



HAL
open science

Diffusive limits of Lipschitz functionals of Poisson measures

Eustache Besançon, Laure Coutin, Laurent Decreusefond, Pascal Moyal

► **To cite this version:**

Eustache Besançon, Laure Coutin, Laurent Decreusefond, Pascal Moyal. Diffusive limits of Lipschitz functionals of Poisson measures. *The Annals of Applied Probability*, 2024, 34 (1A), pp.555-584. 10.1214/23-AAP1972 . hal-03283778v2

HAL Id: hal-03283778

<https://telecom-paris.hal.science/hal-03283778v2>

Submitted on 15 Dec 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DIFFUSIVE LIMITS OF LIPSCHITZ FUNCTIONALS OF POISSON MEASURES

E. BESANÇON, L. COUTIN, L. DECREUSEFOND, AND P. MOYAL

ABSTRACT. Continuous Time Markov Chains, Hawkes processes and many other interesting processes can be described as solution of stochastic differential equations driven by Poisson measures. Previous works, using the Stein's method, give the convergence rate of a sequence of renormalized Poisson measures towards the Brownian motion in several distances, constructed on the model of the Kantorovitch-Rubinstein (or Wasserstein-1) distance. We show that many operations (like time change, convolution) on continuous functions are Lipschitz continuous to extend these quantified convergences to diffusive limits of Markov processes and long-time behavior of Hawkes processes.

1. INTRODUCTION

Limit theorems for Continuous-Time Markov Chains (CTMC's) have proven to be useful tools to approximate the dynamics of the processes under consideration. Fluid approximations can be used to determine ergodicity conditions, a first order approximation of the mean dynamics of the process, or to analyze the dynamics of discrete-event systems around saturation. Likewise, diffusion approximations (also called, along the various communities, Functional Central Limit Theorems or invariance principles) usually lead to the weak approximation of the (properly scaled) difference between the original CTMC and its fluid limit, to a diffusion process. Such convergence results allow to assess the speed of convergence to the fluid limit, and thereby, to gain insights on the behavior of the considered process when the state space is large, and/or to more easily simulate its paths whenever the dynamics of the original discrete-event CTMC is too intricate.

The literature regarding weak (fluid or diffusion) approximations of CTMC's is vast, and so is the range of their fields of applications: for instance, queueing networks (see e.g. [6, 36, 26, 33] and references therein), biology and epidemics (e.g. [10, 20, 17] and references therein), physics [12], and so on. From the mathematical standpoint, we can enumerate at least three approaches to prove the corresponding convergence theorems. The most classical one relies on the so-called Dynkin's Lemma (see e.g. Chapter 7 in [21], or Chapter 7 in [19]), and more generally on the semi-martingale decomposition of the considered CTMC, together with the usual weak convergence theorems for martingales, to derive the limiting process of the properly scaled CTMC. An alternative way consists in representing the Markov process as the sum of time-changed Poisson processes and then to use the

Key words and phrases. Approximation diffusion, Hawkes processes, CTMC, Stein's method.

well-known limit theorems for such processes, see e.g. Chapter 6 in [21]. A third, alternative way is to represent the CTMC as the solution of a stochastic differential equation (SDE) driven by some independent Poisson random measures (see e.g. [10]). We mention these different approaches because not all of them behave nicely for what has to be done here.

As was already mentioned, the diffusion approximation of the CTMC at hand provides a crucial tool for simulating an approximate behavior of the system under consideration by a diffusion process. Among a wide variety of examples and applications, let us set the ideas by considering the classical case of an infinite server queue of traffic load ρ : it is well-known that the fluid limit of the process of initial value ρ is constant, and that the limiting behavior of the scaled version of the process is an Ornstein-Uhlenbeck process (see e.g. Section 6.6 of [36] for an exhaustive survey on the matter). In such cases, it then becomes of crucial interest to assess the *rate of convergence* in such diffusion approximations. This is the main object of the present work. We present hereafter a unified framework, based on the third aforementioned approach, namely, on the representation of CTMC's as solution of SDE's driven by Poisson random measures, to derive bounds for the convergence in the diffusion approximations of a wide class of CTMC's, under various mild conditions on the integrand of these Poisson integrals. By considering a wide range of cases study, from queueing systems to biological models and epidemiological processes, we show that these assumptions are met by many processes that are prevalent in practice. In many cases, we retrieve existing results concerning the diffusion approximations of the considered processes, and then go one step further, by establishing bounds for the latter convergence. We also show that the same procedure can be applied to study the long-run behavior of Hawkes processes.

For a sequence of processes $(X_n, n \geq 1)$ with values in a complete, separable, metric space which converges to a process X , estimating the rate of convergence amounts to computing φ such that

$$(1) \quad \text{dist}_{\mathcal{C}}(\mathbf{P}_{X_n}, \mathbf{P}_X) := \sup_{f \in \mathcal{C}} \mathbf{E}[f(X_n)] - \mathbf{E}[f(X)] \leq \varphi(n),$$

where \mathcal{C} is the set of test functions. The minimum regularity required for the left-hand-side of (1) to define a distance is to suppose f Lipschitz continuous, but not necessarily bounded.

If we take for \mathcal{C} the set of Lipschitz bounded functions, we obtain a distance which generates the same topology as that of the Prokhorov distance. In the seventies, many papers (see e.g. [25, 37], and references therein) derived the rates of convergence of functional CLT's such as Donsker's theorem, for this metric. They generally obtained $\varphi(n) = O(n^{-1/4})$ via the Skorohod representation theorem and ad-hoc subtle computations on the sample-paths themselves.

In the nineties, in his pioneering paper [2], Barbour constructed a Malliavin-like apparatus to estimate the rate of convergence in the Donsker theorem on the Skorohod space \mathbb{D} . The set \mathcal{C} under consideration is the set of three times Fréchet differentiable functions on \mathbb{D} with additional boundedness properties. Once this functional framework is setup, we can proceed similarly to the

Stein method in finite dimension (see [32] for a new application of this approach to the Moran model). Let us also mention several recent applications of the Stein method in finite dimension, assessing the rate of convergence of the stationary distributions of various queueing processes: Erlang-A and Erlang-C systems in [9]; a system with reneging and phase-type service time distributions in [8], and single-server queues in heavy-traffic in [24].

It is only recently that in [13], Barbour's result was extended to the convergence in some fractional Sobolev spaces, instead of \mathbb{D} . This result was then improved in [15] and [14] by allowing, at last, test functions that are only Lipschitz continuous. When \mathcal{C} is the set of Lipschitz continuous functions, the induced distance is stronger than the Prokhorov distance: not only does it imply the convergence in distribution, but also the convergence of the first moments, see [40].

For these test functions, from the theoretical point of view two novelties arise. As could be expected, their reduced regularity induces additional technicalities but more strikingly, it also yields different rates of convergence. The benefit is that the applicability is enriched, for this new set of test functions embraces many more functions of interest in practice. This can be used, for instance, to derive the convergence rate for the maximum of a random walk towards the maximum of the Brownian motion; a result which cannot be established by the basic Stein method, as we do not have a convenient characterization of the law of the maximum of the Brownian motion.

Furthermore, the set of Lipschitz functions is remarkably stable with respect to many operations like time-change, convolution, reflection, etc., so that we can deduce from a master Theorem, many new convergence rates which do not seem to be accessible from scratch.

In [4], Stein's method was used to study the rate of convergence in the diffusion approximations of the $M/M/1$ and the $M/M/\infty$ queues. The two models involved very different ad-hoc techniques and proved to be difficult to generalize, but led to a satisfying estimate of the speed of convergence ($n^{-1/2}$, where n is the scaling factor of the respective models). The approach of the present paper is more general, at the expense of a lower rate of convergence ($n^{-1/6}$), but covers a much wider class of processes. Our results can be used to assess the precision of the approximation of functionals of sample-path of a Markov process by the same functional for the limit diffusion. For instance, in [22], the excursions of the telegraph process are approximated by those of the limiting Ornstein-Uhlenbeck, the distribution of which is known. The rate of convergence for the limit theorem for the Hawkes process can also be used to obtain a functional version of the theorems in [28] for potential applications to mathematical finance.

During the preparation of this work, we became aware of [3] which addresses a similar problem. The test functions in [3] are much more regular than simply Lipschitz but the scope of the method seems wider as it can handle the convergence of some non Markovian processes. The price to pay seems to be a lower rate of convergence.

The paper is organized as follows. In Section 2, we give some estimates of the distance between a multivariate point process and its affine interpolation,

depending on the intensity of its jumps. In Section 3, we establish that sets of Lipschitz functionals on some function spaces enjoy some remarkable stability properties. These properties are crucial to transfer the convergence rate established in the master Theorem 4.7, to more general processes. In Section 5, we provide a general result regarding the rate of convergence in the diffusion approximations of a wide class of CTMC's, and then apply this result to various practical processes in queueing, biology and epidemiology. We then quantify the convergence of some functionals of Hawkes processes in Section 6.

2. PRELIMINARIES

Throughout the paper, we fix a time horizon $T > 0$. For a fixed integer d , we denote by \mathbb{D}_T the Skorohod space (i.e. the space of right continuous with left limits (rcll) functions from $[0, T]$ into \mathbf{R}^d). It is viewed as the canonical probability space and equipped with a filtration $\mathcal{F} = (\mathcal{F}_t, t \in [0, T])$ which satisfies the usual hypothesis. All the forthcoming processes are defined on this filtered probability space and all processes (but the affine interpolations) are assumed to be adapted to \mathcal{F} . The space \mathbb{D}_T contains \mathcal{C}_T , the space of continuous functions on $[0, T]$. We denote the sup-norm over $[0, T]$ by

$$\|f\|_{\infty, T} = \sup_{t \in [0, T]} \|f(t)\|_{\mathbf{R}^d},$$

for $f \in \mathbb{D}_T$. In what follows, inequalities will be valid up to irrelevant multiplicative constants, and we write

$$a \lesssim_{\alpha} b$$

to mean that there exists $c > 0$ which depends only on α such that $a \leq cb$. Likewise, for simplicity we often just write $a \lesssim b$, in cases where a and b both depend on the scaling parameter n , and $a \leq cb$ for a c that does not depend on n , but possibly on other parameters that are clear from the context.

Remark 1. *In the forthcoming examples, we consider processes whose sample-paths are only right-continuous-with-left-limits (rcll) and we wish to compare them to the Brownian motion (BM) or other diffusions whose sample-paths are continuous. In the usual proof of the Donsker theorem, the common probability space on which the convergence is proved is the Skorohod space of rcll functions on $[0, T]$, denoted by \mathbb{D}_T . We do not follow this path for several reasons. When we began this line of research in [13], we were viewing our processes as random variables in some fractional Sobolev spaces of not necessarily continuous functions because these spaces are separable Banach spaces and then yield a convenient framework (see also [7, 38]) to generalize the Stein's method to infinite dimension. Since \mathbb{D}_T equipped with its usual distance is a metric and not a Banach space, this procedure could not be applied to the Skorohod space. Because of the infinite dimensional aspects, we had to estimate the limit of a trace term, a quantity notably known to be difficult to handle. In [15], we realized that by considering affine interpolations, which are natural objects to deal with when working on random walks, we could bypass these delicate computations. By doing so, we can make all the calculations in some finite dimensional spaces but we still need the*

infinite dimensional structure given by the Wiener space to define the test functions as weakly differentiable functions as in Definition 2.5 (a notion of differentiability much weaker than Fréchet differentiability). Since \mathbb{D}_T when equipped with the sup norm is a non separable Banach space [31], it cannot be furnished with a Wiener space structure so we cannot define the concept of weakly differentiable functions on this space hence our restriction to the space of continuous functions.

Our first step is thus to replace the initial processes we want to study by their affine interpolations to work within $\mathcal{C}([0, T]; \mathbf{R}^d)$ instead of \mathbb{D}_T . The estimate of the induced error is easy via martingales techniques and can be done at a small cost in the convergence rate. Another advantage of the space of continuous functions is that since it is embedded in the Skorohod space, the set of Lipschitz functions defined on $\mathcal{C}([0, T]; \mathbf{R}^d)$ is larger than the set of Lipschitz functions defined on \mathbb{D}_T . Finally, by working with affine interpolations, we can detect which factor is responsible for the slowest rate of convergence: the difference of regularity or the difference in the dynamics. Actually, this is the equilibrium to be found between these rates that yields the unusual order of convergence in $n^{-1/6}$.

Definition 2.1. A partition π of $[0, T]$ is a sequence

$$\pi = \{0 = t_0 < t_1 < \dots < t_{l(\pi)} = T\},$$

where $l(\pi)$ is the number of subintervals defined by π . We denote by $|\pi|$ its mesh

$$|\pi| = \sup_{i \in \llbracket 0, l(\pi) - 1 \rrbracket} |t_{i+1} - t_i| + f(0).$$

We denote by Σ_T , the set of partitions of $[0, T]$.

For any function $f \in \mathbb{D}_T$ and any $\pi \in \Sigma_T$, we denote by $\Xi_\pi f$ the affine interpolation of f on $[0, T]$ along π , namely for all $t \in [0, T]$,

$$\begin{aligned} (\Xi_\pi f)(t) &= \sum_{i=0}^{l(\pi)-1} \left[\frac{f(t_{i+1}) - f(t_i)}{t_{i+1} - t_i} (t - t_i) + f(t_i) \right] \mathbf{1}_{[t_i, t_{i+1})}(t) \\ &= \sum_{i=0}^{l(\pi)-1} \frac{f(t_{i+1}) - f(t_i)}{\sqrt{t_{i+1} - t_i}} h_i^\pi(t) + f(0), \end{aligned}$$

where

$$h_i^\pi(t) = \frac{1}{\sqrt{t_{i+1} - t_i}} \int_0^t \mathbf{1}_{[t_i, t_{i+1})}(s) ds, \quad i \in \llbracket 0, l(\pi) - 1 \rrbracket.$$

When $\pi = \{iT/n, i \in \llbracket 0, n \rrbracket\}$, we denote Ξ_π by Ξ_n and h_i^π by h_i^n for all i .

When a point process has not too many jumps per subinterval of a partition π , its affine interpolation along π does not deviate too much from its nominal path. More precisely we have the following result,

Theorem 2.2. Let $m \in \mathbf{N}^*$, and consider $(X(t), t \in [0, T])$ a \mathbf{R}^d -valued point process admitting the representation

$$(2) \quad X(t) = \sum_{k=1}^m \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \varphi_k(s, X(s^-))\}} d\mathcal{N}_k(s, z) \right) \zeta_k, \quad t \in [0, T],$$

where the \mathcal{N}_k 's are independent Poisson measures of respective intensity measures $\rho_k \, ds \otimes dz$, $k \in \llbracket 1, m \rrbracket$, and for any $k \in \llbracket 1, m \rrbracket$, $\zeta_k \in \mathbf{R}^d$ and φ_k is a bounded function $[0, T] \times \mathbf{R}^d \rightarrow \mathbf{R}^+$. Then, for any $\pi \in \Sigma_T$ we have that

$$\mathbf{E} [\|X - \Xi_\pi X\|_{\infty, T}] \lesssim \sum_{k=1}^m \|\zeta_k\|_{\mathbf{R}^d} \Psi\left(l(\pi), \rho_k |\pi| \|\varphi_k\|_{\infty}\right),$$

where for all $(q, \nu) \in \mathbf{N}^* \times \mathbf{R}_+$,

$$\Psi(q, \nu) = \frac{\log(qe^{-\nu})}{\log((\nu e)^{-1} \log(qe^{-\nu})) - \log \log((\nu e)^{-1} \log(qe^{-\nu}))}.$$

Proof of Theorem 2.2. Let $\pi = \{t_i, i \in \llbracket 0, n \rrbracket\}$. For any $t \in [0, T[$ there exists $i \leq n-1$ such that $t \in [t_i, t_{i+1})$ and

$$\begin{aligned} \|X(t) - \Xi_\pi X(t)\|_{\mathbf{R}^d} &= \left\| X(t) - X(t_i) - \frac{X(t_{i+1}) - X(t_i)}{t_{i+1} - t_i} (t - t_i) \right\|_{\mathbf{R}^d} \\ &\leq 2 \sup_{t \in [t_i, t_{i+1}]} \|X(t) - X(t_i)\|_{\mathbf{R}^d}. \end{aligned}$$

Now notice that for all $k \in \llbracket 1, m \rrbracket$, for all $t \in [t_i, t_{i+1}]$,

$$\begin{aligned} \int_{t_i}^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \varphi_k(s, X(s^-))\}} \, d\mathcal{N}_k(s, z) &\leq \int_{t_i}^{t_{i+1}} \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \|\varphi_k\|_{\infty}\}} \, d\mathcal{N}_k(s, z) \\ &:= M_{k,i}^\pi. \end{aligned}$$

Hence,

$$\sup_{t \in [t_i, t_{i+1}]} \|X(t) - X(t_i)\|_{\mathbf{R}^d} \leq \sum_{k=1}^m \|\zeta_k\|_{\mathbf{R}^d} M_{k,i}^\pi$$

so that

$$\mathbf{E} [\|X - \Xi_\pi X\|_{\infty, T}] \leq 2 \sum_{k=1}^m \|\zeta_k\|_{\mathbf{R}^d} \mathbf{E} \left[\max_{i \in \llbracket 0, n-1 \rrbracket} M_{k,i}^\pi \right].$$

But the $M_{k,i}^\pi$, $i \in \llbracket 0, l(\pi)-1 \rrbracket$ are independent Poisson random variables of respective parameters $\rho_k (t_{i+1} - t_i) \|\varphi_k\|_{\infty}$, so they are strongly dominated by a family of $l(\pi)$ independent Poisson random variables of respective parameters $\rho_k |\pi| \|\varphi_k\|_{\infty}$. The result then follows from Proposition 7.1 below. \square

As to the affine interpolation of the Brownian motion, according to [23, Proposition 13.20], there exists $c > 0$ such that for any partition π

$$(3) \quad \mathbf{E} \left[\|\Xi_\pi B - B\|_{W_{\eta,p}}^p \right]^{1/p} \lesssim |\pi|^{1/2-\eta}.$$

2.1. Malliavin gradient. We give the minimum elements of Malliavin calculus to understand the sequel (see [16] for details). More advanced material is necessary to prove Theorem 4.2, see [14]. We denote by \mathcal{H}_T the Hilbert space, often quoted as the Cameron-Martin space,

$$\mathcal{H}_T := \left\{ h \in \mathcal{C}_T, \exists \dot{h} \in L^2([0, T], ds) \text{ such that } h(t) = \int_0^t \dot{h}(s) \, ds \right\}.$$

The function \dot{h} is unique so that we can define

$$\|h\|_{\mathcal{H}_T}^2 = \int_0^T \dot{h}(s)^2 \, ds.$$

By convention, we identify \mathcal{H}_T with its dual $(\mathcal{H}_T)^*$.

Consider $(Z_n, n \geq 1)$, a sequence of independent, standard Gaussian random variables and let $(z_n, n \geq 1)$ be a complete orthonormal basis of \mathcal{H}_T . Then, we know from [29] that

$$(4) \quad \sum_{n=1}^N Z_n z_n \xrightarrow{N \rightarrow \infty} B := \sum_{n=1}^{\infty} Z_n z_n \text{ in } \mathcal{C}_T \text{ with probability 1,}$$

where B is a Brownian motion. We clearly have the diagram

$$(5) \quad \mathcal{C}_T^* \xrightarrow{\mathbf{e}^*} (\mathcal{H}_T)^* \simeq \mathcal{H}_T \xrightarrow{\mathbf{e}} \mathcal{C}_T,$$

where \mathbf{e} is the canonical embedding from \mathcal{H}_T into \mathcal{C}_T . We denote by μ the law of B on \mathcal{C}_T , and by $L^2(\mathcal{C}_T; \mu)$, the space of functions F from \mathcal{C}_T into \mathbf{R} such that

$$\mathbf{E}[F(B)^2] < \infty.$$

Definition 2.3 (Wiener integral). *The Wiener integral, denoted as δ , is the isometric extension of the map*

$$\begin{aligned} \delta : \mathbf{e}^*(\mathcal{C}_T^*) \subset \mathcal{H}_T &\longrightarrow L^2(\mathcal{C}_T; \mu) \\ \mathbf{e}^*(\eta) &\longmapsto \langle \eta, y \rangle_{\mathcal{C}_T^*, \mathcal{C}_T}. \end{aligned}$$

This means that if $h = \lim_{n \rightarrow \infty} \mathbf{e}^*(\eta_n)$ in \mathcal{H}_T ,

$$\delta h(B) = \lim_{n \rightarrow \infty} \langle \eta_n, y \rangle_{\mathcal{C}_T^*, \mathcal{C}_T} \text{ in } L^2(\mu).$$

Definition 2.4. *Let X be a Banach space. A function $F : \mathcal{C}_T \rightarrow X$ is said to be cylindrical if it is of the form*

$$F(y) = \sum_{j=1}^k f_j(\delta h_1(y), \dots, \delta h_k(y)) x_j,$$

where for any $j \in \llbracket 1, k \rrbracket$, f_j belongs to the Schwartz space on \mathbf{R}^k , (h_1, \dots, h_k) are elements of \mathcal{H}_T and (x_1, \dots, x_k) belong to X . The set of such functions is denoted by \mathbb{X} .

For $h \in \mathcal{H}_T$,

$$\langle \nabla F, h \rangle_{\mathcal{H}_T} = \sum_{j=1}^k \sum_{l=1}^k \partial_l f(\delta h_1(y), \dots, \delta h_k(y)) \langle h_l, h \rangle_{\mathcal{H}_T} x_j,$$

which is equivalent to say

$$\nabla F = \sum_{j,l=1}^k \partial_j f(\delta h_1(y), \dots, \delta h_k(y)) h_l \otimes x_j.$$

The space $\mathbb{D}_{1,2}(X)$ is the closure of cylindrical functions with respect to the norm of $L^2(\mathcal{C}_T; \mathcal{H}_T \otimes X)$. An element of $\mathbb{D}_{1,2}(X)$ is said to be Gross-Sobolev differentiable and ∇F belongs to $\mathcal{H}_T \otimes X$ with probability 1.

We can iterate the construction to higher order gradients and thus define $\nabla^{(k)} F$ for any $k \geq 1$, provided that F is sufficiently regular.

We actually need a stronger notion of weak differentiability. The ordinary notion of gradient we have just defined, induces that ∇F belongs almost surely to \mathcal{H}_T . Hereafter we also need that it belongs to a smaller space, namely the dual of $L^2([0, T], ds)$. Recall that we have identified \mathcal{H}_T with itself, so that we cannot identify L^2 with its dual. It is proved in [15] that

$$\begin{aligned} \mathcal{J}_T &:= (L^2([0, T], ds))^* \\ &\simeq \left\{ h \in \mathcal{C}_T, \exists \hat{h} \in L^2([0, T], ds) \text{ such that } h(t) = \int_0^t \int_s^1 \hat{h}(u) du ds \right\}. \end{aligned}$$

Definition 2.5. We denote by Υ_T the subset of functions F in $\mathbb{D}_{2,2}(\mathbf{R})$ which satisfy

$$(6) \quad \left| \left\langle \nabla^{(2)} F(x) - \nabla^{(2)} F(x + g), h \otimes k \right\rangle_{\mathcal{H}_T} \right| \leq \|g\|_{\mathcal{C}_T} \|h\|_{L^2} \|k\|_{L^2},$$

for any $x \in \mathcal{C}_T$, $g \in \mathcal{H}_T$, $h, k \in L^2([0, T], ds)$. This means that $\nabla^{(2)} F$ belongs to $\Upsilon_T^{\otimes 2}$ and is \mathcal{H}_T -Lipschitz continuous on \mathcal{C}_T .

3. LIPSCHITZ FUNCTIONALS

Definition 3.1. Let (E, d_E) and (G, d_G) be two metric spaces. A function $F : E \rightarrow G$ is said to be Lipschitz continuous whenever there exists $c > 0$ such that for any $x, y \in E$,

$$(7) \quad d_G(F(x), F(y)) \leq c d_E(x, y).$$

The minimum value of c such that (7) holds, is the Lipschitz norm of F . We denote by $\text{Lip}_\alpha(E \rightarrow G, d_E)$ the set of Lipschitz continuous functions from E to G having Lipschitz norm less than α .

When E is a functional space, the set of Lipschitz functions is rich enough to be stable by some interesting transformations.

Lemma 3.2. Let r be a positive integrable function on $[0, T]$ and set

$$(8) \quad \gamma(t) = \int_0^t r(s) ds.$$

Then, the map

$$\begin{aligned} \Gamma : \mathcal{C}_T &\longrightarrow \mathcal{C}_{\gamma(T)} \\ f &\longmapsto f \circ \gamma^{-1} \end{aligned}$$

is invertible and Lipschitz continuous with Lipschitz norm 1. Moreover, if F belongs to Υ_T then $F \circ \Gamma$ belongs to $\Upsilon_{\gamma(T)}$.

Proof. The first part is straightforward. Since Γ is a linear continuous and bijective map from \mathcal{C}_T to $\mathcal{C}_{\gamma(T)}$, which maps \mathcal{H}_T (respectively \mathcal{J}_T) bijectively onto $\mathcal{H}_{\gamma(T)}$ (respectively $\mathcal{J}_{\gamma(T)}$), the second assertion follows immediately. \square

Let us now fix $m, d \in \mathbf{N}^*$, and let $\mathfrak{M}_{d,m}(\mathbf{R})$ be the space of $d \times m$ matrices with real entries. Consider the integral equation

$$(9) \quad y(t) = \int_0^t \mathbf{A}y(s) ds + f(t), \quad t \geq 0,$$

where $\mathbf{A} \in \mathfrak{M}_{d,m}(\mathbf{R})$ and $f \in \mathcal{C}_T(\mathbf{R}^d)$. The unique solution of (9) can be written as

$$(10) \quad \Theta_{\mathbf{A}}(f)(t) = f(t) + \int_0^t \mathbf{A}e^{(t-r)\mathbf{A}} f(r) \, dr.$$

We have the following,

Theorem 3.3. *For $\mathbf{A} \in \mathfrak{M}_{d,m}(\mathbf{R})$, the map $\Theta_{\mathbf{A}}$ defined by (10) is Lipschitz-continuous from $\mathcal{C}_T(\mathbf{R}^d)$ to itself. Moreover, if F belongs to Υ_T , then so does $F \circ \Theta_{\mathbf{A}}$.*

Proof. Since $\Theta_{\mathbf{A}}$ is linear, we just have to prove that for any $f \in \mathcal{C}_T$,

$$\|\Theta_{\mathbf{A}}(f)\|_{\infty,T} \lesssim_{\mathbf{A}} \|f\|_{\infty,T}.$$

From (10), we have

$$\|\Theta_{\mathbf{A}}(f)\|_{\infty,T} \lesssim (1 + \|\mathbf{A}\|)e^{T\|\mathbf{A}\|} \|f\|_{\infty,T}.$$

Equation (10) also entails that $\Theta_{\mathbf{A}}$ is linear and that

$$\Theta_{\mathbf{A}}(\mathcal{H}_T) \subset \mathcal{H}_T \text{ and } \Theta_{\mathbf{A}}(\mathcal{J}_T) \subset \mathcal{J}_T,$$

hence Υ_T is stable by $\Theta_{\mathbf{A}}$. The proof is thus complete. \square

Recall (see e.g. Chapter D in [36]), that for any $T > 0$ and any $Y \in \mathbb{D}_T$ such that $Y(0) \geq 0$, there exists a unique pair of functions X_Y and R_Y in \mathbb{D}_T such that $X_Y(t)$ is non-negative, R_Y is non-decreasing, $R_Y(0) = 0$ and for all $t \leq T$,

$$\begin{cases} X_Y(t) = Y(t) + R_Y(t), \\ \int_0^t X_Y(s) \, dR_Y(s) = 0. \end{cases}$$

Define the mapping

$$(11) \quad \text{Sko} : \mathbb{D}_T \longrightarrow \mathbb{D}_T \\ f \longmapsto (s \mapsto f(s) + \|f^-\|_{\infty,s}),$$

usually referred to as the *Skorokhod reflection map* of f . Then, it is well known that in the particular case $d = 1$, X_Y has the explicit form $X_Y = \text{Sko}(Y)$. We have the following results,

Theorem 3.4. *The mapping*

$$\begin{aligned} \max : \mathbb{D}_T &\longrightarrow \mathbb{D}_T \\ f &\longmapsto (s \mapsto \|f\|_{\infty,s}), \end{aligned}$$

the local time map

$$\begin{aligned} \ell^0 : \mathbb{D}_T &\longrightarrow \mathbb{D}_T \\ f &\longmapsto (s \mapsto \|f^-\|_{\infty,s}) \end{aligned}$$

and the Skorokhod reflection map Sko are all Lipschitz continuous, and so is for any $\varepsilon > 0$ the continuity modulus mapping

$$\begin{aligned} \alpha_{\varepsilon} : \mathbb{D}_T &\longrightarrow \mathbf{R}^+ \\ f &\longmapsto \sup_{|s-s'| \leq \varepsilon} \|f(s) - f(s')\|_{\mathbf{R}^d}. \end{aligned}$$

Proof. The first three assertions readily follow from the left triangular inequality, entailing that for all f, g , and all $t \in [0, T]$,

$$\left| \|f\|_{\infty, t} - \|g\|_{\infty, t} \right| \leq \|f - g\|_{\infty, t}.$$

Regarding the last assertion, we clearly have for all $s \leq T$,

$$\|f(s) - f(s')\|_{\mathbf{R}^d} \leq \|f(s) - g(s)\|_{\mathbf{R}^d} + \|g(s) - g(s')\|_{\mathbf{R}^d} + \|g(s') - f(s')\|_{\mathbf{R}^d}.$$

Hence for all ε ,

$$\alpha_\varepsilon(f) \leq 2\|f - g\|_{\infty, T} + \alpha_\varepsilon(g).$$

The same holds with the role of f and g permuted, hence

$$\|\alpha_\varepsilon(f) - \alpha_\varepsilon(g)\|_{\mathbf{R}^d} \leq 2\|f - g\|_{\infty, T},$$

and the proof is complete. \square

4. KANTOROVITCH-RUBINSTEIN DISTANCES

Definition 4.1. For μ and ν two probability measures on a metric space (E, d_E) , the Kantorovitch-Rubinstein (or Wasserstein-1) distance between μ and ν is defined as

$$\mathfrak{K}_E(\mu, \nu) := \sup_{F \in \text{Lip}_1(E \rightarrow \mathbf{R}, d_E)} \int_E F \, d\mu - \int_E F \, d\nu.$$

It is the well known [40] that $(\mu_n, n \geq 1)$ tends to ν in the Kantorovitch-Rubinstein topology if and only if μ_n converges in law to ν and the sequence of first order moments converges: For some $x_0 \in E$ and then all $x_0 \in E$

$$(12) \quad \int_E d_E(x, x_0) \, d\mu_n(x) \xrightarrow{n \rightarrow \infty} \int_E d_E(x, x_0) \, d\nu(x).$$

In [14], it is proved that

Theorem 4.2. Let $(X_i, i \geq 0)$ be a sequence of independent and identically centered, with unit variance and finite third moment distributed random variables belonging to L^p for some $p \geq 3$. Then, for any $n \geq 1$,

$$(13) \quad \mathfrak{K}_{\mathcal{C}_T} \left(\sum_{i=0}^{n-1} X_i h_i^n, B \right) \lesssim_{T, \mathbf{E}[|X_1|^3]} n^{-1/6} \log(n).$$

To explain the somehow surprising exponent $-1/6$, we quickly describe the proof of this result. We proceed to a sort of bias-variance decomposition: for any $N < n$

$$\begin{aligned} \mathfrak{K}_{\mathcal{C}_T} \left(\sum_{i=0}^{n-1} X_i h_i^n, B \right) &\leq \mathfrak{K}_{\mathcal{C}_T} \left(\sum_{i=0}^{n-1} X_i h_i^n, \Xi_N \left(\sum_{i=0}^{n-1} X_i h_i^n \right) \right) \\ &\quad + \mathfrak{K}_{\mathcal{C}_T} \left(\Xi_N \left(\sum_{i=0}^{n-1} X_i h_i^n \right), \Xi_N B \right) \\ &\quad + \mathfrak{K}_{\mathcal{C}_T} (\Xi_N B, B) \\ &= A_1 + A_2 + A_3. \end{aligned}$$

We have seen in (3) that the rightmost term is bounded by $n^{-1/2}$. The proof of the latter inequality is based on the scaling invariance and Hölder

continuity of the Brownian motion, hence there are strong reasons to believe that this rate is optimal. Direct computations show that A_1 is also bounded by $n^{-1/2}$.

It remains to bound the median term A_2 . For this, we remark that the two processes under study belong to the finite dimensional space

$$\mathcal{H}_N = \text{span} \{h_i^N, i \in \llbracket 0, N-1 \rrbracket\}.$$

This leads to the definition of finite rank functional,

Definition 4.3. For $\pi \in \Sigma_T$, let

$$\mathcal{H}_\pi = \text{span} \{h_i^\pi, i = 0, \dots, l(\pi) - 1\}.$$

A function $F : \mathcal{C}_T \rightarrow \mathbf{R}$ is then said to have a finite rank if there exists a partition $\pi \in \Sigma_T$ and a function $\varphi_F : \mathcal{H}_\pi \rightarrow \mathbf{R}$ such that

$$(14) \quad F = \varphi_F \circ \Xi_\pi.$$

It amounts to saying that $F(x)$ depends only on the value of x by its value at the points of π . For any $\pi \in \Sigma_T$, we denote by \mathcal{F}^π the set of functions of finite rank associated to π , that is, such that (14) holds.

A straightforward adaptation of the proof of the main theorem in [14] yields

Theorem 4.4. For any $\pi \in \Sigma_T$, set

$$\text{Lip}_1^\pi = \text{Lip}_1(\mathcal{C}_T \rightarrow \mathbf{R}, \|\cdot\|_{\infty, T}) \cap \mathcal{F}^\pi.$$

Then, we have for all $n \in \mathbf{N}^*$,

$$(15) \quad \sup_{F \in \text{Lip}_1^\pi} \left\{ \mathbf{E} \left[F \left(\Xi_\pi \left(\sum_{i=0}^{n-1} X_i h_i^n \right) \right) \right] - \mathbf{E} [F(\Xi_\pi B)] \right\} \lesssim_{T, \mathbf{E}[|X_1|^3]} \frac{|\pi|^{-1}}{\sqrt{n}} \log(n).$$

If we apply this theorem to Ξ_N , we are in a position similar (but not equivalent) to that of bounding the Wasserstein-1 distance between a sum of independent (but not identically distributed) random vectors in dimension N , and the standard Gaussian distribution in \mathbf{R}^N . In this finite dimensional situation, the best known results [35, 5] show that the multiplying constant of the factor $n^{-1/2} \log(n)$ depends linearly on the dimension, a fact that we retrieve here.

The exponent $-1/6$ is then obtained by choosing the optimal N as a function of n . We have strong confidence that each of the two steps gives the optimal rate, and consequently, that this overall rate is optimal. For many practical applications, test functions in Lip_1^π are sufficient, see e.g. the simple and practical functionals addressed in Section 3. This leads us to introduce the following distance,

$$\mathfrak{F}_{\mathcal{C}_T}^\pi(\mu, \nu) = \sup_{F \in \text{Lip}_1^\pi} \left\{ \int_{\mathcal{C}_T} F \, d\mu - \int_{\mathcal{C}_T} F \, d\nu \right\},$$

for μ, ν two probability measures.

If we allow to take the supremum over a smaller set of test functions like three times Fréchet differentiable with bounded derivatives (as in [2, 32]),

we get a convergence rate bounded by $n^{-1/2}$. However, we can get this rate even for much less regular test functions. Let

$$\text{Lip}_1^\Upsilon = \text{Lip}_1(\mathcal{C}_T \rightarrow \mathbf{R}, \| \cdot \|_{\infty, T}) \cap \Upsilon_T.$$

Then, setting for all μ, ν ,

$$\mathfrak{J}_{\mathcal{C}_T}(\mu, \nu) = \sup_{F \in \text{Lip}_1^\Upsilon} \left\{ \int_{\mathcal{C}_T} F \, d\mu - \int_{\mathcal{C}_T} F \, d\nu \right\},$$

it is shown in [15] that

$$\mathfrak{J}_{\mathcal{C}_T} \left(\Xi_\pi \left(\sum_{i=0}^{n-1} X_i h_i^n \right), B \right) \lesssim_{T, \pi} \frac{1}{\sqrt{n}}.$$

In what follows, we deal with renormalized stochastic integrals with respect to Poisson measures as in (2). One of the trick of our proof is to replace the random integrands by deterministic and supposedly close functions. This technical step introduces an error which converges to zero at rate $n^{-1/4}$ (see (30)), so much slower than $n^{-1/2} \log(n)$. We thus redefine the usual rates at which the convergence hold.

Definition 4.5. *A sequence of rcll processes $(X_n, n \geq 1)$ is said to converge in distribution at the usual rates to a process Z if for any $\pi \in \Sigma_T$ and any $n \geq 1$,*

$$(†) \quad \begin{cases} \mathfrak{K}_{\mathcal{C}_T}(\Xi_n X_n, Z) & \lesssim_T n^{-1/6} \log(n), \\ \mathfrak{F}_{\mathcal{C}_T}^\pi(\Xi_n X_n, Z) & \lesssim_T |\pi|^{-1} n^{-1/4}, \\ \mathfrak{J}_{\mathcal{C}_T}(\Xi_n X_n, Z) & \lesssim_T n^{-1/4}. \end{cases}$$

A straightforward application of the previous results leads to the following,

Theorem 4.6. *Let P_n be a Poisson process on $[0, T]$ of intensity n . Then, the sequence of processes $(\bar{P}_n, n \geq 1)$ defined by*

$$(16) \quad \bar{P}_n(t) := \frac{P_n(t) - nt}{\sqrt{n}}, \quad t \leq T,$$

converges at the usual rates to a Brownian motion. Furthermore, for all $n \geq 1$,

$$\mathbf{E} [\| \bar{P}_n - \Xi_n \bar{P}_n \|_{\infty, T}] \lesssim_T \frac{\log(n)}{\sqrt{n}}.$$

Proof. Fix $n \geq 1$. According to Theorem 2.2, we get that

$$\mathbf{E} [\| \bar{P}_n - \Xi_n \bar{P}_n \|_{\infty, T}] \lesssim_T \frac{\Psi(n, 1)}{\sqrt{n}} \lesssim_T \frac{\log(n)}{\sqrt{n}}.$$

Furthermore,

$$\begin{aligned} \Xi_n \bar{P}_n &= \sum_{i=0}^{n-1} \frac{P_n((i+1)/n) - P_n(i/n) - 1}{\sqrt{n} \sqrt{1/n}} h_i^n \\ &\stackrel{\text{dist.}}{=} \sum_{i=0}^{n-1} (X_i - 1) h_i^n \end{aligned}$$

where $(X_k, k \geq 1)$ is an IID sequence of Poisson random variables of parameter 1. The result then follows from (13) and (15). \square

Remark 2. *In our computations, we do not follow the values of the constants which appear in the different inequalities. It is a tedious and lengthy work to compute each of them so that we decided not to include these calculations in this paper for the sake of clarity. One can still ponder how the constants vary with the dimension of the process, or equivalently of the limiting Brownian motion. If we look closely at the proof of the distance between $\Xi_\pi B$ and B in [23], we easily see that the constant is proportional to $d^{1/2}$ where d is the dimension. For the approximation of the linear interpolation of the initial process, Theorem 2.2 says the constants depend on the growth of $\sup_k \|\zeta_k\|_{\mathbf{R}^d}$ when d is increasing. It is thus impossible to have a generic answer taking into account all situations. If the ζ 's are well behaved, i.e. if their components do not grow too much with the dimension ($\sup_d \sup_{k=1, \dots, m} \sup_{i=1, \dots, d} |\zeta_k(i)| < \infty$) then the constants is linear with the dimension. For the core of the Stein's method, each component is handled individually, i.e., there is no compensation between the coordinates, hence the constants vary as the dimension. At the end, the constants are of the order of the dimension when the latter is large.*

We are now in a position to state our main result. It consists of an extension of the last Theorem to stochastic integral with respect to Poisson measures,

Theorem 4.7. *Let r be a positive integrable function on $[0, T]$, and γ be defined by (8). Let \mathcal{N}^n be a Poisson measure on $[0, T] \times \mathbf{R}^+$ of intensity measure $n \, dt \otimes dz$. Consider the compensated Poisson measure*

$$(17) \quad d\tilde{\mathcal{N}}^n(t, z) = d\mathcal{N}^n(t, z) - n \, dt \otimes dz.$$

Define also the process \bar{R}_n by

$$\bar{R}_n(t) = \frac{1}{\sqrt{n}} \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq r(s)\}} d\tilde{\mathcal{N}}^n(s, z), \quad t \in [0, T].$$

Then, $(\bar{R}_n, n \geq 1)$ converges at the usual rates to $B \circ \gamma$.

Proof. According to [11, Theorem 16], for any $n \geq 1$, $\bar{R}_n \circ \gamma^{-1}$ has the distribution of \bar{P}_n defined in (16). Therefore, from Theorem 4.6, $(\bar{R}_n \circ \gamma^{-1}, n \geq 1)$ converges at the usual rates to B . The result then follows from Lemma 3.2. \square

We can now use the last theorem as a blackbox that yields the convergence rate in numerous applications. The main limitation is that Theorem 4.7 requires the integrand of \bar{R}_n to be deterministic when most often we have to deal with stochastic integrands. It turns out that in limit theorems, these stochastic integrands tend to some deterministic process. The idea is then to control the error made by replacing these integrands by their deterministic limit and then to apply Theorem 4.7.

5. APPLICATION TO CONTINUOUS TIME MARKOV CHAINS

5.1. General settings. Numerous Continuous time Markov chains (CTMC's) can be described as solutions of stochastic differential equations with respect to a finite family of Poisson measures. For any $m \in \mathbf{N}^*$, any family $(\zeta_1, \dots, \zeta_m)_{1 \leq k \leq m}$ of elements of \mathbf{R}^d and any array $(\rho_k)_{1 \leq k \leq m}$ of mappings

from $[0, T] \times \mathbb{D}_T$ to \mathbf{R} , consider the \mathbf{R}^d -valued process X defined as the solution of the SDE

$$X(t) = X(0) + \sum_{k=1}^m \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \rho_k(s, X)\}} d\mathcal{N}_k(s, z) \right) \zeta_k, \quad t \leq T,$$

where $X(0) \in \mathbf{R}^d$ is fixed, and $(\mathcal{N}_k)_{1 \leq k \leq m}$ denote m independent Poisson measures of unit intensity $ds \otimes dz$.

Fix $n \in \mathbf{N}^*$ and an array $\alpha := (\alpha_1, \dots, \alpha_m) \in (\mathbf{R})^m$. We scale the process X by replacing for all k , the measure \mathcal{N}_k by a Poisson measure $\mathcal{N}_k^{\alpha_k}$ of intensity $(n^{\alpha_k} ds) \otimes dz$, and normalizing in space by n . Then the process $\bar{X}_n := n^{-1}X_n$ is the solution of the following SDE: For any $t \geq 0$,

$$(18) \quad \bar{X}_n(t) = \bar{X}_n(0) + \frac{1}{n} \left(\sum_{k=1}^m \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \rho_k(s, n\bar{X}_n)\}} d\mathcal{N}_k^{\alpha_k}(s, z) \right) \zeta_k.$$

The key assumption on our scaling is the following:

Assumption 1 (Law of large numbers scaling). *For any $k \in \llbracket 1, m \rrbracket$, there exists a mapping $r_k \in \text{Lip}_c(\mathbf{R}^d \rightarrow \mathbf{R}, \|\cdot\|_{\mathbf{R}^d})$ such that for all $n \in \mathbf{N}^*$, all $x \in \mathbf{R}^d$ and $t \leq T$,*

$$n^{\alpha_k - 1} \rho_k(t, nx) = r_k(t, x),$$

and such that

$$(19) \quad \sup_{n \geq 1} \mathbf{E} \left[\sup_{t \leq T} |r_k(t, \bar{X}_n(t))| \right] \lesssim_{k, T} 1.$$

Observe that crucially, under Assumption 1 the r_k 's do not depend on n . We denote for any k , by $\tilde{\mathcal{N}}_k^{\alpha_k}$, the compensated Poisson measures of $\mathcal{N}_k^{\alpha_k}$, that is, we let

$$d\tilde{\mathcal{N}}_k^{\alpha_k}(s, z) = d\mathcal{N}_k^{\alpha_k}(s, z) - n^{\alpha_k} (ds \otimes dz).$$

Then, for all n and all k , the \mathbf{R}^d -valued process M_{n, k, \bar{X}_n} , defined for all $t \leq T$ by

$$(20) \quad M_{n, k, \bar{X}_n}(t) = \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq n^{1-\alpha_k} r_k(s, \bar{X}_n(s-))\}} d\tilde{\mathcal{N}}_k^{\alpha_k}(s, z) \right) \zeta_k$$

is a martingale with respect to the natural filtration of the Poisson measures. It then follows from (18) that for all n and t ,

$$(21) \quad \bar{X}_n(t) = \bar{X}_n(0) + \sum_{k=1}^m \left(\int_0^t r_k(s, \bar{X}_n(s)) ds \right) \zeta_k + \sum_{k=1}^m n^{-1} M_{n, k, \bar{X}_n}(t).$$

In view of Assumption 1 and the Cauchy-Lipschitz Theorem, there exists a unique solution Λ in $\mathcal{C}([0, T]; \mathbf{R}^d)$ to the integral equation

$$(22) \quad \Lambda(t) = \Lambda(0) + \sum_{k=1}^m \left(\int_0^t r_k(s, \Lambda(s)) ds \right) \zeta_k, \quad t \geq 0,$$

for a fixed $\Lambda(0) \in \mathbf{R}^d$. We have the following law of large numbers,

Theorem 5.1. *Assume that there exists a solution Λ of (22) on $[0, T]$. Suppose that Assumption 1 holds. Then, there exists $c > 0$ such that*

$$\sup_{n \geq 1} \mathbf{E} \left[\sup_{t \leq T} \|\bar{X}_n(t)\| \right] \leq e^{cT}$$

and for all $n \geq 1$,

$$(23) \quad \mathbf{E} \left[\sup_{t \leq T} \|\bar{X}_n(t) - \Lambda(t)\| \right] \lesssim_T \left(\mathbf{E} [\|\bar{X}_n(0) - \Lambda(0)\|] + n^{-1/2} \right) e^{cT}.$$

Proof. By denoting for all $k \in \llbracket 1, m \rrbracket$, by λ_k the Lipschitz constant of the mapping r_k , we readily get that for all $t \geq 0$,

$$\begin{aligned} \|\bar{X}_n(t) - \Lambda(t)\| &\leq \|\bar{X}_n(0) - \Lambda(0)\| \\ &\quad + \left(\sum_{k=1}^m \lambda_k \|\zeta_k\| \right) \int_0^t \|\bar{X}_n(s) - \Lambda(s)\| \, ds + \sum_{k=1}^m \left\| n^{-1} M_{n,k,\bar{X}_n}(t) \right\|, \end{aligned}$$

and it is then a classical consequence of Gronwall Lemma that for $c := \sum_{k=1}^m \lambda_k \|\zeta_k\|$,

$$\begin{aligned} \mathbf{E} \left[\sup_{t \leq T} \|\bar{X}_n(t) - \Lambda(t)\| \right] \\ \leq \left(\mathbf{E} [\|\bar{X}_n(0) - \Lambda(0)\|] + \mathbf{E} \left[\sup_{t \leq T} \sum_{k=1}^m \left\| n^{-1} M_{n,k,\bar{X}_n}(t) \right\| \right] \right) e^{cT}. \end{aligned}$$

We conclude using the Burkholder-Davis-Gundy inequality in view of (19). \square

We now turn to the so-called diffusion scaling of the process X . We study the sequence of processes $(U_n, n \geq 1)$ defined for all n by

$$U_n = n^{1/2} (\bar{X}_n - \Lambda), \quad n \geq 1.$$

In view of (20), we see that the integrands in the M_{n,k,\bar{X}_n} 's, $k \in \llbracket 1, m \rrbracket$, are random processes that depend on n . In this form, we cannot directly use Theorem 4.7. The key idea is to introduce intermediate martingales with deterministic integrands. The error made with this additional process is easily controlled by sample-paths estimates.

Introduce for all n , the \mathbf{R}^d -valued martingales $(M_{n,k,\bar{X}_n})_{k \in \llbracket 1, m \rrbracket}$, $(M_{n,k,\Lambda})_{k \in \llbracket 1, m \rrbracket}$ and $\bar{M}_{n,\Lambda}$, respectively defined for all $t \leq T$ by

$$\begin{cases} M_{n,k,\Lambda}(t) &= \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq n^{1-\alpha_k} r_k(s, \Lambda(s))\}} \, d\tilde{\mathcal{N}}_k^{\alpha_k}(s, z) \right) \zeta_k; \\ \bar{M}_{n,\bar{X}_n}(t) &= n^{-1/2} \sum_{k=1}^m M_{n,k,\bar{X}_n}(t); \\ \bar{M}_{n,\Lambda}(t) &= n^{-1/2} \sum_{k=1}^m M_{n,k,\Lambda}(t). \end{cases}$$

Observe that the crucial difference between \bar{M}_{n,\bar{X}_n} and $\bar{M}_{n,\Lambda}$ is that the indicator functions appearing in the latter involve deterministic processes

and thus $\overline{M}_{n,\Lambda}$ has independent increments. From (21) and (22), for all n , for any $t \in [0, T]$, we have that

$$(24) \quad U_n(t) = U_n(0) + n^{1/2} \sum_{k=1}^m \left(\int_0^t (r_k(s, \overline{X}_n(s)) - r_k(s, \Lambda(s))) \, ds \right) \zeta_k + \overline{M}_{n, \overline{X}_n}(t).$$

To obtain a formal functional central limit theorem for the sequence $(U_n, n \geq 1)$, we make the following additional assumptions.

Assumption 2 (Diffusion scaling).

(i) *The initial conditions are such that*

$$(25) \quad \mathbf{E} [\|U_n(0) - Y(0)\|] \lesssim n^{-1/2}.$$

(ii) *For all n , all $k \in \llbracket 1, m \rrbracket$ and all $t \in [0, T]$,*

$$(26) \quad n^{1/2} (r_k(t, \overline{X}_n(t)) - r_k(t, \Lambda(t))) = \left(\langle L_k, U_n(t) \rangle_{\mathbf{R}^d} + E_{n,k}(t) \right),$$

where for all $k \in \llbracket 1, m \rrbracket$, $L_k \in \mathbf{R}^d$ and

$$\mathbf{E} \left[\|E_{n,k}\|_{\infty, T} \right] \lesssim n^{-1/2}.$$

Theorem 5.2. *Assume that Assumptions 1 and 2 hold. Then $(U_n, n \geq 1)$ converges at the usual rates to*

$$\Theta_{\mathbf{A}} \left(Y(0) + \sum_{k=1}^m (B_k \circ \gamma_k) \zeta_k \right),$$

where $\Theta_{\mathbf{A}}$ is the map associated by (10) to the matrix

$$(27) \quad \mathbf{A} = (L_1, \dots, L_m) \otimes (\zeta_1, \dots, \zeta_m),$$

$(B_k)_{k \in \llbracket 1, m \rrbracket}$ are independent one dimensional Brownian motions, and for any $k \in \llbracket 1, m \rrbracket$, γ_k is defined in function of the mapping $t \mapsto r_k(t, \Lambda(t))$ through (8).

Proof. In view of (24) and (26), we have for all $n \geq 1$,

$$(28) \quad U_n = \Theta_{\mathbf{A}} \left(U_n(0) + \sum_{k=1}^m \left(\int_0^{\cdot} E_{n,k}(s) \, ds \right) \zeta_k + \overline{M}_{n, \overline{X}_n} \right).$$

But first, according to Assumption 2 we get that

$$(29) \quad \mathbf{E} \left[\left\| \sum_{k=1}^m \left(\int_0^{\cdot} E_{n,k}(s) \, ds \right) \zeta_k \right\|_{\infty, T} \right] \lesssim n^{-1/2}.$$

Second, from the Burkholder Davis Gundy inequality, for any k there exists a constant c_k such that

$$\begin{aligned} \mathbf{E} \left[\left\| n^{-1/2} M_{n,k,\bar{X}_n} - n^{-1/2} M_{n,k,\Lambda} \right\|_{\infty,T} \right] \\ \leq c_k n^{-1/2} \|\zeta_k\| n^{1/2} \mathbf{E} \left[\int_0^T |r_k(t, \bar{X}_n(t)) - r_k(t, \Lambda(t))| dt \right]^{1/2} \\ \leq c_k \|\zeta_k\| \mathbf{E} \left[\int_0^T \|\bar{X}_n(t) - \Lambda(t)\|_{\mathbf{R}^d} dt \right]^{1/2}, \end{aligned}$$

so it follows from (23) together with (25), that

$$(30) \quad \mathbf{E} \left[\left\| n^{-1/2} M_{n,k,\bar{X}_n} - n^{-1/2} M_{n,k,\Lambda} \right\|_{\infty,T} \right] \lesssim n^{-1/4}.$$

On another hand, or any $k \in \llbracket 1, m \rrbracket$, $M_{n,k,\Lambda}$ has the distribution of

$$t \longmapsto \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq r_k(s, \Lambda(s))\}} d\tilde{\mathcal{N}}^n(s, z) \right) \zeta_k,$$

where \mathcal{N}^n is a Poisson measure of intensity $n dt \otimes dz$, and it follows from Corollary 4.7 that the process $n^{-1/2} M_{n,k,\Lambda}$ converges at the usual rates to $(B_k \circ \gamma_k) \zeta_k$. The result then follows from the Lipschitz continuity of $\Theta_{\mathbf{A}}$ in view of the representation (28), the fact that the constant process $\bar{X}_n(0) - \Lambda(0)$ converges to zero at the usual rates in view of (25), and Theorem 3.3, together with (29) and (30). \square

As will be shown below, Theorem 5.2 can be used to show the convergence, at the usual rates, of various classical (properly scaled) CTMC's, towards remarkable diffusion processes, among which, Ornstein-Uhlenbeck processes (see Remark 3) or reflected Brownian motions, see Proposition 5.5.

5.2. The Telegraph process. Let Y_i , $i \in \llbracket 1, n \rrbracket$ be an IID family of CTMC's taking values in $\{0, 1\}$, with transition intensity σ_0 (resp., σ_1) from state 0 to state 1 (resp., from state 1 to state 0), and let X_n be the process defined by

$$X_n(t) = \sum_{i=1}^n Y_i(t), \quad t \geq 0.$$

The processes Y_i , $i \in \llbracket 1, n \rrbracket$ are often called *telegraph processes*, and model various phenomena in finance, physics, networking and biology. The states 0 and 1 can be respectively interpreted as 'Off' and 'On' modes, and so X_n counts the number of telegraph processes in 'On' mode at each time. (It is often itself called, a telegraph process.) Denote by π_0 and π_1 , the common stationary probability of the Y_i 's, $i \in \llbracket 1, n \rrbracket$, namely

$$\pi_0 = \frac{\sigma_1}{\sigma_0 + \sigma_1} \quad \text{and} \quad \pi_1 = \frac{\sigma_0}{\sigma_0 + \sigma_1}.$$

It is immediate to observe that for any n , the limiting distribution $X_n(\infty)$ of the process X_n is binomial of parameters n , π_1 , and that

$$\sqrt{n} \left(\frac{X_n(\infty)}{n} - \pi_1 \right) \Longrightarrow \mathcal{N}(0, \pi_0 \pi_1).$$

At the process level, it is shown in [22] that under suitable assumptions on the initial conditions, for all $T > 0$,

$$U_n = \sqrt{n} \left(\frac{X_n}{n} - \Lambda \right) \Longrightarrow \Theta(Y) \quad \text{in } \mathbb{D}([0, T], \mathbf{R}),$$

where

- For a fixed $\Lambda(0) \in \mathbf{R}_+$,

$$(31) \quad \Lambda(t) = \pi_1 + (\Lambda(0) - \pi_1) \exp(-(\sigma_1 + \sigma_0)t), \quad t \geq 0;$$

- For any f , $\Theta(f)$ denotes the unique solution in $\mathbb{D}([0, T], \mathbf{R})$ of the equation

$$g(t) = -(\sigma_1 + \sigma_0) \int_0^t g(s) \, ds + f(t), \quad t \geq 0;$$

- The process Y is defined by

$$Y(t) = Y(0) + \int_0^t \sqrt{\sigma_0(1 - \Lambda(s))} \, dB_1(s) - \int_0^t \sqrt{\sigma_1 \Lambda(s)} \, dB_2(s), \quad t \geq 0,$$

for B_1 and B_2 , two independent standard Brownian motions.

We have the following result,

Proposition 5.3. *Suppose that condition (i) of Assumption 2 is satisfied. Then the convergence of $(U_n, n \geq 1)$ to $\Theta(Y)$ occurs at the usual rates.*

Proof. Let for all $n \geq 1$, $\bar{X}_n(t) = X_n(t)/n$, $t \geq 0$. Then by the very definition of X_n we have the equality in distribution

$$\begin{aligned} \bar{X}_n(t) \stackrel{(d)}{=} & \bar{X}_n(0) + \int_0^t \int_{\mathbf{R}_+} \mathbf{1}_{\{z \leq n\sigma_0(1 - \bar{X}_n(s^-))\}} \, d\mathcal{N}_1(s, z) \\ & - \int_0^t \int_{\mathbf{R}_+} \mathbf{1}_{\{z \leq n\sigma_1 \bar{X}_n(s^-)\}} \, d\mathcal{N}_2(s, z), \quad t \geq 0, \end{aligned}$$

where \mathcal{N}_1 and \mathcal{N}_2 denote two independent Poisson random measures of common intensity $ds \otimes dz$ on $(\mathbf{R}_+)^2$, representing the overall “up” and “down” jumps, respectively. So \bar{X}_n satisfies the SDE (18) for $d = 1$, $m = 2$, $\alpha_1 = \alpha_2 = 0$, $\zeta_1 = 1$, $\zeta_2 = -1$, and for the mappings $\rho_1 : (t, y) \mapsto \sigma_0(n - y)$ and $\rho_2 : (t, y) \mapsto \sigma_1 y$. It is then clear that Assumption 1 is satisfied for

$$r_1 : (t, x) \mapsto \sigma_0(1 - x) \quad \text{and} \quad r_2 : (t, x) \mapsto \sigma_1 x,$$

which are obviously Lipschitz continuous with respect to their second variable. Plainly, Λ defined by (31) is the unique solution of (22) in the present case. As r_1 and r_2 are linear in their second variable, condition (ii) in Assumption 2 is clearly satisfied for $L_1 = -\sigma_0$, $L_2 = \sigma_1$ and $E_{n,1} \equiv E_{n,2} \equiv 0$ for all n , and so the corresponding operator $\Theta_{\mathbf{A}}$ defined by (10) for \mathbf{A} given in (27) is linear, continuous and therefore Lipschitz continuous. But it follows from (9) that $\Theta = \Theta_{\mathbf{A}}$. We conclude using Theorem 5.2. \square

5.3. Two classical queueing systems. It is a simple matter to show that the single-server queue $M/M/1$ and the infinite server queue $M/M/\infty$, for which the speed of convergence in the functional central limit theorem was already addressed in [4], can be studied in the present framework. The convergence rate obtained hereafter is $n^{1/6}$, slightly slower than that obtained in [4] ($n^{1/2}$), where we used representation methods that are specific to each of these models.

5.3.1. The infinite server queue. We consider an $M/M/\infty$ queue: a potentially unlimited number of servers attend customers that enter the system following a Poisson process of intensity $\lambda > 0$, requesting service times that are exponentially distributed of parameter $\mu > 0$. Let for all $t \geq 0$, $X(t)$ denote the number of customers in the system at time t . It is well known that X is an ergodic Markov process with stationary distribution Poisson of parameter λ/μ . Let us scale this process in space and time, by dividing the number of customers by n , while multiplying the intensity of arrivals by n . Namely, we denote for all $t \geq 0$, $\bar{X}_n(t) = X(t)/n$. Then, it is a classical result (see e.g. [6], Theorem 6.14 in [36], or a measure-valued extension of the result in [18]), that under suitable assumptions on the initial conditions, for all $T > 0$,

$$U_n = \sqrt{n} (\bar{X}_n - \Lambda) \Longrightarrow \Theta(Y) \quad \text{in } \mathbb{D}([0, T], \mathbf{R}),$$

where

- For a fixed $\Lambda(0) \in \mathbf{R}_+$,

$$(32) \quad \Lambda(t) = \frac{\lambda}{\mu} - \left(\Lambda(0) - \frac{\lambda}{\mu} \right) \exp(-\mu t), \quad t \geq 0;$$

- For any stochastic process f , $\Theta(f)$ denotes the unique solution in $\mathbb{D}([0, T], \mathbf{R})$ of the SDE

$$g(t) = -\mu \int_0^t g(s) ds + f(t), \quad t \geq 0;$$

- The process Y is defined by

$$Y(t) = Y(0) + \sqrt{\lambda} B_1(t) - \int_0^t \sqrt{\mu \Lambda(s)} dB_2(s), \quad t \geq 0,$$

for B_1 and B_2 , two independent standard Brownian motions and $Y(0)$, a r.v. independent of everything else,

We have the following result,

Proposition 5.4. *Suppose that condition (i) of Assumption 2 is satisfied. Then the convergence of $(U_n, n \geq 1)$ to $\Theta(Y)$ occurs at the usual rates.*

Proof. Then, for all $n \geq 1$ it is easily checked that the resulting scaled process \bar{X}_n satisfies the SDE

$$\begin{aligned} \bar{X}_n(t) = \bar{X}_n(0) + \frac{1}{n} \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \lambda\}} d\mathcal{N}_1^1(s, z) \\ - \frac{1}{n} \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu \bar{X}_n(s^-)\}} d\mathcal{N}_2^1(s, z), \quad t \geq 0, \end{aligned}$$

where \mathcal{N}_1^1 and \mathcal{N}_2^1 denote two independent random Poisson measures of intensity $n \, ds \otimes dz$ on $(\mathbb{R}_+)^2$. This precisely means that the process \overline{X}_n satisfies (18) for $d = 1$, $m = 2$, $\alpha_1 = \alpha_2 = 1$, $\zeta_1 = 1$, $\zeta_2 = -1$, and for the mappings $\rho_1 : (t, y) \mapsto \lambda$ and $\rho_2 : (t, y) \mapsto \frac{\mu y}{n}$. It is then clear that Assumption 1 is satisfied for $r_1 : (t, y) \mapsto \lambda$ and $r_2 : (t, y) \mapsto \mu y$. Clearly, Λ defined by (32) is the unique solution of the differential equation (22) in the present context. Then, plainly, condition (ii) in Assumption 2 is satisfied for $L_1 = 0$, $L_2 = \mu$ and $E_{n,1} \equiv E_{n,2} \equiv 0$ for all n , and we get that $\Theta_{\mathbf{A}} = \Theta$ in (10). We conclude again using Theorem 5.2. \square

Remark 3. *In the special case $\Lambda(0) = \lambda/\mu$, Λ is the constant process equal to λ/μ and in turn, Y equals $Y_0 + \sqrt{2\lambda}B$ in distribution, for B a standard brownian motion. Then, it is remarkable that the limiting process $\Theta(Y)$ is an Ornstein-Uhlenbeck process. To check this, observe that from (10) and then integrating by parts, for all t we get that*

$$\begin{aligned} \Theta(Y)(t) &= Y(t) - \mu \int_0^t e^{-\mu(t-r)} Y(r) \, dr \\ &= e^{-\mu t} \left\{ Y(t) e^{\mu t} - \mu \int_0^t e^{\mu r} Y(r) \, dr \right\} \\ &= e^{-\mu t} \left\{ Y(0) + \int_0^t e^{\mu r} \, dY(r) \right\} \\ &= Y(0) e^{-\mu t} + \sqrt{2\lambda} \int_0^t e^{-\mu(t-r)} \, dB(r). \end{aligned}$$

This is precisely the Ornstein-Uhlenbeck process that solves the SDE

$$Z(0) = Y(0), \quad dZ(t) = -\mu Z(t) \, dt + \sqrt{2\lambda} \, dB(t),$$

see also the discussion in [36], Theorem 6.14.

5.3.2. The single server queue. In this section we consider a M/M/1 queue: a single-server attends without vacations, customers that enter following a Poisson process of intensity $\lambda > 0$, and the service times are IID with exponential distribution of parameter $\mu > 0$. It is then immediate that the process $(L(t), t \geq 0)$ counting the number of customers in the system at all times, is a birth and death process, that is ergodic if and only if $\lambda/\mu < 1$. This process can be represented as follows: for all $t \geq 0$,

$$\begin{aligned} L(t) &= x + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq \lambda\}} \, d\mathcal{N}_1(s, z) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq \mu\}} \mathbf{1}_{\{L(s^-) > 0\}} \, d\mathcal{N}_2(s, z) \\ &= x + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq \lambda\}} \, d\mathcal{N}_1(s, z) - \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq \mu\}} \, d\mathcal{N}_2(s, z) \\ &\quad + \int_0^t \int_{\mathbb{R}_+} \mathbf{1}_{\{z \leq \mu\}} \mathbf{1}_{\{L(s^-) = 0\}} \, d\mathcal{N}_2(s, z), \end{aligned}$$

where x is the number-in-system at time 0, and \mathcal{N}_1 , \mathcal{N}_2 and \mathcal{N}_3 stand again for three independent random Poisson measures of intensity $ds \otimes dz$ on $(\mathbb{R}_+)^2$. The standard Law-of-Large-Numbers scaling of this process is performed by multiplying the arrival and service intensities by a factor n , while

increasing the number of customers in the initial state by the same multiplicative factor, and dividing the number of customers in the system at any time by n : equivalently, for all $t \geq 0$ we set

$$(33) \quad \bar{L}_n(t) = x + \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \lambda\}} d\mathcal{N}_1^1(s, z) - \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \mu\}} d\mathcal{N}_2^1(s, z) \\ + \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \mu\}} \mathbf{1}_{\{\bar{L}_n(s^-) = 0\}} d\mathcal{N}_2^1(s, z),$$

where \mathcal{N}_1^1 and \mathcal{N}_2^1 denote two Poisson random measures of intensity $n ds \otimes dz$. The following result completes the classical diffusion limit presented e.g. in Proposition 5.16 in [36], and Theorem 1 in [4],

Proposition 5.5. *Let*

$$(34) \quad \Lambda(t) = (x + (\lambda - \mu)t)^+, \quad t \geq 0,$$

and Z be the process defined by

$$Z(t) = Z(0) + \sqrt{\lambda + \mu} B(t), \quad t \geq 0,$$

for B a standard Brownian motion and $Z(0)$, a r.v. independent of everything else. Then the following convergence holds at the usual rate,

$$U_n = \sqrt{n} (\bar{L}_n - \Lambda) \implies \text{Sko}(Z),$$

where the mapping Sko is defined by (11).

Proof. Define the processes \bar{X}_n and \bar{R}_n as follows: for all $t \geq 0$,

$$\bar{X}_n(t) = x + \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \lambda\}} d\mathcal{N}_1^1(s, z) - \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \mu\}} d\mathcal{N}_2^1(s, z); \\ \bar{R}_n(t) = \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \mu\}} \mathbf{1}_{\{\bar{L}_n(s^-) = 0\}} d\mathcal{N}_2^1(s, z).$$

Then in view of (33) we easily get that for all t ,

$$\begin{cases} \bar{X}_n(t) = \bar{L}_n(t) + \bar{R}_n(t), \\ \int_0^t \bar{X}_n(s) d\bar{R}_n(s) = 0, \end{cases}$$

and thus we get that $\bar{L}_n = \text{Sko}(\bar{X}_n)$, see (11). It is then immediate that both Theorems 5.1 and 5.2 can be applied to the sequence of processes $\{\bar{X}_n\}$ for $d = 1$, $m = 2$, $\alpha_1 = \alpha_2 = 1$, $\zeta_1 = 1$, $\zeta_2 = -1$ and for all (t, x) , $r_1(t, x) = \lambda$ and $r_2(t, x) = \mu$. We obtain (for $\Theta_{\mathbf{A}} := \mathbf{Id}$), that the convergence

$$V_n := \sqrt{n} (\bar{X}_n - \Gamma) \implies Z$$

occurs at the usual rate, where for all $t \geq 0$, we define

$$\Gamma(t) = x + (\lambda - \mu)t,$$

and observe the equality in distribution

$$Z(t) = Z(0) + \sqrt{\lambda} B_1(t) - \sqrt{\mu} B_2(t),$$

for B_1 and B_2 , two independent standard Brownian motions. It is immediate that $\Lambda = \text{Sko}(\Gamma)$, so the result follows from the Lipschitz continuity of the mapping Sko , proven in Theorem 3.4. \square

5.4. SIR epidemics. We consider a population of constant size n , in which individuals can go through three states, susceptible, infectious and then removed. The duration of any infection follows an exponential distribution of parameter γ and for each couple infectious/susceptible, a contagion occurs from the former at an exponential rate λ . All the involved r.v.'s are assumed independent. At any time $t \geq 0$, we let $S^n(t), I^n(t)$ and $R^n(t)$ denote respectively the number of Susceptible, Infected and Recovered individuals, and let $X^n(t) = \begin{pmatrix} S^n(t) \\ I^n(t) \end{pmatrix}$. The processes are scaled by defining for all $t \geq 0$, $\bar{S}_n(t) = S^n(t)/n$, $\bar{I}_n(t) = I^n(t)/n$ and $\bar{R}^n(t) = R^n(t)/n$, which represent respectively the proportions of Susceptible, Infected and Recovered individuals in the whole population at time t . We also let $\bar{X}^n(t) = \begin{pmatrix} \bar{S}^n(t) \\ \bar{I}^n(t) \end{pmatrix}$, $t \geq 0$.

A large-graph limit and a functional central limit theorem for the process \bar{X} are given in Chapter 2 of [10], together with similar results regarding the related SEIR, SIRS and SIS models. Observe that a hydrodynamic limit for a SIR process propagating on a heterogeneous population (meaning that a - non necessarily complete - graph connects susceptible to infectious individuals) is provided in [17], completing the result in [41]. The following result makes precise the speed of convergence in the functional CLT for the complete-graph case, given in [10]. We are confident that similar results hold for the other related models addressed in Chapter 2 of [10], however we only consider here the SIR case for brevity,

Proposition 5.6. *Let $\Lambda : t \mapsto \begin{pmatrix} s(t) \\ i(t) \end{pmatrix}$ be the unique solution of the system of ODE's*

$$(35) \quad \begin{cases} s'(t) &= -\lambda s(t)i(t) \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t), \end{cases} \quad t \geq 0.$$

Let $Y(0)$ be a r.v. independent of everything else, and $\Theta(Y)$ denote the unique solution in \mathcal{C}_T of the following SDE of unknown $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$,

$$g(t) = \lambda \int_0^t (i(u)g_1(u) + s(u)g_2(u)) du \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \gamma \int_0^t g_2(u) du \begin{pmatrix} 0 \\ -1 \end{pmatrix} + Y(t), \quad t \geq 0,$$

where for all $t \geq 0$,

$$Y(t) = Y(0) + \int_0^t \sqrt{\lambda s(u)i(u)} dB_1(u) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \int_0^t \sqrt{\gamma i(u)} dB_2(u) \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

for B_1 and B_2 , two independent standard Brownian motions. Then, if assertion (i) of Assumption 2 is satisfied, the following convergence holds at the usual rates,

$$U_n = \sqrt{n} (\bar{X}_n - \Lambda) \Longrightarrow \Theta(Y).$$

Proof. By the very definition of the SIR dynamics, for any n the process \bar{X}^n admits the following representation: for all $t \geq 0$,

$$\begin{aligned} \bar{X}_n(t) = \bar{X}_n(0) &+ \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \lambda n \bar{S}^n(u^-) \bar{I}^n(u^-)\}} d\mathcal{N}_1(u, z) \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &+ \frac{1}{n} \int_0^t \int_{\mathbb{R}^+} \mathbf{1}_{\{z \leq \gamma n \bar{I}^n(u^-)\}} d\mathcal{N}_2(u, z) \begin{pmatrix} 0 \\ -1 \end{pmatrix}, \end{aligned}$$

for \mathcal{N}_1 and \mathcal{N}_2 , two Poisson random measures of unit intensity, so we fall again into the settings of Section 5.1 for $d = 2$, $m = 2$, $\alpha_1 = \alpha_2 = 0$, $\zeta_1 = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and $\zeta_2 = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$. It is then clear that Assumption 1 is satisfied for the mappings

$$\begin{aligned} r_1 : \mathbf{R}_+ \times [0, 1]^2 &\longrightarrow \mathbf{R}, \\ &\left(t, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \longmapsto \lambda y_1 y_2; \\ r_2 : \mathbf{R}_+ \times [0, 1]^2 &\longrightarrow \mathbf{R}, \\ &\left(t, \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \longmapsto \gamma y_2. \end{aligned}$$

Indeed, to check that r_1 is Lipschitz continuous on its second variable, just observe that for all $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in [0, 1]^2$, $y' = \begin{pmatrix} y'_1 \\ y'_2 \end{pmatrix} \in [0, 1]^2$ and for all t ,

$$|r_1(t, y') - r_1(t, y)| = \lambda |y'_2(y'_1 - y_1) + y_1(y'_2 - y_2)| \leq \lambda \|y' - y\|,$$

so Theorem 5.1 is satisfied. Now, define again for all t and n , $U^n(t) = n^{1/2} (\bar{X}_n(t) - \Lambda(t))$. To check that assumption 2 holds, let us also observe that for all y, y', t as above, we also have that

$$r_1(t, y') - r_1(t, y) = \lambda (y_2(y'_1 - y_1) + y_1(y'_2 - y_2) + (y'_2 - y_2)(y'_1 - y_1)),$$

entailing that for all n and t ,

$$n^{1/2} (r_1(t, \bar{X}_n(t)) - r_1(t, \Lambda(t))) = \langle L_1(t), U_n(t) \rangle_{\mathbf{R}^2} + E_{1,n}(t),$$

for

$$L_1(t) = \lambda \begin{pmatrix} i(t) \\ s(t) \end{pmatrix} \quad \text{and} \quad E_{1,n}(t) = \lambda (\bar{S}^n(t) - s(t)) (\bar{I}^n(t) - i(t)).$$

Thus, observing that both $|\bar{S}^n(t) - s(t)|$ and $|\bar{I}^n(t) - i(t)|$ are less than one for all t , we obtain that

$$\begin{aligned} \mathbf{E} [\| E_{1,n} \|_{T, \infty}] &\leq \lambda \mathbf{E} \left[\sup_{0 \leq t \leq T} (|\bar{S}^n(t) - s(t)| + |\bar{I}^n(t) - i(t)|) \right] \\ &\lesssim \lambda \mathbf{E} \left[\sup_{0 \leq t \leq T} \| \bar{X}^n(t) - \Lambda(t) \| \right] \\ &\lesssim n^{-1/2}, \end{aligned}$$

from (23) together with assertion (i) of Assumption 2. Likewise, it is immediate that r_2 is Lipschitz continuous in its second variable, and we get for

all n and t that

$$n^{1/2} (r_2(t, \bar{X}_n(t)) - r_2(t, \Lambda(t))) = \langle L_2(t), U_n(t) \rangle_{\mathbf{R}^2} \quad \text{for } L_2(t) := \begin{pmatrix} 0 \\ \gamma \end{pmatrix},$$

so assumptions 2 hold, and we apply again Theorem 5.2. \square

5.5. The Moran model. In this section we consider a biological model known as the Moran model : in a population of size n , each individual bears a gene liable to take two forms : A and B . Each individual has one single parent and its child inherits the genetic form of its parent. To each couple of individuals is associated an exponential clock of unit rate, and each time the clock of a given couple rings, one element of the couple, drawn uniformly at random, dies, while the other one gives birth to another individual bearing the same gene. In addition, every gene of type A mutates independently to type B at rate ν_1 and every gene of type B mutates independently to type A at rate ν_2 . For all $t \geq 0$, we let $X_n(t)$ denote the number of individuals bearing gene A in the population at time t . The process X_n is scaled by dividing by n the exponential rates, together with the number of individuals, so that $\bar{X}_n(t) = X_n(t)/n$ represents the proportion of individuals carrying the gene of type A at time t . A functional Stein method is applied to this model in [32]. The following result is based on an alternative representation of the process X_n ,

Proposition 5.7. *Let Λ be the solution in $\mathbb{C}([0, T])$ of the integral equation*

$$\Lambda(t) = \Lambda(0) + \int_0^t (\nu_2 - (\nu_1 + \nu_2)\Lambda(s)) \, ds, \quad t \geq 0.$$

with $\Lambda(0)$ such that (i) of assumption 2 is satisfied. Then the following convergence holds at the usual rate,

$$U_n = \sqrt{n} (\bar{X}_n - \Lambda) \implies \Theta(Y),$$

where for any process f , $\Theta(f)$ is the only solution of the SDE

$$y(t) = (\nu_1 + \nu_2) \int_0^t y(s) \, ds + f(t), \quad t \geq 0,$$

and where

$$Y(t) = Y(0) + \int_0^t \sqrt{2\Lambda(s)(1 - \Lambda(s))} \, dB(s), \quad t \geq 0,$$

for B a standard Brownian motion.

Proof. For all n , at each given time s there are $X_n(s)(n - X_n(s))$ couples gathering an ‘ A -individual’ and a ‘ B -individual’. Thus for all $t \geq 0$, $X_n(t)$

can be represented as

$$\begin{aligned}
X_n(t) &= X_n(0) + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq X_n(s)(n-X_n(s))\}} d\mathcal{N}_1(s, z) \\
&\quad - \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq X_n(s)(n-X_n(s))\}} d\mathcal{N}_2(s, z) \\
&\quad + \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq \nu_2(n-X_n(s))\}} d\mathcal{N}_3(s, z) \\
&\quad - \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq \nu_1 X_n(s)\}} d\mathcal{N}_4(s, z),
\end{aligned}$$

where $\mathcal{N}_i, i \in \llbracket 1, 4 \rrbracket$, denote four independent Poisson random measures of unit intensity. Consequently we get that

$$\begin{aligned}
\bar{X}_n(t) &= \bar{X}_n(0) + \frac{1}{n} \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq n\bar{X}_n(s)(n-n\bar{X}_n(s))\}} d\mathcal{N}_1^{-1}(s, z) \\
&\quad - \frac{1}{n} \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq n\bar{X}_n(s)(n-n\bar{X}_n(s))\}} d\mathcal{N}_2^{-1}(s, z) + \frac{1}{n} \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq \nu_2(n-n\bar{X}_n(s))\}} d\mathcal{N}_3^{-1}(s, z) \\
&\quad - \frac{1}{n} \int_0^t \int_{\mathbb{R}} \mathbf{1}_{\{z \leq \nu_1 n\bar{X}_n(s)\}} d\mathcal{N}_4^{-1}(s, z),
\end{aligned}$$

where $\mathcal{N}_i^{-1}, i \in \llbracket 1, 4 \rrbracket$, are four independent Poisson measures of intensity $n^{-1} ds \otimes dz$. We fall back into the settings of Section 5.1 for $d = 1, m = 4, \zeta_1 = \zeta_3 = 1, \zeta_2 = \zeta_4 = -1$ and $\alpha_1 = \dots = \alpha_4 = -1$. Assumption 1 holds for

$$r_1(t, x) = r_2(t, x) = x(1 - x), \quad t \geq 0, \quad x \in [0, 1],$$

which are Lipschitz continuous in their second variable, as for all $x, y \in [0, 1]$ and all $t \geq 0$ we get

$$|x(1 - x) - y(1 - y)| \leq 3|x - y|.$$

Assumption 1 also obviously holds for

$$r_3(t, x) = \nu_2(1 - x) \text{ and } r_4(t, x) = \nu_1 x \quad t \geq 0, \quad x \in [0, 1].$$

On another hand, an immediate computation shows that for all $x, y \in [0, 1]$,

$$y(1 - y) - x(1 - x) = (y - x)(1 - 2x) - (y - x)^2,$$

so (26) holds for

$$\begin{cases} L_1(t) = L_2(t) = (1 - 2\Lambda(t)), & t \geq 0, \\ L_3(t) = \nu_2, L_4(t) = -\nu_1, & t \geq 0, \\ E_{n,1}(t) = E_{n,2}(t) = \sqrt{n}(\bar{X}_n(t) - \Lambda(t))^2, & t \geq 0, \\ E_{n,3}(t) = E_{n,4}(t) = 0, & t \geq 0, \end{cases}$$

and from (23) we obtain that

$$\mathbf{E} [\| E_{i,n} \|_{T,\infty}] \lesssim n^{-1/2}, \quad i \in \{1, 2\}.$$

So Assumption 2 holds, and we apply again Theorem 5.2. \square

6. LIMIT THEOREM FOR HAWKES PROCESSES

In this section we turn to the cases of Hawkes processes. We let $\varphi : \mathbf{R}^+ \rightarrow \mathbf{R}^+$ be an integrable function, such that

$$(36) \quad \kappa := \int_0^\infty \varphi(t) dt < 1 \text{ and } \int_0^\infty t^{1/2} \varphi(t) dt < \infty.$$

We denote by Φ the first primitive of φ :

$$\Phi(t) = \int_0^t \varphi(s) ds, \quad t \geq 0.$$

We also need to consider the iterated convolution products of φ with itself:

$$\varphi^{(1)} = \varphi \text{ and } \varphi^{(k)} = \varphi * \varphi^{(k-1)}.$$

The function

$$\psi = \sum_{k=1}^{\infty} \varphi^{(k)}$$

plays an important role in the representation of the process to be defined hereafter. Note that

$$\int_0^\infty \psi(t) dt = \sum_{k \geq 1} \kappa^k = \frac{\kappa}{1 - \kappa}.$$

For $\mu > 0$, according to [27, 34], there exists a point process N (unique in distribution) such that N admits the intensity

$$t \longmapsto \mu + \int_0^t \varphi(t-s) dN(s).$$

In [1], it is proved that

$$(37) \quad \mathbf{E} \left[\sup_{v \in [0,1]} \left| \frac{1}{n} N(nv) - \rho v \right|^2 \right] \lesssim \frac{1}{n},$$

where $\rho = (1 - \kappa)^{-1} \mu$. Furthermore, it is shown that

$$\sqrt{n} \left(\frac{1}{n} N(nv) - \rho v \right) \xrightarrow[n \rightarrow \infty]{\text{dist. in } \mathbb{D}} B \left(\frac{\rho}{(1 - \kappa)^2} v \right).$$

Our goal is to assess the rate of this convergence. Note that a similar problem, at fixed time but for more general situation, was tackled by a different method in [39]. To that end, we use a particular construction of N based on a Poisson measure \mathcal{N} of intensity measure $ds \otimes dz$. For all $t \geq 0$, we know from [34] and references therein, that we can write

$$N(t) = \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + \int_0^s \varphi(s-u) dN(u)\}} d\mathcal{N}(s, z).$$

Denote also

$$W(t) = \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + \int_0^s \varphi(s-u) dN(u)\}} (d\mathcal{N}(s, z) - ds dz),$$

the corresponding compensated integral, so that W is a local martingale with respect to the filtration induced by \mathcal{N} :

$$\mathcal{F}_t = \sigma \{ \mathcal{N}([0, s] \times A), 0 \leq s \leq t, A \in \mathcal{B}(\mathbf{R}^+) \}, t \geq 0.$$

For a process Z , we denote for all n ,

$$Z^{(n)}(t) = Z(nt), \quad \bar{Z}^{(n)}(t) = \frac{1}{\sqrt{n}}Z^{(n)}(t), \quad \tilde{Z}^{(n)}(t) = \frac{1}{n}Z^{(n)}(t), \quad t \geq 0.$$

From [30, Theorem 10.27], we also know that for $t \geq 0$,

$$W^{(n)}(t) = \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\left\{z \leq \mu+n \int_0^s \varphi(ns - nu) d\tilde{N}^{(n)}(u)\right\}} \left(d\mathcal{N}^n(s, z) - n ds dz \right),$$

where \mathcal{N}^n is a Poisson measure of intensity $n dt \otimes dz$. These considerations result in the following lemma,

Lemma 6.1. *For all $n \geq 1$, we have*

$$\mathfrak{K}_{\mathcal{C}_1} \left(\Xi_n(\bar{W}^{(n)}), \Xi_n(B \circ \gamma) \right) \lesssim n^{-1/6} \log(n),$$

where $\gamma(t) = \rho t$, $t \geq 0$.

Proof. Fix $n \geq 1$. According to (37), $\tilde{N}^{(n)}$ is asymptotically close to ρ , hence we consider the process

$$\begin{aligned} R^{(n)}(t) &= \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\left\{z \leq \mu+n \int_0^s \varphi(ns - nu) \rho du\right\}} \left(d\mathcal{N}^n(s, z) - n ds dz \right) \\ &= \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\left\{z \leq \mu+ \int_0^{ns} \varphi(v) \rho dv\right\}} d\tilde{\mathcal{N}}^n(s, z), \end{aligned}$$

where $\tilde{\mathcal{N}}^n$ is the compensated Poisson measure of \mathcal{N}^n . We can apply Theorem 4.7 to

$$\begin{aligned} \hat{R}^{(n)}(t) &= \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\left\{z \leq \mu+ \int_0^{+\infty} \varphi(v) \rho dv\right\}} d\tilde{\mathcal{N}}^n(s, z) \\ &= \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \rho\}} d\tilde{\mathcal{N}}^n(s, z), \end{aligned}$$

in view of the definition of κ and ρ , implying that

$$(38) \quad \mathfrak{K}_{\mathcal{C}_1} \left(\Xi_n \left(\frac{1}{\sqrt{n}} \hat{R}^{(n)} \right), \Xi_n(B \circ \gamma) \right) \lesssim n^{-1/6} \log(n),$$

We then control the distance between $R^{(n)}$ and $\hat{R}^{(n)}$. For this, observe that the Markov inequality implies that

$$\int_{ns}^{\infty} \varphi(v) dv \leq \frac{1}{\sqrt{ns}} \int_0^{\infty} \sqrt{v} \varphi(v) dv,$$

and hence

$$\int_0^1 \int_{ns}^{\infty} \varphi(v) dv \lesssim \frac{1}{\sqrt{n}}.$$

According to the Burkholder-Davis-Gundy inequality, we obtain that

$$\begin{aligned} \mathbf{E} \left[\sup_{t \in [0,1]} \frac{|\hat{R}^{(n)}(t) - R^{(n)}(t)|}{\sqrt{n}} \right]^2 &\leq \mathbf{E} \left[\frac{|\hat{R}^{(n)}(1) - R^{(n)}(1)|^2}{n} \right] \\ &\lesssim \rho \int_0^1 \int_{ns}^\infty \varphi(v) \, dv \, ds \\ &\lesssim \frac{1}{\sqrt{n}}. \end{aligned}$$

We now want to estimate the error made by considering $\bar{R}^{(n)}$ instead of $\bar{W}^{(n)}$. Let for all $t \geq 0$,

$$\begin{aligned} r(t) &= \mu + n \int_0^t \varphi(nt - ns) \, d\tilde{N}^{(n)}(s), \\ \check{r}(t) &= \mu + n \int_0^t \varphi(nt - ns) \rho \, ds. \end{aligned}$$

We have

$$\begin{aligned} &\|\Xi_n(\bar{R}^{(n)}) - \Xi_n(\bar{W}^{(n)})\|_{\infty,1} \\ &\leq \sup_{i \in \llbracket 0, n-1 \rrbracket} \left| (\bar{R}^{(n)} - \bar{W}^{(n)}) \left(\frac{i+1}{n} \right) - (\bar{R}^{(n)} - \bar{W}^{(n)}) \left(\frac{i}{n} \right) \right| \\ &\quad + \sup_{i \in \llbracket 0, n-1 \rrbracket} \left| (\bar{R}^{(n)} - \bar{W}^{(n)}) \left(\frac{i}{n} \right) \right| \\ &\lesssim \frac{1}{\sqrt{n}} \sup_{i \in \llbracket 0, n-1 \rrbracket} \left| \int_0^{i/n} \int_{\mathbf{R}^+} (\mathbf{1}_{\{z \leq r(s)\}} - \mathbf{1}_{\{z \leq \check{r}(s)\}}) (d\mathcal{N}^n(s, z) - n \, ds \, dz) \right|. \end{aligned}$$

Apply the Doob inequality to discrete martingale

$$k \mapsto \left(\int_0^{k/n} \int_{\mathbf{R}^+} (\mathbf{1}_{\{z \leq r(s)\}} - \mathbf{1}_{\{z \leq \check{r}(s)\}}) (d\mathcal{N}^n(s, z) - n \, ds \, dz), \, i \in \llbracket 0, n-1 \rrbracket \right),$$

to obtain that

$$\begin{aligned} \mathbf{E} \left[\|\Xi_n(\bar{R}^{(n)}) - \Xi_n(\bar{W}^{(n)})\|_{\infty,1} \right] &\lesssim \frac{1}{\sqrt{n}} \mathbf{E} \left[\int_0^1 |r(s) - \check{r}(s)| n \, ds \right]^{1/2} \\ &= \mathbf{E} \left[\int_0^1 |r(s) - \check{r}(s)| \, ds \right]^{1/2}. \end{aligned}$$

With the particular expression of r and \check{r} , we get that

$$\begin{aligned}
& \mathbf{E} \left[\|\Xi_n(\bar{R}^{(n)}) - \Xi_n(\bar{W}^{(n)})\|_{\infty,1} \right] \\
& \leq \frac{1}{\sqrt{n}} \mathbf{E} \left[\left\| \int_0^1 n \left| \int_0^s \varphi(ns - nu) (d\tilde{N}^{(n)}(u) - \rho \, du) \right| \, ds \right\| \right]^{1/2} \\
& \leq \frac{1}{\sqrt{n}} \mathbf{E} \left[\int_0^1 n \int_0^s \varphi(ns - nu) \, d|\tilde{N}^{(n)}(u) - \rho \, du| \, ds \right]^{1/2} \\
& \leq \frac{1}{\sqrt{n}} \mathbf{E} \left[\int_0^1 \Phi(n(1-s)) \, d|\tilde{N}^{(n)}(s) - \rho s| \right]^{1/2}
\end{aligned}$$

where the last two integrals are understood as integrals with respect to the total variation of the process $s \mapsto \tilde{N}^{(n)}(s) - \rho s$. By integration by parts, we then obtain

$$\begin{aligned}
& \frac{1}{\sqrt{n}} \mathbf{E} \left[\int_0^1 \Phi(n(1-s)) \, d|\tilde{N}^{(n)}(s) - \rho s| \right]^{1/2} \\
& = \mathbf{E} \left[\int_0^1 |\tilde{N}^{(n)}(s) - \rho s| \varphi(n(1-s)) \, ds \right]^{1/2}.
\end{aligned}$$

Thus, we have that

$$\mathbf{E} \left[\|\Xi_n(\bar{R}^{(n)}) - \Xi_n(\bar{W}^{(n)})\|_{\infty,1} \right] \lesssim_{\kappa} \mathbf{E} \left[\|\tilde{N}^{(n)} - \rho\|_{\infty,1} \right]^{1/2} \frac{1}{\sqrt{n}},$$

where $\kappa = \int_0^\infty \varphi(u) \, du$. Note that since \tilde{N} is an non decreasing process,

$$\|\tilde{N}^{(n)} - \rho\|_{\infty,1} \leq \tilde{N}^{(n)}(1) + \rho = \tilde{N}^{(n)}(1) - \rho + 2\rho.$$

In view of (37), we get

$$\mathbf{E} \left[\|\Xi_n(\bar{R}^{(n)}) - \Xi_n(\bar{W}^{(n)})\|_{\infty,1} \right] \lesssim \frac{1}{\sqrt{n}}.$$

Thus, we can substitute $\bar{W}^{(n)}$ to $\hat{R}^{(n)}/\sqrt{n}$ in (38), and the result follows. \square

The convergence of $\bar{N}^{(n)}$ follows from the representation formula established in [1]. Let

$$X^{(n)}(t) = N^{(n)}(t) - \mathbf{E} \left[N^{(n)}(t) \right], \quad t \geq 0.$$

According to [1, Lemma 4], we have for all $t \geq 0$,

$$(39) \quad \bar{X}^{(n)}(t) = \bar{W}^{(n)}(t) + \int_0^t n\psi(ns) \bar{W}^{(n)}(t-s) \, ds.$$

The analysis of this identity is a bit tricky because the two terms of the integrand do depend on n , so we cannot invoke the Lipschitz continuity of a well chosen map.

Theorem 6.2. *For all $n \geq 1$ we have that*

$$\mathfrak{R}_{\mathcal{C}_1} \left(\Xi_n(\bar{X}^{(n)}), \Xi_n(B \circ \zeta) \right) \lesssim n^{-1/6} \log(n)$$

where for all $t \geq 0$,

$$\zeta(t) = \frac{1}{(1 - \kappa)^2} \gamma(t), \quad t \geq 0,$$

for $\gamma(t) = \rho t$.

Remark 4. We are here limited to the Kantorovitch-Rubinstein distance because we are going to use the convergence of first moments induced by the \mathfrak{K} -topology, see (12), which is not valid for the other distances investigated above.

The proof follows closely the lines of the proof of [1].

Proof of Theorem 6.2. Fix $n \geq 1$, and take for granted that

$$(40) \quad I_n := \mathbf{E} \left[\sup_{t \in [0,1]} \left| \int_0^t n \psi(ns) \overline{W}^{(n)}(t-s) \, ds - \frac{\kappa}{1-\kappa} \overline{W}^{(n)}(v) \right| \right] \lesssim n^{-1/4}.$$

Then according to (39),

$$\mathfrak{K} \left(\Xi_n \left(\overline{X}^{(n)} \right), \Xi_n (B \circ \zeta) \right) \lesssim \mathfrak{K} \left(\frac{1}{1-\kappa} \Xi_n (\overline{W}^{(n)}), \Xi_n (B \circ \zeta) \right).$$

The result then follows from Lemma 6.1.

We now establish (40). According to the decomposition given in [1], for any $0 < \delta < \eta$, we have

$$I_n \leq \mathbf{E} \left[\|\overline{W}^{(n)}\|_{\infty,1} \right] \int_{\delta n}^{\infty} \psi(t) \, dt + \mathbf{E} \left[\alpha_{\eta}(\overline{W}^{(n)}) \right] \int_0^{\infty} \psi(t) \, dt.$$

Since the \mathfrak{K} -convergence implies the convergence of first order moments,

$$\sup_n \mathbf{E} \left[\|\overline{W}^{(n)}\|_{\infty,1} \right] \lesssim \mathbf{E} [\|B \circ \gamma\|_{\infty,1}].$$

In view of Theorem 3.4, we also have that

$$\sup_n \mathbf{E} \left[\alpha_{\eta}(\overline{W}^{(n)}) \right] \lesssim_{\epsilon} \mathbf{E} [\alpha_{\eta}(B \circ \gamma)] \lesssim \eta^{1/2-\epsilon}$$

for any $\epsilon > 0$. The integrability conditions (36) on φ and Lemma 5 of [1] show that $(t \mapsto \sqrt{t}\psi(t))$ belongs to $L^1(\mathbf{R}^+)$ hence, the Markov inequality entails that

$$\int_{\delta n}^{\infty} \psi(t) \, dt \lesssim \frac{1}{\sqrt{\delta n}}.$$

If we choose $\eta = n^{-1/2}$ and $\delta = \eta/2$, we get

$$I_n \lesssim n^{-1/4}.$$

The proof is thus complete. \square

7. AUXILIARY RESULT

Proposition 7.1. Let $(X_i, i = 1, \dots, n)$ be an array of Poisson random variables of parameter ν . Then we have that

$$\mathbf{E} \left[\max_{i=1, \dots, n} X_i \right] \leq \frac{\log(ne^{-\nu})}{\log\left(\frac{\log(ne^{-\nu})}{\nu e}\right) - \log\log\left(\frac{\log(ne^{-\nu})}{\nu e}\right)},$$

for all $n \geq e^{\nu(1+e^2)}$.

Proof. Let $Z_i = X_i - \nu$ for all $i = 1, \dots, n$. By a straightforward calculation, for all i and all $u \in \mathbf{R}$,

$$\mathbf{E} [e^{uZ_i}] = e^{-u\nu} \sum_{k=0}^{\infty} e^{uk} e^{-\nu} \frac{\nu^k}{k!} = \exp(\nu(e^u - u - 1)).$$

By Jensen's inequality, we obtain that

$$\begin{aligned} \exp\left(u \mathbf{E} \left[\max_{i=1, \dots, n} Z_i \right]\right) &\leq \mathbf{E} \left[\exp\left(u \max_{i=1, \dots, n} Z_i\right) \right] \\ &= \mathbf{E} \left[\max_{i=1, \dots, n} \exp(uZ_i) \right] \\ &\leq \mathbf{E} \left[\sum_{i=1}^n \exp(uZ_i) \right] \\ &= n \mathbf{E} [\exp(uZ_1)] = n \exp(\nu(e^u - u - 1)). \end{aligned}$$

Taking the log, for any u in $(0, +\infty)$,

$$u \mathbf{E} \left[\max_{i=1, \dots, n} Z_i \right] - \nu(e^u - u - 1) \leq \log n,$$

so that

$$\mathbf{E} \left[\max_{i=1, \dots, n} Z_i \right] \leq \inf_{u \in \mathbf{R}^+} \left(\frac{\log n + \nu(e^u - u - 1)}{u} \right).$$

By differentiation, it is easy to check that the infimum above is reached at $u^* \in (0, +\infty)$ satisfying

$$(41) \quad \nu(u^* - 1)e^{u^*} + \nu = \log n,$$

implying in turn that the infimum equals

$$(42) \quad \nu e^{u^*} - \nu.$$

Now, let W be the Lambert function, defined by the equation $W(x)e^{W(x)} = x$ for all $x \in [-1/e, \infty)$. Then, as (41) amounts to writing that

$$(u^* - 1)e^{u^* - 1} = \log(ne^{-\nu})/(\nu e),$$

we obtain that

$$u^* - 1 = W(a),$$

for $a = \log(ne^{-\nu})/(\nu e)$. Plugging this into (42), and by the very definition of $W(a)$, we obtain that the infimum equals

$$\nu e^{1+W(a)} - \nu = \nu e e^{W(a)} - \nu = \frac{\nu e a}{W(a)} - \nu,$$

which implies that

$$\mathbf{E} \left[\max_{i=1, \dots, n} X_i \right] = \mathbf{E} \left[\max_{i=1, \dots, n} Z_i \right] + \nu \leq \frac{\nu e a}{W(a)}.$$

We conclude by observing that $W(z) = \log z - \log(W(z))$ for all $z > 0$, implying in turn that $W(z) \geq \log z - \log \log z$ for all $z > e$. Therefore, for all $n \geq e^{\nu(1+e^2)}$ we get that $W(a) \geq \log a - \log \log a$, which concludes the proof. \square

REFERENCES

- [1] E. Bacry, S. Delattre, M. Hoffmann, and J.F. Muzy, *Some limit theorems for hawkes processes and application to financial statistics*, Stochastic Processes and their Applications **123** (2013), no. 7, 2475 – 2499.
- [2] A. D. Barbour, *Stein’s method for diffusion approximations*, Probability Theory and Related Fields **84** (1990), no. 3, 297–322.
- [3] A. D. Barbour, N. Ross, and G. Zheng, *Stein’s method, gaussian processes and palm measures, with applications to queueing*, 2021.
- [4] E. Besançon, L. Decreusefond, and P. Moyal, *Stein’s method for diffusive limit of Markov processes*, Queueing Systems **95** (2020), 173–201.
- [5] T. Bonis, *Stein’s method for normal approximation in Wasserstein distances with application to the multivariate central limit theorem*, Probab. Theory Related Fields **178** (2020), no. 3-4, 827–860.
- [6] AA Borovkov, *On limit laws for service processes in multi-channel systems*, Siberian Mathematical Journal **8** (1967), no. 5, 746–763.
- [7] Solesne Bourguin and Simon Campese, *Approximation of hilbert-valued gaussians on dirichlet structures*, Electron. J. Probab. **25** (2020), 30 pp.
- [8] Anton Braverman and J. G. Dai, *Stein’s method for steady-state diffusion approximations of $m/ph/n + m$ systems*, The Annals of Applied Probability **27** (2017), no. 1, 550–581.
- [9] Anton Braverman, J. G. Dai, and Jiekun Feng, *Stein’s method for steady-state diffusion approximations: an introduction through the Erlang-A and Erlang-C models*, Stochastic Systems **6** (2016), no. 2, 301–366.
- [10] T. Britton, E. Pardoux, F. Ball, C. Laredo, D. Sirl, and V.C. Tran, *Stochastic epidemic models with inference*, Springer, 2019.
- [11] P. Brémaud, *Point processes and queues, martingale dynamics*, Springer-Verlag, New York, 1981.
- [12] P. M. Chaikin and T. C. Lubensky, *Principles of condensed matter physics*, Cambridge University Press, 1995.
- [13] L. Coutin and L. Decreusefond, *Stein’s method for Brownian approximations*, Communications on Stochastic Analysis **7** (2013), no. 3, 349–372.
- [14] L. Coutin and L. Decreusefond, *Donsker’s theorem in Wasserstein-1 distance*, Electron. Commun. Probab. **25** (2020), 1–13.
- [15] L. Coutin and L. Decreusefond, *Stein’s method for rough paths*, Potential Analysis **50** (2020), 387–406.
- [16] L. Decreusefond, *Selected topics in Malliavin calculus*, Bocconi Springer Ser., vol. 10, 2022 (English).
- [17] L. Decreusefond, J.-S. Dhersin, P. Moyal, and V. C. Tran, *Large graph limit for an sir process in random network with heterogeneous connectivity*, Ann. Appl. Probab. **22** (2012), no. 2, 541–575.
- [18] L. Decreusefond and P. Moyal, *A functional central limit theorem for the $m/gi/\infty$ queue*, The Annals of Applied Probability **18** (2008), no. 6, 2156–2178.
- [19] Laurent Decreusefond and Pascal Moyal, *Stochastic modeling and analysis of telecom networks*, John Wiley & Sons, 2012.
- [20] A. Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer, Heidelberg, 2011.
- [21] S.N. Ethier and T.G. Kurtz, *Markov processes : Characterizations and convergence*, Wiley, 1986.
- [22] M.-P. Etienne, L. Decreusefond, G. Lang, and S. Robin, *Detecting genomic alteration in genomic profiles: the infinite population case*, IWAP (International Workshop on Applied Probability) (Antaly, Turkey), 2014.
- [23] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge University Press, 2010.
- [24] Robert E. Gaunt and Neil Walton, *Stein’s method for the single server queue in heavy traffic*, Statistics & Probability Letters **156** (2020).

- [25] E. Haeusler, *On the rate of convergence in the invariance principle for real-valued functions of doebelin processes*, Journal of Multivariate Analysis **15** (1984), 73–90.
- [26] J. M. Harrison and R. J. Williams, *Brownian models of open queueing networks with homogeneous customer populations*, Stochastics **22** (1987), no. 2, 77–115.
- [27] A. G. Hawkes and D. Oakes, *A cluster process representation of a self-exciting process*, Journal of Applied Probability **11** (1974), no. 3, 493–503.
- [28] Caroline Hillairet, Lorick Huang, Mahmoud Khabou, and Anthony Réveillac, *The malliavin-stein method for hawkes functionals*, arXiv preprint arXiv:2104.01583 (2021).
- [29] K. Itô and M. Nisio, *On the convergence of sums of independent Banach space valued random variables*, Osaka Journal of Mathematics **5** (1968), 35–48.
- [30] J. Jacod, *Calcul stochastique et problèmes de martingales*, Springer-Verlag, 1979.
- [31] A. Jakubowski, *The Skorokhod space in functional convergence: a short introduction*, <https://kpsc.umk.pl/Content/39953/PDF/kievtologies.pdf>.
- [32] M.J. Kasprzak, *Diffusion approximations via Stein’s method and time changes*, arXiv:1701.07633 [math] (2017).
- [33] L. Kruk, J. Lehoczyk, K. Ramanan, and S. Shreve, *Heavy traffic analysis for EDF queues with reneging*, The Annals of Applied Probability **21** (2011), no. 2, 484–545.
- [34] L. Massoulié and P. Brémaud, *Stability of nonlinear hawkes processes*, The Annals of Probability **24** (1996), no. 3, 1563–1588.
- [35] Martin Raič, *A multivariate central limit theorem for lipschitz and smooth test functions*, arXiv preprint arXiv:1812.08268 (2018).
- [36] P. Robert, *Stochastic networks and queues*, french ed., Applications of Mathematics, vol. 52, Springer-Verlag, Berlin, 2003.
- [37] S. Sawyer, *Rates of convergence for some functionals in probability*, Annals of Mathematical Statistics **43** (1972), 273–284 (English).
- [38] Hsin-Hung Shih, *On Stein’s method for infinite-dimensional Gaussian approximation in abstract Wiener spaces*, Journal of Functional Analysis **261** (2011), no. 5, 1236–1283.
- [39] G. L. Torrisi, *Gaussian approximation of nonlinear Hawkes processes*, Ann. Appl. Probab. **26** (2016), no. 4, 2106–2140.
- [40] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
- [41] Erik Volz, *Sir dynamics in random networks with heterogeneous connectivity*, Journal of mathematical biology **56** (2008), no. 3, 293–310.

LTCI, TELECOM PARIS, I.P. PARIS, FRANCE
Email address: `eustache.besancon@mines-telecom.fr`

INSTITUT MATHÉMATIQUE DE TOULOUSE, UNIVERSITÉ P. SABATIER, TOULOUSE, FRANCE
Email address: `coutin@math.univ-toulouse.fr`

LTCI, TELECOM PARIS, I.P. PARIS, FRANCE
Email address: `laurent.decreusefond@mines-telecom.fr`

UNIVERSITÉ DE LORRAINE, FRANCE
Email address: `pascal.moyal@univ-lorraine.fr`