n^d} \xrightarrow[n\to\infty]{} 0.$$

Since $\lfloor sn\rfloor/n \to s$, $N^n_{\lfloor sn\rfloor/n} \to 0$ uniformly on $\{0 \le s \le t\}$ in probability for all $t > 0$.

In addition, we recall that $F(z, \{\|w\| > 1\} \le \zeta(1+\|z\|)$ under Assumption 2.1. Thus,

$$\sup_{s\le t} |K'^n_s| \le \zeta(1 + \|x\|)\xi_n \Delta S'^{n,\Delta,\eta_n}_t \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0$$

in case (i); and for all $t > 0$ in case (ii). Again, we observe that N'^n is a martingale w. r. t. the filtration $(\mathscr{F}_{sn\Delta})_{s\ge 0}$. Its predictable quadratic variation satisfies

$$\langle N'^n, N'^n\rangle_s = \frac{\Delta^2\xi_n^2}{v_{n\Delta}^2\eta_n^{2d}}|h_n(X_-)|^2\mathbb{1}_{\|w\|>1} \star \mathfrak{n}_{sn\Delta} \le \frac{\zeta(1 + \|x\|)\Delta^2\xi_n^2}{v_{n\Delta}\eta_n^d} S'^{n,\Delta,\eta_n}_s \xrightarrow[n\to\infty]{} 0.$$

Thus, $N'^n_{\lfloor sn\rfloor/n} \to 0$ uniformly on $\{0 \le s \le t\}$ in probability in case (i); and for all $t > 0$ in case (ii). ⋄

(d) Let $(M'^n_s)_{s\ge 0}$ and $(M''^n_s)_{s\ge 0}$ denote the $\mathscr{F}_{sn\Delta}$-martingales given by

$$M'^n_s := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\int_0^{sn\Delta} \varphi_\Delta(r)\nabla h_n(X_r)^\intercal\sigma(X_r)\mathrm{d}W_r,$$
$$M''^n_s := \frac{\xi_n}{v_{n\Delta}\eta_n^d}\varphi_\Delta(r)(h_n(X_- + w) - h_n(X_-))\mathbb{1}_{\|w\|\le 1} \star (\mathfrak{m} - \mathfrak{n})_{sn\Delta},$$

where $\varphi_\Delta(r) := \Delta - (r - \lfloor r/\Delta\rfloor\Delta)$. The predictable quadratic variation of M'^n satisfies

$$\langle M'^n, M'^n\rangle_s = \frac{\xi_n^2}{v_{n\Delta}^2\eta_n^{2d}}\int_0^{sn\Delta} \varphi_\Delta(r)^2\nabla h_n(X_r)^\intercal c(X_r)\nabla h_n(X_r)\mathrm{d}t$$
$$\le \frac{\zeta(1 + \|x\|^2)\Delta^2\xi_n^2}{v_{n\Delta}\eta_n^{d+2}} S'^{n,\Delta,\eta_n}_s.$$

As $\Delta\xi_n\eta_n^{-2} \to 0$ and $v_{n\Delta}\eta_n^d \to \infty$, $M'^n_s \to 0$ uniformly on $\{0 \le s \le t\}$ in probability in case (i); and for all $t > 0$ in case (ii).

In addition, the predictable quadratic variation of M''^n satisfies

$$\langle M''^n, M''^n\rangle_s = \frac{\xi_n^2}{v_{n\Delta}^2\eta_n^{2d}}\varphi_\Delta(r)^2(h_n(X_- + w) - h_n(X_-))^2\mathbb{1}_{\|w\|\le 1} \star \mathfrak{n}_{sn\Delta}$$
$$\le \frac{\Delta^2\xi_n^2}{v_{n\Delta}^2\eta_n^{2d}}\int_0^{sn\Delta}\mathrm{d}r\int_{\|w\|\le 1} F(X_r, \mathrm{d}w)(h_n(X_r + w) - h_n(X_r))^2.$$

Let $\zeta' > 1$ and $\kappa = 1 \wedge 2/(\beta + d)$ be as in step (b). By (5.13),

$$\left|\int_{\|w\|\le 1} F(z, \mathrm{d}w)(h_n(z + w) - h_n(z))^2\right|$$
$$\le \begin{cases} \zeta(1 + \|z\|)\eta_n^{-2}\int_{\|w\|\le 1}\bar{F}(\mathrm{d}w)\|w\|^2, & \text{for } \|z - x\| \le \zeta'\eta_n^\kappa, \\ \zeta(1 + \|z\|)\eta_n^{-\kappa\beta}\int_{\|w\|\le 1}\bar{F}(\mathrm{d}w)\|w\|^\beta, & \text{for } \zeta'\eta_n^\kappa < \|z - x\| \le 1 + \eta_n, \\ 0, & \text{else.}\end{cases}$$

Therefore,

$$\langle M''^n, M''^n \rangle_s \le \frac{\zeta(1+\|x\|)\Delta\xi_n}{v_{n\Delta}\eta_n^d}\left(\frac{\Delta\xi_n\eta_n^{\kappa d}}{\eta_n^{d+2}}S_s'^{n,\Delta,\zeta'\eta_n^\kappa} + \frac{\Delta\xi_n}{\eta_n^{d+\kappa\beta}}S_s'^{n,\Delta,1+\eta_n}\right).$$

Again since $\Delta\xi_n\eta_n^{-2-d(1-\kappa)} \to 0$, $M_s''^n \to 0$ uniformly on $\{0 \le s \le t\}$ in probability in case (i); and for all $t > 0$ in case (ii). □

5.2. Auxiliary martingale limit theorem

The theorem presented in this subsection serves as a preliminary result for the proof of our central limit theorem (Theorem 2.10 and Corollary 2.11). It is a non-standard limit theorem for a triangular, martingale array scheme.

Here, we work on the extension (2.11) of the probability space, L denotes the Mittag-Leffler process of order $0 < \delta \le 1$, and $W = (W^i)_{i\in I}$ denotes an I-dimensional standard Wiener process such that L, W and $\mathscr{F}$ are independent.

5.5 Theorem. *For $n \in \mathbb{N}^*$, let $(\mathscr{G}_s^n)_{s>0}$ be the filtration given by $\mathscr{G}_s^n := \mathscr{F}_{\lfloor sn\rfloor\Delta}$, and I be a finite index set. Moreover, let $h_n : E\times E \to \mathbb{R}^I$ be such that $\|h_n\|_\infty \to 0$ as $n\to\infty$. Grant Assumptions 2.2 and 2.4, and suppose that the process M^n given by*

$$M_s^n := \sum_{k=1}^{\lfloor sn\rfloor} h_n(X_{(k-1)\Delta}, \Delta_k^n X) \tag{5.14}$$

is a $\mathscr{G}_s^n$-martingale such that the predictable quadratic co-variation $\langle M^{ni}, M^{nj}\rangle$ is identically zero for every $i \ne j$ and all n large enough. If $(\langle M^{ni}, M^{ni}\rangle)_{i\in I}$ converges stably in law in $\mathcal{D}(\mathbb{R}^I)$ to $(\varsigma_i^2 L)_{i\in I}$, then

$$M^n \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} (\varsigma_i W_L^i)_{i\in I}.$$

Proof. Let $\delta = 1$. Then we have $L_s = s$. Therefore, the convergence of M^n to $(\varsigma_i^2 W^i)_{i\in I}$ follows directly from standard results (see section VIII.3c of Jacod and Shiryaev (2003)).

For the remainder, let $0 < \delta < 1$. We consider the processes L^{ni}, $\bar{L}^n$, K^n and N^n given by

$$L_s^{ni} := \langle M^{ni}, M^{ni}\rangle_s = \sum_{k=1}^{\lfloor sn\rfloor} \mathbb{E}^{X_{(k-1)\Delta}}\, h_n^i(X_{(k-1)\Delta}, \Delta_k^n X)^2,$$

$$\bar{L}_s^n := \sum_{i\in I} L_s^{ni}, \quad K_u^n := \inf\left\{s > 0 : \bar{L}_s^n > u\right\} \quad \text{and} \quad N_s^n := M_{K_s^n}^n.$$

We emphasise that $N^n(\bar{L}_s^n) = M_s^n + \Delta M_{K^n(\bar{L}_s^n)}^n$ holds for all s. As $\|\Delta M^n\| \le \|h_n\| \to 0$, it is sufficient to prove that we have the following stable convergence in law in $\mathcal{D}(\mathbb{R}\times\mathbb{R}^I)$:

$$(\bar{L}^n, N^n) \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left(\bar{\varsigma}^2 L, \left((\varsigma_i/\bar{\varsigma})W^i\right)_{i\in I}\right), \quad \text{where } \bar{\varsigma}^2 := \sum_{i\in I}\varsigma_i^2. \tag{5.15}$$

First, by the continuous mapping theorem, we obtain

$$\left(\bar{L}^n, \left(L^{ni}\right)_{i\in I}\right) \xrightarrow[n\to\infty]{\mathscr{L}-\text{st}} \left(\bar{\varsigma}^2 L, \left(\varsigma_i^2 L\right)_{i\in I}\right). \tag{5.16}$$

Second, we remark that K_u^n is a predictable $\mathscr{G}_s^n$-stopping time for all $u \geq 0$. Thus, N^n is a martingale w. r. t. the time-changed filtration $\mathscr{H}_s^n := \mathscr{G}_{K_s^n}^n$. Moreover, we observe that its predictable quadratic variation satisfies

$$\langle N^{ni}, N^{ni}\rangle_s = L_{K_s^n}^{ni}.$$

By (5.16), we have that $|L^{ni} - (\varsigma_i/\bar{\varsigma})^2 \bar{L}^n| \to 0$ uniformly on compacts in probability for all $i \in I$. We note that the (scaled) Mittag-Leffler process $\bar{\varsigma}^2 L$ is a. s. continuous. Its right-inverse K given by $K_u := \inf\{s > 0 : \bar{\varsigma}^2 L_s > u\}$ is a (deterministically time-changed) δ-stable Lévy process, hence, without fixed time of discontinuity. By (3.2) of Höpfner et al. (1990), therefore, $L_{K_s^n}^{ni} \to (\varsigma_i/\bar{\varsigma})^2 s$ in law for every $s \geq 0$; hence, in probability. By construction, we have that $\|\Delta N_s^n\|$ is bounded above by $\|h_n\|_\infty$. This bound converges to zero. By standard results (see above), consequently,

$$N^n \xrightarrow[n\to\infty]{\mathscr{L}-\text{st}} \left((\varsigma_i/\bar{\varsigma})W^i\right)_{i\in I}. \tag{5.17}$$

In analogy to the proof of Lemma 4.12 and Steps 6 and 7 on pp. 122–124 of Höpfner et al. (1990), we obtain that the pair $(\bar{L}^n, N^n)$ converges in law in $\mathcal{D}(\mathbb{R} \times \mathbb{R}^I)$ to $(\bar{\varsigma}^2 L, ((\varsigma_i/\bar{\varsigma})W^i)_{i\in I})$. Finally, the stable convergence in law and the independence from $\mathscr{F}$ follows in analogy to Lemma 4.13. □

5.3. Proof of Theorem 2.9

Throughout the remainder of section 5, we work under the law $\mathbb{P}^\pi$ for some inital probability π on E, and we denote $E_\oplus := \{x \in E : \mu'(x) > 0, F(x, E) > 0\}$.

We consider the processes $G^{n,\Delta,\eta}$ and $R^{n,\Delta,\eta}$ given by

$$G_s^{n,\Delta,\eta}(x, y) := \frac{1}{v_{n\Delta}} \sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta,x}(X_{(k-1)\Delta}) g_2^{\eta,y}(\Delta_k^n X), \tag{5.18}$$

$$R_s^{n,\Delta,\eta}(x) := \frac{\Delta}{v_{n\Delta}} \sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta,x}(X_{(k-1)\Delta}). \tag{5.19}$$

5.6 Lemma. *Grant Assumptions 2.1 to 2.3. Let $\eta_n = \eta_{1,n}$ be such that (2.7) holds, and let $x \in E_\oplus$.*

(i) If $n\Delta^2 \to 0$, then,

$$\text{the family } \left\{\mathscr{L}\left(R_1^{n,\Delta,\eta_n}(x) \mid \mathbb{P}^\pi\right) : n \in \mathbb{N}^*\right\} \text{ is tight.} \tag{5.20}$$

(ii) Grant Assumption 2.4 in addition. If $(n\Delta)^{1-\delta}\Delta \to 0$, then, (5.20) holds as well.

In both cases, each limit point of the family in (5.20) is the law $\mathscr{L}(\mu'(x)\tilde{L})$ for some positive random variable $\tilde{L}$.

Proof. Let $S_s^{t,\eta}(x) := v_t^{-1}\int_0^{st} g_2^{\eta,x}(X_r)\mathrm{d}r$. By Lemma 4.8, the family $\{\mathscr{L}(S_1^{n\Delta,\eta_n}(x) \mid \mathbb{P}^\pi) : n \in \mathbb{N}^*\}$ is tight; moreover, each of its limit points is the law $\mathscr{L}(\mu'(x)\tilde{L})$ for some random variable $\tilde{L} > 0$. In both cases (i) and (ii), since η_n is such that (2.7) holds, we have

$$\left|S_1^{n\Delta,\eta_n}(x) - R_1^{n,\Delta,\eta_n}(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}^\pi} 0$$

by Proposition 5.2. Consequently, the family $\{\mathscr{L}(R_1^{n,\Delta,\eta}(x) \mid \mathbb{P}^\pi) : n \in \mathbb{N}^*\}$ is tight; moreover, each of its limit points is a limit point of the family $\{\mathscr{L}(S_1^{t,\eta}(x) \mid \mathbb{P}^\pi) : t > 0\}$, hence, the law $\mathscr{L}(\mu'(x)\tilde{L})$ for some random variable $\tilde{L} > 0$. □

5.7 Lemma. *Grant Assumptions 2.1 and 2.3. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that $\eta_{1,n} \to 0$, $\eta_{2,n} \to 0$ and $\Delta\eta_{2,n}^{-2\vee(\beta+d)} \to 0$. Moreover, let $(x,y) \in E_\oplus \times E^*$, and let g be a $\mathcal{C}^2$-function with compact support. Then*

$$\lim_{n\to\infty} \sup_{z\in B_{\eta_{1,n}}(x)} \left|\frac{1}{\Delta}\,\mathbb{E}^z\, g^{\eta_n,y}(\Delta_1^n X) - f(x,y)\int g(w)\mathrm{d}w\right| = 0. \tag{5.21}$$

Proof. First, by Proposition 5.1 – where we choose m large enough – we have

$$\lim_{n\to\infty} \sup_{z\in B_{\eta_{1,n}}(x)} \left|\Delta^{-1}\,\mathbb{E}^z\, g^{\eta,x}(\Delta_1^n X) - Fg^{\eta,y}(z)\right| = 0.$$

Second, under Assumption 2.3, $f \in \mathcal{C}^\alpha_{\mathrm{loc}}(E\times E^*)$ for some $\alpha > 0$. Therefore,

$$\lim_{n\to\infty} \sup_{z\in B_{\eta_{1,n}}(x)} \left|Fg^{\eta_n,y}(z) - Fg^{\eta_n,y}(x)\right| \le \lim_{n\to\infty} \zeta\eta_{1,n}^{\alpha\wedge 1} = 0.$$

Third, by Lebesgue's differentiation theorem, we observe

$$\lim_{n\to\infty} \left|Fg^{\eta_n,y}(x) - f(x,y)\int g(w)\mathrm{d}w\right| = 0.$$

□

5.8 Lemma. *Grant Assumptions 2.1 to 2.3. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.7) holds. Moreover, let $(x,y) \in E_\oplus \times E^*$. Then, in both cases as in Lemma 5.6,*

$$\text{the family } \left\{\mathscr{L}\left(G_1^{n,\Delta,\eta_n}(x,y), R_1^{n,\Delta,\eta_n}(x) \mid \mathbb{P}^\pi\right) : n \in \mathbb{N}^*\right\} \text{ is tight.} \tag{5.22}$$

Moreover, each limit point of the family in (5.22) is the law $\mathscr{L}(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L})$ for some positive random variable $\tilde{L}$.

Proof. We note that $G_s^{n,\Delta,\eta}(x,y) = f(x,y)R_s^{n,\Delta,\eta}(x) + H_s^{n,\Delta,\eta}(x,y) + M_s^{n,\Delta,\eta}(x,y)$, where

$$H_s^{n,\Delta,\eta}(x,y) = \frac{1}{v_{n\Delta}} \sum_{k=1}^{\lfloor sn \rfloor} g_1^{\eta,x}(X_{(k-1)\Delta}) \left(\mathbb{E}^{X_{(k-1)\Delta}}[g_2^{\eta,x}(\Delta_1^n X)] - \Delta f(x,y) \right), \tag{5.23}$$

$$M_s^{n,\Delta,\eta}(x,y) = \frac{1}{v_{n\Delta}} \sum_{k=1}^{\lfloor sn \rfloor} g_1^{\eta,x}(X_{(k-1)\Delta}) \left(g_2^{\eta,y}(\Delta_k^n X) - \mathbb{E}^{X_{(k-1)\Delta}}[g_2^{\eta,x}(\Delta_1^n X)] \right). \tag{5.24}$$

By Lemma 5.6, it is sufficient to prove that $H_1^{n,\Delta,\eta_n}(x,y)$ and $M_1^{n,\Delta,\eta_n}(x,y)$ converge to zero in probability as $n \to \infty$.

(H) We observe

$$\left| H_1^{n,\Delta,\eta}(x,y) \right| \leq \sup_{z \in B_{\eta_1}(x)} \left| \Delta^{-1}\, \mathbb{E}^z[g_2^{\eta,x}(\Delta_1^n X)] - f(x,y) \right| v_{n\Delta}^{-1} \sum_{k=1}^{n} \Delta h^{\eta,x}(X_{(k-1)\Delta}), \tag{5.25}$$

where h is a $\mathcal{C}^2$-function dominating $|g_1|$. The sequence $(v_{n\Delta}^{-1} \sum_{k=1}^n \Delta h^{\eta_n,x}(X_{(k-1)\Delta}))_{n \in \mathbb{N}^*}$ is tight in analogy to Lemma 5.6. As $\sup_{z \in B_{\eta_{1,n}}(x)} |\Delta^{-1}\, \mathbb{E}^z[g_2^{\eta_n,x}(\Delta_1^n X)] - f(x,y)| \to 0$ by Lemma 5.7, we have $H_1^{n,\Delta,\eta_n}(x,y) \to 0$ in law, hence, in probability. ⋄

(M) We observe that $M^{n,\Delta,\eta}$ is an $\mathscr{F}_{\lfloor sn \rfloor \Delta}$-martingale. We note $\sup_{s \leq 1} \|\Delta M_s^{n,\Delta,\eta_n}\| \leq (v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d)^{-1} \|g_1\|_\infty \|g_2\|_\infty \to 0$ by (2.7). By Theorem VIII.2.4 of Jacod and Shiryaev (2003), thus, it is sufficient to show that the predictable quadratic variation of M^{n,Δ,η_n} at time one, denoted $\langle M^{n,\Delta,\eta_n}, M^{n,\Delta,\eta_n} \rangle_1$, converges to zero in probability.

We observe

$$\left\langle M^{n,\Delta,\eta}, M^{n,\Delta,\eta} \right\rangle_1 \leq \frac{\|g_1\|_\infty}{v_{n\Delta}\eta_1^d \eta_2^d} \sup_{z \in B_{\eta_1}(x)} \left| \frac{\eta_2^d}{\Delta}\, \mathbb{E}^z\, g_2^{\eta,y}(\Delta_1^n X)^2 \right| v_{n\Delta}^{-1} \sum_{k=1}^{n} \Delta h^{\eta,x}(X_{(k-1)\Delta}).$$

By Lemma 5.7, $\sup_{z \in B_{\eta_{1,n}}(x)} |\Delta^{-1}\, \mathbb{E}^z\, \eta_{2,n}^d g_2^{\eta_n,x}(\Delta_1^n X)^2| \to f(x,y) \int g_1(w)^2 \mathrm{d}w$. In analogy to step (H), since $v_{n\Delta}\eta_{1.n}^d \eta_{2,n}^d \to \infty$, we have $\langle M^{n,\Delta,\eta_n}, M^{n,\Delta,\eta_n} \rangle_1 \to 0$ in law, hence, in probability. □

Proof (of Theorem 2.9). We recall the results from Lemma 5.8. Let $\tilde{L} > 0$ be a random variable such that the law $\mathscr{L}(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L})$ is a limit point of the family in (5.22), and let $(n_k)_{k \in \mathbb{N}^*}$ be a sequence such that

$$\left(G_1^{n_k,\Delta,\eta_{n_k}}(x,y), R_1^{n_k,\Delta,\eta_{n_k}}(x) \right) \xrightarrow[k \to \infty]{\mathscr{L}} \left(f(x,y)\mu'(x)\tilde{L}, \mu'(x)\tilde{L} \right).$$

Since $\mu'(x) > 0$, by the continuous mapping theorem, we conclude

$$\hat{f}_{n_k}^{\Delta,\eta_{n_k}}(x,y) = \frac{G_1^{n_k,\Delta,\eta_{n_k}}(x,y)}{R_1^{n_k,\Delta,\eta_{n_k}}(x)} \xrightarrow[k \to \infty]{\mathscr{L}} f(x,y).$$

As this limit is unique and independent of the particular limit point of the family in (5.22), we have that $\hat{f}_n^{\Delta,\eta_n}(x,y)$ converges to $f(x,y)$ in law, hence, in probability. □

5.4. Proofs of Theorem 2.10 and Corollary 2.11

Throughout this subsection, we work on the extension (2.11) of the probability space, L denotes the Mittag-Leffler process of order $0 < \delta \leq 1$, and $W = (W^i)_{i\in I}$ denotes an I-dimensional standard Wiener process such that L, W and $\mathscr{F}$ are independent.

We consider the processes $G^{n,\Delta,\eta}$ and $R^{n,\Delta,\eta}$ given by (5.18) and (5.19), and the processes $U^{n,\Delta,\eta}$ and $R'^{n,\Delta,\eta}$ given by

$$U^{n,\Delta,\eta}_s(x,y) := \sqrt{v_{n\Delta}\eta_1^d\eta_2^d}\left(G^{n,\Delta,\eta}_s(x,y) - \frac{\mu(g_1^{\eta,x}Fg_2^{\eta,y})}{\mu(g_1^{\eta,x})}R^{n,\Delta,\eta}_s(x)\right) \tag{5.26}$$

$$R'^{n,\Delta,\eta}_s(x) := \frac{\Delta}{v_{n\Delta}}\sum_{k=1}^{\lfloor sn\rfloor}\eta_1^d g_1^{\eta,x}(X_{(k-1)\Delta})^2. \tag{5.27}$$

We recall that, under Darling–Kac's condition, we have Théorème 3 of Touati (1987) at hand (see Proposition 4.10). First, we obtain an extension of Lemma 5.6.

5.9 Lemma. *Grant Assumptions 2.1 to 2.4. Let $\eta_n = \eta_{1,n}$ be such that (2.7) and (2.9a) hold, and let $(x_i)_{i\in I}$ be a family of pairwise distinct points in $E_\oplus$. If $(n\Delta)^{1-\delta}\Delta \to 0$, then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law in $\mathcal{D}(\mathbb{R}^{2I})$:*

$$\left(R^{n,\Delta,\eta_n}(x_i), R'^{n,\Delta,\eta_n}(x_i)\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left(\mu'(x_i)L, \mu'(x_i)\int g_2(w)^2\mathrm{d}wL\right)_{i\in I}. \tag{5.28}$$

Proof. Let $S^{t,\eta}_s(x) := v_t^{-1}\int_0^{st} g_1^{\eta,x}(X_r)\mathrm{d}r$ and $S'^{t,\eta}_s(x) := v_t^{-1}\int_0^{st}\eta^d g_1^{\eta,x}(X_r)^2\mathrm{d}r$. We note that $\mu(g_1^{\eta_n,x}) \to \mu'(x)$ and $\mu(\eta_n^d(g_1^{\eta_n,x})^2) \to \mu'(x)\int g_1(w)^2\mathrm{d}w$ for all x. By Theorem 4.2 and Proposition 4.10, we deduce – in analogy to Corollary 4.14 – that

$$\left(S^{n\Delta,\eta_n}(x_i), S'^{n\Delta,\eta_n}(x_i)\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{t\to\infty}{\Longrightarrow}} \left(\mu'(x_i)L, \mu'(x_i)\int g_2(w)^2\mathrm{d}wL\right)_{i\in I}.$$

For every x, moreover, we deduce from Proposition 5.2 that

$$\left|R^{n,\Delta,\eta_n}(x) - S^{n\Delta,\eta_n}(x)\right| \underset{n\to\infty}{\overset{\text{ucp}}{\Longrightarrow}} 0 \quad\text{and}\quad \left|R'^{n,\Delta,\eta_n}(x) - S'^{n\Delta,\eta_n}(x)\right| \underset{n\to\infty}{\overset{\text{ucp}}{\Longrightarrow}} 0.$$

Consequently, we obtain (5.28). □

In view of Theorem 5.5, we obtain the following preliminary result.

5.10 Lemma. *Grant Assumptions 2.1 to 2.5. Let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.7) and (2.9) hold, and let $(x_i, y_i)_{i\in I}$ be a finite family of pairwise distinct points in $E_\oplus \times E^*$. If $(n\Delta)^{1-\delta}\Delta \to 0$, then, under any law $\mathbb{P}^\pi$, we have the following stable convergence in law in $\mathcal{D}(\mathbb{R}^I)$:*

$$\left(R^{n,\Delta,\eta_n}(x_i), U^{n,\Delta,\eta_n}(x_i,y_i)\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left(\mu'(x_i)L, \sigma(x_i,y_i)\mu'(x_i)W^i_L\right)_{i\in I}, \tag{5.29}$$

where $\sigma(x,y)^2$ is given by (2.12).

Proof. Let $(\mathscr{G}_s^n)_{s\geq 0}$ be given by $\mathscr{G}_s^n = \mathscr{F}_{\lfloor sn\rfloor\Delta}$, and let the process $M^{n,\Delta,\eta}$ be given by

$$M_s^{n,\Delta,\eta}(x,y) := \sqrt{\frac{\eta_1^d\eta_2^d}{v_{n\Delta}}}\sum_{k=1}^{\lfloor sn\rfloor} g_1^{\eta,x}(X_{(k-1)\Delta})\left(g_2^{\eta,y}(\Delta_k^n X) - \mathbb{E}^{X_{(k-1)\Delta}}\, g_2^{\eta,y}(\Delta_1^n X)\right).$$

We note that $M^{n,\Delta,\eta}$ is a $\mathscr{G}_s^n$-martingale of the form (5.14). The proof is divided into four steps: First, we prove

$$\left|U^{n,\Delta,\eta_n}(x,y) - M^{n,\Delta,\eta_n}(x,y)\right| \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} 0. \tag{5.30}$$

Second, we show that the predictable quadratic variation of $M^{n,\Delta,\eta}(x,y)$ satisfies

$$\left(\left\langle M^{n,\Delta,\eta_n}(x_i,y_i), M^{n,\Delta,\eta_n}(x_i,y_i)\right\rangle\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left([\sigma(x_i,y_i)\mu'(x_i)]^2 L\right)_{i\in I} \tag{5.31}$$

in $\mathcal{D}(\mathbb{R}^I)$. Third, we show that $\langle M^{n,\Delta,\eta_n}(x_i,y_i), M^{n,\Delta,\eta_n}(x_j,y_j)\rangle$ vanishes for all n large enough if $i\neq j$. Last, we argue

$$\left(R^{n,\Delta,\eta_n}(x_i), \left\langle M^{n,\Delta,\eta_n}(x_i,y_i), M^{n,\Delta,\eta_n}(x_i,y_i)\right\rangle\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left(\mu'(x_i)L, [\sigma(x_i,y_i)\mu'(x_i)]^2 L\right)_{i\in I}$$

in $\mathcal{D}(\mathbb{R}^{2I})$. By Theorem 5.5 and (3.5) of Höpfner et al. (1990), we then have (5.29).

(i) We note $U_s^{n,\Delta,\eta}(x,y) - M_s^{n,\Delta,\eta}(x,y) = H_s^{n,\Delta,\eta}(x,y) + H_s'^{n,\Delta,\eta}(x,y)$ with

$$H_s^{n,\Delta,\eta}(x,y) := \sqrt{v_{n\Delta}\eta_1^d\eta_2^d}\,\frac{\Delta}{v_{n\Delta}}\sum_{k=1}^{\lfloor sn\rfloor} g_1(X_{(k-1)\Delta})\left(Fg_2^{\eta,y}(X_{(k-1)\Delta}) - \frac{g_1^{\eta,x}Fg_2^{\eta,y}}{\mu(g_1^{\eta,x})}\right),$$

$$|H_s'^{n,\Delta,\eta}(x,y)| \leq \sqrt{v_{n\Delta}\eta_1^d\eta_2^d}\sup_{z\in B_{\eta_1}(x)}\left|\frac{1}{\Delta}\,\mathbb{E}^z\, g_2^{\eta,y}(\Delta_1^n X) - Fg_2^{\eta,y}(z)\right| R_s''^{n,\Delta,\eta}(x),$$

where $R_s''^{n,\Delta,\eta}(x) = \Delta v_{n\Delta}^{-1}\sum_{k=1}^{\lfloor sn\rfloor} h^{\eta,x}(X_{(k-1)\Delta})$ for some $\mathcal{C}^2$-function h, dominating $|g_1|$. Under Assumption 2.5, $Fg_2^{\eta,y}$ is twice continuously differentiable. Since (2.9) holds, by Proposition 5.2 and step (i) in the proof of Lemma 4.9, $H^{n,\Delta,\eta_n}(x,y) \Rightarrow 0$ in ucp as $n\to\infty$. By Proposition 5.1 – where we choose m large enough – we have

$$\sup_{z\in B_{\eta_1}(x)}\left|\frac{1}{\Delta}\,\mathbb{E}^z\, g_2^{\eta,y}(\Delta_1^n X) - Fg_2^{\eta,y}(z)\right| \leq \zeta\left(\sqrt{\Delta} + \Delta\eta_2^{-2\vee(\beta+d)}\right)$$

since (2.9a) holds. Since, moreover, (2.9) holds, therefore,

$$\sqrt{v_{n\Delta}\eta_{1,n}^d\eta_{2,n}^d}\sup_{z\in B_{\eta_{1,n}}(x)}\left|\frac{1}{\Delta}\,\mathbb{E}^z\, g_2^{\eta_n,y}(\Delta_1^n X) - Fg_2^{\eta_n,y}(z)\right| \underset{n\to\infty}{\longrightarrow} 0. \tag{5.32}$$

In analogy to Lemma 5.9, $R''^{n,\Delta,\eta_n}(x)$ converges stably in law. Thus, $|H_s'^{n,\Delta,\eta}(x,y)| \Rightarrow 0$ in ucp as $n\to\infty$. Consequently, (5.30) holds. ⋄

(ii) We note $\langle M^{n,\Delta,\eta}(x,y), M^{n,\Delta,\eta}(x,y)\rangle_s = K_s^{n,\Delta,\eta}(x,y) - K_s'^{n,\Delta,\eta}(x,y)$, where

$$K_s^{n,\Delta,\eta}(x,y) = \frac{\eta_1^d \eta_2^d}{v_{n\Delta}} \sum_{k=1}^{\lfloor sn \rfloor} g_1^{\eta,x}(X_{(k-1)\Delta})^2 \left(\mathbb{E}^{X_{(k-1)\Delta}}\, g_2^{\eta,y}(\Delta_1^n X)^2\right),$$

and

$$|K_s'^{n,\Delta,\eta}(x,y)| \leq \sup_{z\in B_{\eta_1}(x)} \left| \frac{1}{\Delta^2} \left(\mathbb{E}^{X_{(k-1)\Delta}}\, g_2^{\eta,y}(\Delta_1^n X)\right)^2 \right| \Delta \eta_2^d R_s'^{n,\Delta,\eta}(x).$$

By Lemma 5.7 and the continuous mapping theorem,

$$\sup_{z\in B_{\eta_{1,n}}(x)} \left| \frac{1}{\Delta^2} \left(\mathbb{E}^z\, g_2^{\eta_n,y}(\Delta_1^n X)\right)^2 \right| \xrightarrow[n\to\infty]{} f(x,y)^2.$$

By Lemma 5.9, $R_s'^{n,\Delta,\eta_n}(x)$ converges stably in law. Since $\Delta\eta_{2,n}^d \to 0$, we observe that $|K_s'^{n,\Delta,\eta_n}(x,y)|$ converges to zero uniformly on compacts in probability as $n\to\infty$.

Again by Lemma 5.7,

$$\sup_{z\in B_{\eta_{1,n}}(x)} \left| \frac{\eta_{2,n}^d}{\Delta}\, \mathbb{E}^{X_{(k-1)\Delta}}\, g_2^{\eta_n,y}(\Delta_1^n X)^2 - f(x,y)\int g_2(w)^2 \mathrm{d}w \right| \xrightarrow[n\to\infty]{} 0.$$

In analogy to $K'^{n,\Delta,\eta}(x,y)$, therefore,

$$\left| K^{n,\Delta,\eta_n}(x,y) - f(x,y)\int g_1(w)^2 \mathrm{d}w R'^{n,\Delta,\eta_n}(x) \right| \overset{\text{ucp}}{\underset{n\to\infty}{\Longrightarrow}} 0. \tag{5.33}$$

By Lemma 5.9, consequently,

$$\left(K^{n,\Delta,\eta_n}(x_i,y_i)\right)_{i\in I} \overset{\mathscr{L}-\text{st}}{\underset{n\to\infty}{\Longrightarrow}} \left(f(x_i,y_i)\int g_1(w)^2 \mathrm{d}w \mu'(x_i) \int g_2(z)^2 \mathrm{d}z L\right)_{i\in I};$$

hence, (5.31) holds. ⋄

(iii) Let $i,j\in I$. We note that for all n large enough such that $\eta_{1,n},\eta_{2,n}$ are small enough, we have $g_1^{\eta_n,x_i} g_1^{\eta_n,x_j} \equiv 0$ whenever $x_i \neq x_j$, and $g_2^{\eta_n,y_i} g_2^{\eta_n,y_j} \equiv 0$ whenever $y_i \neq y_j$. For all ω and n large enough, thus, $\langle M^{n,\Delta,\eta_n}(x_i,y_i), M^{n,\Delta,\eta_n}(x_j,y_j)\rangle_s \equiv 0$ if $i\neq j$. ⋄

(iv) By Lemma 5.9 and (5.33), we obtain the joint convergence of $(R^{n,\Delta,\eta_n}(x_i))_{i\in I}$ and $\langle M^{n,\Delta,\eta_n}(x_i,y_i), M^{n,\Delta,\eta_n}(x_i,y_i)\rangle)_{i\in I}$ to the required limit. □

Proof (of Theorem 2.10). For every n, and $(x,y)\in E_\oplus \times E^*$, we have

$$\sqrt{v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d}\left(\hat{f}_n^{\Delta,\eta_n}(x,y) - \bar{f}^{\eta_n}(x,y)\right) = \frac{U_1^{n,\Delta,\eta_n}(x,y)}{R_1^{n,\Delta,\eta_n}(x)},$$

where $\bar{f}^\eta(x,y) := \mu(g_1^{\eta,x} F g_2^{\eta,y})/\mu(g_1^{\eta,x})$. Since L and W are independent, $V(x_i,y_i) := L_1^{-1/2} W_{L_1}^i$ defines an I-dimensional standard Gaussian random vector such that L, V and $\mathscr{F}$ are independent. By the continuous mapping theorem and Lemma 5.10, consequently,

$$\sqrt{v_{n\Delta}\eta_{1,n}^d \eta_{2,n}^d}\left(\hat{f}_n^{\Delta,\eta_n}(x_i,y_i) - \bar{f}^{\eta_n}(x_i,y_i)\right)_{i\in I} \xrightarrow[n\to\infty]{\mathscr{L}-\text{st}} \left(\sigma(x_i,y_i)V(x_i,y_i)L_1^{-1/2}\right)_{i\in I},$$

where $\sigma(x,y)^2$ is given by (2.12).

In addition, let $\eta_n = (\eta_{1,n}, \eta_{2,n})$ be such that (2.8) holds as well. It remains to prove $(v_{n\Delta}\eta_{1,n}^d\eta_{2,n}^d)^{1/2}(\bar{f}^{\eta_n}(x,y) - f(x,y)) \to \gamma(x,y)$. This, however, follows in analogy to the proof of Theorem 3.7. □

Proof (of Corollary 2.11). In analogy to the proof of Theorem 2.10, by Lemma 5.10 it remains to show that $(v_{n\Delta}\eta_{1,n}^d\eta_{2,n}^d)^{1/2}\hat{\gamma}_n^{\eta_n}(x,y)$ is a consistent estimator for $\gamma(x,y)$. This, however, follows in analogy to the proof of Corollary 3.8. □

A. On the auxiliary Markov chains *Z* and *Z'*

In this appendix, we derive an explicit representation for the transition kernel Φ of the auxiliary process Z', and (in-)equalities for expectations of the form $\mathbb{E}^x(\int_0^{T_1} h(X_s)\mathrm{d}s)^k$. In addition, we derive representations for the stationary probability measures ψ and φ of the processes Z and Z'.

We invoke technical results on resolvents of semi-groups. The *resolvent* $(R_\lambda)_{\lambda>0}$ of a semi-group $(P_t)_{t\geq 0}$ is given by $R_\lambda := \int_0^\infty \exp(-\lambda t)P_t\mathrm{d}t$. For bounded measurable functions h, the generalised resolvent kernel R_h is given by

$$R_h(x,A) := \mathbb{E}^x \int_0^\infty e^{-\int_0^t h(X_s)\mathrm{d}s}\mathbb{1}_A(X_t)\mathrm{d}t \qquad \forall x \in E, A \in \mathscr{E}.$$

These kernels were first introduced by Neveu (1972). For a comprehensive interpretation, we refer to section 4 of Down, Meyn, and Tweedie (1995).

A.1 Lemma. *Let $(R_\lambda)_{\lambda>0}$ be the resolvent of X, and let $(R^*_\lambda)_{\lambda>0}$ be given by*

$$R^*_\lambda := R_\lambda \sum_{k=0}^\infty \Big((\mathrm{I}_q - \mathrm{I}_q\bar{\Pi})R_\lambda\Big)^k, \quad \text{where} \quad \mathrm{I}_q h(x) := q(x)h(x). \tag{A.1}$$

Then $(R^*_\lambda)_{\lambda>0}$ is the resolvent of a positive contraction semi-group. For its corresponding process X^*, we have that the laws of $X^*\mathbb{1}_{[\![0,T_1[\![}$ and $X\mathbb{1}_{[\![0,T_1[\![}$ are equal.

Proof. Since $\mathrm{I}_q\bar{\Pi}$ is a bounded kernel, $(R^*_\lambda)_{\lambda>0}$ is the resolvent of a positive contraction semi-group by Theorem 4.2 of Bass (1979). It follows from Sawyer (1970) and Chapter 6 of Bass (1979) that, for the process X^* (corresponding to $(R^*_\lambda)_{\lambda>0}$), we have that the laws of $X^*\mathbb{1}_{[\![0,T_1[\![}$ and $X\mathbb{1}_{[\![0,T_1[\![}$ are equal. □

A.2 Lemma. *Let h be a measurable function on E. Then*

$$\mathbb{E}^x\, h(Z'_1) = R^*_q I_q h(x) \quad \text{and} \quad \mathbb{E}^x \int_0^{T_1} h(X_s)\mathrm{d}s = R^*_q h(x), \tag{A.2}$$

where R^*_q denotes the generalised resolvent kernel associated with the modified resolvent $(R^*_\lambda)_{\lambda>0}$ and the function q. For every $\lambda_q \geq \|q\|_\infty$, we have $R^*_q = \sum_{k=0}^\infty R^*_{\lambda_q}(\mathrm{I}_{\lambda_q - q}R^*_{\lambda_q})^k$.

Proof. We recall that the laws of $X^*\mathbb{1}_{[\![0,T_1[\![}$ and $X\mathbb{1}_{[\![0,T_1[\![}$ are equal. The expectation of $h(Z_1')$ under $\mathbb{P}^x$, therefore, coincides with the expectation of $h(X^*)$ sampled at an independent killing time according to the multiplicative functional $\exp(-\int_0^\cdot q(X_s^*)\mathrm{d}s)$. In formulas, we have

$$\mathbb{E}^x\, h(Z_1') = \mathbb{E}^x \int_0^\infty e^{-\int_0^t q(X_s^*)\mathrm{d}s} q(X_t^*)h(X_t^*)\mathrm{d}t.$$

By eq. (19) of Down et al. (1995), hence, $\mathbb{E}^x\, h(Z_1') = R_q^*\mathrm{I}_q h(x)$, where R_q^* denotes the generalised resolvent kernel associated with the modified resolvent $(R_\lambda^*)_{\lambda>0}$. By Chapter 7 of Neveu (1972), $R_q^* = \sum_{k=0}^\infty R_{\lambda_q}^*(\mathrm{I}_{\lambda_q-q}R_{\lambda_q}^*)^k$ holds for every $\lambda_q \geq \|q\|_\infty$.

Similarly, we observe

$$\mathbb{E}^x \int_0^{T_1} h(X_s)\mathrm{d}s = \mathbb{E}^x \int_0^\infty e^{-\int_0^t q(X_u^*)\mathrm{d}u} q(X_t^*) \int_0^t h(X_s^*)\mathrm{d}s\mathrm{d}t. \tag{A.3}$$

By Fubini's theorem (cf., eq. (20) of Down et al. (1995)), consequently,

$$\mathbb{E}^x \int_0^{T_1} h(X_s)\mathrm{d}s = \mathbb{E}^x \int_0^\infty e^{-\int_0^t q(X_s^*)\mathrm{d}s} h(X_t^*)\mathrm{d}t = R_q^* h(x).$$

□

Remark. It is immediate from Lemma 4.3 that $\Phi = \bar{\Pi}R_q^*\mathrm{I}_q$.

We obtain two corollaries:

A.3 Corollary. *Let $h_1, \ldots, h_k$ be measurable functions on E. Then*

$$\mathbb{E}^x \prod_{j=1}^k \int_0^{T_1} h_j(X_s)\mathrm{d}s = \sum_{j=1}^k \mathbb{E}^x \int_0^\infty e^{-\int_0^t q(X_u^*)\mathrm{d}u} h_j(X_t^*) \prod_{l\neq j} \int_0^t h_l(X_s^*)\mathrm{d}s\mathrm{d}t. \tag{A.4}$$

Proof. In analogy to (A.3), we observe

$$\mathbb{E}^x \prod_{j=1}^k \int_0^{T_1} h_j(X_s)\mathrm{d}s = \mathbb{E}^x \int_0^\infty e^{-\int_0^t q(X_u^*)\mathrm{d}u} q(X_t^*) \prod_{j=1}^k \int_0^t h_j(X_s^*)\mathrm{d}s\mathrm{d}t.$$

By the Leibniz rule, moreover,

$$\prod_{j=1}^k \int_0^t h_j(X_s^*)\mathrm{d}s = \sum_{j=1}^k \int_0^t h_j(X_s^*) \prod_{l\neq j} \int_0^s h_l(X_r^*)\mathrm{d}r\mathrm{d}s.$$

By Fubini's theorem, therefore, we have (A.4). □

A.4 Corollary. *Let h be a bounded measurable function on E. For all $k \in \mathbb{N}^*$, if $\inf_{x\in\mathrm{supp}(h)} q(x) > 0$, then*

$$\mathbb{E}^x \left(\int_0^{T_1} h(X_s)\mathrm{d}s \right)^k \leq \frac{k!\|h\|_\infty^{k-1}}{(\inf_{x\in\mathrm{supp}(h)} q(x))^{k-1}} R_q^*|h|(x). \tag{A.5}$$

Proof (by induction). By Lemma A.2, we immediately have (A.5) for $k = 1$. We assume that (A.5) holds for some $k \in \mathbb{N}^*$. Then we deduce from Corollary A.3 and $|h| \leq q\|h\|_\infty / (\inf_{x \in \mathrm{supp}(h)} q(x))$ that (A.5) holds for $k + 1$. □

A.5 Lemma. $\mu \mathrm{I}_q \bar{\Pi} R_q^* = \mu$.

Proof. By Theorem 4.2 of Bass (1979) and Section 7 of Neveu (1972), we have

$$(\mathrm{I}_q \bar{\Pi} - (\mathrm{I} - R_1^{-1})) R_q^* = \mathrm{I},$$

where the formal inverse of R_1 is defined by $R_1^{-1} := \sum_{k=0}^{\infty} (\mathrm{I} - R_1)^k$. Since μ is invariant w. r. t. $(P_t)_{t \geq 0}$, we also have $\mu R_1 = \mu$ and $\mu = \mu R_1^{-1}$. Hence, $\mu \mathrm{I}_q \bar{\Pi} = \mu(\mathrm{I}_q \bar{\Pi} - (\mathrm{I} - R_1^{-1}))$. Therefore, $\mu \mathrm{I}_q \bar{\Pi} R_q^* = \mu$. □

A.6 Corollary. *The measures $\varphi := (\mu(q))^{-1} \mu \mathrm{I}_q$ and $\psi := \varphi \Psi$ are the invariant probability measures w. r. t. Φ and Ψ.*

Proof. Since $\Phi = \bar{\Pi} R_q^* \mathrm{I}_q$, we observe $\mu \mathrm{I}_q \Phi = \mu \mathrm{I}_q$. By (4.10), $\varphi \Psi^{k+1} = \varphi \Phi^k \Psi = \varphi \Psi$ in addition. □

Acknowledgment

I am grateful to Jean Jacod for introducing me to this topic, illuminating discussions, and helpful comments on earlier versions of this manuscript.

References

Azéma, J., Kaplan-Duflo, M., and Revuz, D. (1967) Measure invariante sur les classes recurrents des processus de Markov. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 8:157–181.

Bass, R. F. (1979) Adding and subtracting jumps from Markov processes. *Trans. Amer. Math. Soc.*, 255:363–376.

Benveniste, A. and Jacod, J. (1973) Systèmes de Lévy des processus de Markov. *Invent. Math.*, 21:183–198.

Comte, F. and Genon-Catalot, V. (2011) Estimation for Lévy processes from high frequency data within a long time interval. *Ann. Statist.*, 39:803–837.

Darling, D. A. and Kac, M. (1957) On occupation times for Markoff processes. *Trans. Amer. Math. Soc.*, 84:444–458.

Down, D., Meyn, S. P., and Tweedie, R. L. (1995) Exponential and uniform ergodicity of Markov processes. *Ann. Appl. Probab.*, 23:1671–1691.

Fan, J. and Yim, T. H. (2004) A crossvalidation method for estimating conditional densities. *Biometrika*, 91:819–834.

Figueroa-López, J. E. (2011) Sieve-based confidence intervals and bands for Lévy densities. *Bernoulli*, 17:643–670.

Getoor, R. K. (1975) *Markov Processes: Ray Processes and Right Processes*, *Lecture Notes in Mathematics*, vol. 440. Springer, Berlin.

Greenwood, P. E. and Wefelmeyer, W. (1994) Nonparametric estimators for Markov step processes. *Stochastic Process. Appl.*, 52:1–16.

Gugushvili, S., Klaassen, C. A. J., and Spreij, P. (eds.) (2010) *Special Issue: Statistical Inference for Lévy Processes with Applications to Finance*, *Stat. Neerl.*, vol. 64. Pp. 255–366.

Hall, P., Racine, J., and Li, Q. (2004) Cross-validation and the estimation of conditional probability densities. *J. Amer. Statist. Assoc.*, 99:1015–1026.

Höpfner, R. (1993) Asymptotic inference for Markov step processes: Observation up to a random time. *Stochastic Process. Appl.*, 48:295–310.

Höpfner, R., Jacod, J., and Ladelli, L. (1990) Local asymptotic normality and mixed normality for Markov statistical models. *Probab. Theory Related Fields*, 86:105–129.

Höpfner, R. and Löcherbach, E. (2003) Limit theorems for null recurrent Markov processes. *Mem. Amer. Math. Soc.*, 161:vi+92.

Jacod, J. and Protter, P. (2012) *Discretization of Processes.* Springer, 2012 edn.

Jacod, J. and Shiryaev, A. N. (2003) *Limit Theorems for Stochastic Processes.* Springer, Berlin. 2nd edition.

Karlsen, H. A. and Tjøstheim, D. (2001) Nonparametric estimation in null recurrent time series. *Ann. Statist.*, 29:372–416.

Löcherbach, E. and Loukianova, D. (2008) On Nummelin splitting for continuous time Harris recurrent Markov processes and applications to kernel estimation for multidimensional diffusions. *Stochastic Process. Appl.*, 118:1301–1321.

Meyn, S. P. and Tweedie, R. L. (1993) *Markov Chains and Stochastic Stability.* Springer, London. Online edition, 2005.

Neumann, M. H. and Reiß, M. (2009) Nonparametric estimation for Lévy processes from low-frequency observations. *Bernoulli*, 15:223–248.

Neveu, J. (1972) Potentiel markovien récurrent des chaînes de Harris. *Ann. Inst. Fourier (Grenoble)*, 22:85–130.

Renyi, A. (1963) On stable sequences of events. *Sankhyā Ser. A*, 25:293–302.

Sawyer, S. A. (1970) A formula for semigroups, with an application to branching diffusion processes. *Trans. Amer. Math. Soc.*, 152:1–38.

Touati, A. (1987) Théorèmes limites pour des processus de Markov récurrents. *C. R. Acad. Sci. Paris Sér. I Math.*, 305:841–844.

Ueltzhöfer, F. A. J. and Klüppelberg, C. (2011) An oracle inequality for penalised projection estimation of Lévy densities from high-frequency observations. *J. Nonparametr. Stat.*, 23:967–989.

Watanabe, S. (1964) On discontinuous additive functionals and Lévy measures of a Markov process. *Japan. J. Math.*, 34:53–70.

Weil, M. (1971) Conditionnement par rapport au passé strict. *Séminaire de Probabiltés (Strasbourg)*, 5:362–372.