Diffusive limits of lipschitz functionals of poisson measures

Eustache Besançon, Laure Coutin, Laurent Decreusefond, Pascal Moyal

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 diffuse 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. [5, 31, 23, 28] and references therein), biology and epidemics (e.g. [8, 18, 15] and references therein), physics [10], 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 [19], or Chapter 7 in [17]), 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

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

of time-changed Poisson processes and then to use the well-known limit theorems for such processes, see e.g. Chapter 6 in [19]. 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. [8]). We mention these different approaches because not all of them behave nicely for what has to be done here.

After fluid and diffusion approximations, the third and next natural step is to evaluate the rate of convergence in the diffusion approximation. 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 $(X_n,\, n\ge 1)$ 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

$$\operatorname{dist}_{\mathcal{C}}(\mathbf{P}_{X_n},\, \mathbf{P}_X) := \sup_{f\in\mathcal{C}} \mathbf{E}\left[f(X_n)\right] - \mathbf{E}\left[f(X)\right] \le \varphi(n), \tag{1}$$

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. [22, 32], 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 [27] 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 [7]; a system with reneging and phase-type service time distributions in [6], and single-server queues in heavy-traffic in [21].

It is only recently that in [11], Barbour's result was extended to the convergence in some fractional Sobolev spaces, instead of $\mathbb{D}$. This result was then improved in [13] and [12] 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 Prokohorov distance: not only does it imply the convergence in distribution, but also the convergence of the first moments, see [33].

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 [3], 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.

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 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 $a \le c\, b$.

2.1. **Affine interpolations.** In the forthcoming examples, we have 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. Here we aim at a more precise result. Actually, our goal is to estimate which factor is responsible for the slowest rate of convergence: the difference of regularity or the difference in the dynamics.

It leads us to consider the distance between the affine interpolation of the processes under scrutiny and the affine interpolation of the BM, instead of the distance between their nominal trajectory and that of the BM.

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

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

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

$$|\pi| = \sup_{i \in [\![0, l(\pi)-1]\!]} |t_{i+1} - t_i|.$$

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

For any function $f \in \mathbb{D}([0,T], \mathbf{R}^d)$ 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 $t \in [0,T]$,

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

where

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

When $\pi = \{iT/n,\, i \in [\![0, n]\!]\}$, 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*

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

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

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

where for all $(n,x)\in\mathbf{N}^*\times\mathbf{R}$,

$$\Psi(n,x)=\frac{\log(ne^{x/n})}{\log(nx^{-1}\log(ne^{x/n}))}.$$

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

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

so that

$$\mathbf{E}\left[\|\ X-\Xi_\pi X\ \|_{\infty,T}\right]\le 2\mathbf{E}\left[\max_{i\in[\![0,n-1]\!]}\|X(t_{i+1})-X(t_i)\|_{\mathbf{R}^d}\right].$$

Now notice that for all $k\in [\![1,m]\!]$,

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

But the $M^\pi_{k,i}$, $i\in [\![0,l(\pi)-1]\!]$ 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. □

As to the affine interpolation of the Brownian motion, according to [20, Proposition 13.20], there exists $c>0$ such that for any partition π

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

2.2. **Malliavin gradient.** We give the minimum elements of Malliavin calculus to understand the sequel. More advanced material is necessary to prove Theorem 4.2, see [12]. We denote by $\mathcal{H}_T$ the Hilbert space

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

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

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

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

Consider $(Z_n,\, n \ge 1)$, a sequence of independent, standard Gaussian random variables and let $(z_n,\, n \ge 1)$ be a complete orthonormal basis of $\mathcal{H}_T$. Then, we know from [25] that

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

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

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

where $\mathfrak{e}$ 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

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

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

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

This means that if $h = \lim_{n\to\infty} \mathfrak{e}^*(\eta_n)$ in $\mathcal{I}_{1,2}$,

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

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

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

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

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

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

which is equivalent to say

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

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 $k \ge 1$, 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 $L^2([0,T],\,\mathrm{d}s)$. Recall that we have identified $\mathcal{H}_T$ with itself, so that we cannot identify L^2 with its dual. It is proved in [13] that

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

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

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

for any $x\in\mathcal{C}_T$, $g\in\mathcal{H}_T$, $h,k\in L^2([0,T],\,\mathrm{d}s)$. 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 $F:E\to G$ is said to be Lipschitz continuous whenever there exists $c>0$ such that for any $x,y\in E$,*

$$d_G(F(x),F(y))\le c\,d_E(x,y). \tag{7}$$

The minimum value of c such that (7) holds, is the Lipschitz norm of F. We denote by $\mathrm{Lip}_\alpha(E\to G,d_E)$ 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*

$$\gamma(t)=\int_0^t r(s)\,\mathrm{d}s. \tag{8}$$

Then, the map

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

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. □

Let us now fix $m,d\in\mathbf{N}^*$, and consider the integral equation

$$y(t)=y(0)+\int_0^t Ay(s)\,\mathrm{d}s+f(t),\quad t\ge 0, \tag{9}$$

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

$$S(f)(t)=Ae^{tA}f(0)-Af(t)+A\int_0^t e^{(t-r)A}f(r)\,\mathrm{d}r. \tag{10}$$

We have the following,

Theorem 3.3. *For $A \in \mathfrak{M}_n(\mathbf{R})$, the map*

$$\begin{aligned} \Theta_A \,:\, \mathcal{C}_T(\mathbf{R}^d) &\longmapsto \mathcal{C}_T(\mathbf{R}^d) \\ f &\longmapsto S(f) \end{aligned} \tag{11}$$

where $S(f)$ is the solution of (9), is Lipschitz continuous. Moreover, if F belongs to Υ_T, then so does $F \circ S$.

Proof. Since Θ_A is linear, we just have to prove that there exists $c > 0$ such that for any $f \in \mathcal{C}_T$,

$$\|S(f)\|_{\infty,T} \lesssim_A \|f\|_{\infty,T}.$$

From(10), we have

$$\|S(f)\|_{\infty,T} \lesssim (\|A\| + \|A\|^2)e^{T\|A\|}\, \|f\|_{\infty,T}.$$

Equation (10) also entails that S is linear and that

$$S(\mathcal{H}_T) \subset \mathcal{H}_T \text{ and } S(\mathcal{J}_T) \subset \mathcal{J}_T,$$

hence Υ_T is stable by S. The proof is thus complete. □

Recall (see e.g. Chapter D in [31]), 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 $t \leq T$,

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

Define the mapping

$$\begin{aligned} \mathrm{Sko} \,:\, \mathbb{D}_T &\longrightarrow \mathbb{D}_T \\ f &\longmapsto \left(s \mapsto f(s) + \|f^-\|_{\infty,s}\right), \end{aligned} \tag{12}$$

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 $X_Y = \mathrm{Sko}(Y)$. 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 Skorohod reflection map Sko *are all Lispchitz 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 $t \in [0, T]$,

$$\Big| \|f\|_{\infty,t} - \|g\|_{\infty,t} \Big| \le \|f-g\|_{\infty,t}.$$

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

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

Hence for all ε,

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

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

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

and the proof is complete. □

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*

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

It is the well known that $(\mu_n,\, n \ge 1)$ 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 $x_0 \in E$

$$\int_E d_E(x,x_0)\,\mathrm{d}\mu_n(x) \xrightarrow{n\to\infty} \int_E d_E(x,x_0)\,\mathrm{d}\nu(x). \tag{13}$$

In [12], it is proved that

Theorem 4.2. *Let $(X_i,\, i \ge 0)$ be a sequence of independent and identically distributed random variables belonging to L^p for some $p \ge 3$. Then, for any $n \ge 1$,*

$$\mathfrak{K}_{\mathcal{C}_T}\left(\sum_{i=0}^{n-1} X_i h_i^n,\ B\right) \lesssim_{T,\mathbf{E}[|X_1|^3]} \ n^{-1/6}\,\log(n). \tag{14}$$

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 $N < n$

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

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

$$\mathcal{H}_N = \operatorname{span}\left\{h_i^N, i \in [\![0, N-1]\!]\right\}.$$

This leads to the definition of finite rank functional,

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

$$\mathcal{H}_\pi = \operatorname{span}\left\{h_i^\pi,\ i = 0, \cdots, l(\pi)-1\right\}.$$

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

$$F = \varphi_F \circ \Xi_\pi. \tag{15}$$

It amounts to saying that F depends only on the value of x by its value at the points of π. For any $\pi \in \Sigma_T$, we denote by $\mathcal{F}^\pi$ the set of functions of finite rank associated to π, that is, such that (15) *holds.*

A straightforward adaptation of the proof of the main theorem in [12] yields

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

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

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

$$\sup_{F \in \operatorname{Lip}_1^\pi} \left\{ \mathbf{E}\left[F\left(\Xi_\pi\left(\sum_{i=0}^{n-1} X_i h_i^n\right)\right)\right] - \mathbf{E}\left[F\left(\Xi_\pi B\right)\right] \right\} \lesssim_{T,\mathbf{E}[|X_1|^3]} \frac{|\pi|^{-1}}{\sqrt{n}} \log(n). \tag{16}$$

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 [30, 4] 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,

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

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

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

Then, setting for all μ, ν,

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

it is shown in [13] 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 (29)), 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 $(X_n, \, n \ge 1)$ is said to converge in distribution at the usual rates to a process Z if for any $\pi \in \Sigma_T$ and any $n \ge 1$,*

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

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 $\left(\overline{P}_n, \, n \ge 1\right)$ defined by*

$$\overline{P}_n(t) := \frac{P_n(t) - nt}{\sqrt{n}} \tag{17}$$

converges at the usual rates to a Brownian motion. Furthermore,

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

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

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

Furthermore,

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

where $(X_k, \, k \ge 1)$ is an IID sequence of Poisson random variables of parameter 1. The result then follows from (14) and (16). □

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 defined by (8). Let $\mathcal{N}^n$ be a Poisson measure on $[0,T]\times\mathbf{R}^+$ of intensity measure $n\ \mathrm{d}t\otimes\mathrm{d}z$. Consider the compensated Poisson measure*

$$\mathrm{d}\widetilde{\mathcal{N}}^n(t,z)=\mathrm{d}\mathcal{N}^n(t,z)-n\ \mathrm{d}t\otimes\mathrm{d}z.$$

Define also the process $\overline{R}_n$ by

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

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

Proof. According to [9, Theorem 16], for any $n\ge 1$, $\overline{R}_n\circ\gamma^{-1}$ has the distribution of $\overline{P}_n$ defined in (17). Therefore, from Theorem 4.6, $(\overline{R}_n\circ\gamma^{-1},\ n\ge 1)$ converges at the usual rates to B. The result then follows from Lemma 3.2. □

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 $(\zeta_1,...,\zeta_m)_{1\le k\le m}$ of elements of $\mathbf{R}^d$ and any array $(\rho_k)_{1\le k\le m}$ 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

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

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

Fix $n\in\mathbf{N}^*$ and an array $\alpha:=(\alpha_1,...,\alpha_m)\in(\mathbf{R})^m$. 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 $(n^{\alpha_k}\ \mathrm{d}s)\otimes\mathrm{d}z$, and normalizing in space by n. Then the process $\overline{X}_n:=n^{-1}X_n$ is the solution of the following SDE: For any $t\ge 0$,

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

The key assumption on our scaling is the following:

Assumption 1 (Law of large numbers scaling). *For any $k\in[\![1,m]\!]$, there exists a mapping $r_k\in\mathrm{Lip}_c(\mathbf{R}^d\to\mathbf{R},\|\,\|_{\mathbf{R}^d})$ such that for all $n\in\mathbf{N}^*$, all $x\in\mathbf{R}^d$ and $t\le T$,*

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

and such that

$$\sup_{n\ge 1}\mathbf{E}\left[\sup_{t\le T}\left|r_k(t,\overline{X}_n(t))\right|\right]\le K_k.\tag{19}$$

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

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

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

$$M_{n,k,\overline{X}_n}(t) = \left(\int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\left\{z \le n^{1-\alpha_k} \; r_k\left(s, \overline{X}_n(s^-)\right)\right\}} \, \mathrm{d}\widetilde{\mathcal{N}}_k^{\alpha_k}(s,z) \right) \zeta_k \tag{20}$$

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

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

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

$$\Lambda(t) = \Lambda(0) + \sum_{k=1}^m \left(\int_0^t r_k(s, \Lambda(s)) \, \mathrm{d}s \right) \zeta_k, \quad t \ge 0. \tag{22}$$

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*

$$\sup_{n \ge 1} \mathbf{E}\left[\sup_{t \le T} \|\overline{X}_n(t)\| \right] \le e^{cT}$$

and for all $n \ge 1$,

$$\mathbf{E}\left[\sup_{t \le T} \left\|\overline{X}_n(t) - \Lambda(t)\right\| \right] \le \left(\mathbf{E}\left[\left\|\overline{X}_n(0) - \Lambda(0)\right\| \right] + K n^{-1/2} \right) e^{cT}. \tag{23}$$

Proof. By denoting for all $k \in [\![1,m]\!]$, by λ_k the Lipschitz constant of the mapping r_k, we readily get that for all $t \ge 0$,

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

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

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

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

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

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

In view of (20), we see that the integrands in the $M_{n,k,\overline{X}_n}$'s, $k \in [\![1,m]\!]$, are random processes that depend on n. In this form, we cannot directly use Corollary 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.

So let us introduce for all n, the $\mathbf{R}^d$-valued martingales $(M_{n,k,\overline{X}_n})_{k\in[\![1,m]\!]}$, $(M_{n,k,\Lambda})_{k\in[\![1,m]\!]}$ and $\overline{M}_{n,\Lambda}$, respectively defined for all $t \leq T$ by

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

Observe that the crucial difference between $\overline{M}_{n,\overline{X}_n}$ and $\overline{M}_{n,\Lambda}$ 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

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

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

$$\mathbf{E}\left[\left\|\overline{X}_n(0) - \Lambda(0)\right\|\right] \lesssim n^{-1/2}, \tag{25}$$

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

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

where for all $k \in [\![1,m]\!]$, $L_k \in \mathbf{R}^d$ and

$$\mathbf{E}\left[\|E_{n,k}\|_{\infty,T}\right] \xrightarrow{n\to\infty} 0.$$

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

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

where $\Theta_{\mathbf{A}}$ is the map associated to the matrix $\mathbf{A} = (L_1, ..., L_m) \otimes (\zeta_1, ..., \zeta_m)$ by (11), $(B_k)_{k\in[\![1,m]\!]}$ are independent one dimensional Brownian motions, *and for any $k \in [\![1,m]\!]$, 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 $n \geq 1$,

$$U_n = \Theta_{\mathbf{A}} \left(\sum_{k=1}^{m} \left(\int_0^t E_{n,k}(s) \, \mathrm{d}s \right) \zeta_k + \overline{M}_{n,\overline{X}_n} \right). \tag{27}$$

But first, according to Assumption 2 we get that

$$\mathbf{E} \left[\left\| \sum_{k=1}^{m} \left(\int_0^{\cdot} E_{n,k}(s) \, \mathrm{d}s \right) \zeta_k \right\|_{\infty,T} \right] \xrightarrow{n\to\infty} 0. \tag{28}$$

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

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

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

$$\mathbf{E} \left[\left\| n^{-1/2} M_{n,k,\overline{X}_n} - n^{-1/2} M_{n,k,\Lambda} \right\|_{\infty,T} \right] \lesssim n^{-1/4}. \tag{29}$$

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

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

where $\mathcal{N}^n$ is a Poisson measure of intensity $n \; \mathrm{d}t \otimes \mathrm{d}z$, 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 (27) and Theorem 3.3, together with (28) and (29). □

5.2. **The Telegraph process.** Let Y_i, $i \in [\![1,n]\!]$ 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, $i \in [\![1,n]\!]$ 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, $i \in [\![1, n]\!]$, 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 [14] 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 $\Lambda(0) \in \mathbf{R}+$,

$$\Lambda(t) = \pi_1 + (\Lambda(0) - \pi_1)\exp(-(\sigma_1 + \sigma_0)t), \quad t \geq 0; \tag{30}$$

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

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

- The process Y is defined by

$$Y(t) = \int_0^t \sqrt{\sigma_0(1 - \Lambda(s))}\, \mathrm{d}B_1(s) - \int_0^t \sqrt{\sigma_1\Lambda(s)}\, \mathrm{d}B_2(s), \quad t \geq 0,$$

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

$$\overline{X}_n(t) \overset{(d)}{=} \overline{X}_n(0) + \int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z \leq n\sigma_0(1-\overline{X}_n(s^-))\}}\, \mathrm{d}\mathcal{N}_1(s,z) \\ - \int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z \leq n\sigma_1\overline{X}_n(s^-)\}}\, \mathrm{d}\mathcal{N}_2(s,z), \quad t \geq 0,$$

where $\mathcal{N}_1$ and $\mathcal{N}_2$ denote two independent Poisson random measures of common intensity $\mathrm{d}s \otimes \mathrm{d}z$ on $(\mathbf{R}_+)^2$, representing the overall "up" and "down" jumps, respectively. So $\overline{X}_n$ 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

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

which are obviously Lipschitz continuous with respect to their second variable. Plainly, Λ defined by (30) 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 $\Theta_{\mathbf{A}} = \Theta$ defined by (11) is linear, continuous and therefore Lipschitz continuous. We conclude using Theorem 5.2. □

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 [3], can be studied in the present framework. The convergence rate obtained hereafter is $n^{1/6}$, slightly slower than that obtained in [3] ($n^{1/2}$), 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/∞ 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 $t \geq 0$, $L^{\#}(t)$ denote the number of customers in the system at time t. It is well known that $L^{\#}$ 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 $t \geq 0$, $\overline{L_n^{\#}}(t) = L^{\#}(t)/n$. Then, it is a classical result (see e.g. [5], Theorem 6.14 in [31], or a measure-valued extension of the result in [16]), that under suitable assumptions on the initial conditions, for all $T > 0$,

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

where

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

$$\frac{\lambda}{\mu} - \left(\Lambda(0) - \frac{\lambda}{\mu}\right)\exp(-\mu t), \quad t \geq 0; \tag{31}$$

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

$$g(t) = -\mu \int_0^t g(s) \, \mathrm{d}s + f(t), \quad t \geq 0;$$

- The process Y is defined by

$$Y(t) = \sqrt{\lambda} B_1(t) - \int_0^t \sqrt{\mu \Lambda(s)} \, \mathrm{d}B_2(s), \quad t \geq 0,$$

  for B_1 and B_2, two independent standard Brownian motions.

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 $\overline{L}_n^{\#}$ satisfies the SDE

$$\overline{L}_n^{\#}(t) = \overline{L}_n^{\#}(0) + \frac{1}{n}\int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z\le\lambda\}} \, \mathrm{d}\mathcal{N}_1^1(s,z) \\ - \frac{1}{n}\int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z\le\mu\overline{L}_n^{\#}(s^-)\}} \, \mathrm{d}\mathcal{N}_2^1(s,z), \quad t \geq 0,$$

where $\mathcal{N}_1^1$ and $\mathcal{N}_2^1$ denote two independent random Poisson measures of intensity $n \, \mathrm{d}s \otimes \mathrm{d}z$ on $(\mathbb{R}_+)^2$. This precisely means that the process $\overline{L}_n^{\#}$ 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 $r_1 : (t,y) \longmapsto \lambda$ and $r_2 : (t,y) \longmapsto \mu y$. Clearly, Λ defined by (31) 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 (11). We conclude again using Theorem 5.2. □

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 $\big(L^\dagger(t),\, t \geq 0\big)$ 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 $t \geq 0$,

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

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 $\mathrm{d}s \otimes \mathrm{d}z$ 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

$$\overline{L}_n^\dagger(t) = x + \frac{1}{n}\int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z\le\lambda\}} \, \mathrm{d}\mathcal{N}_1^1(s,z) - \frac{1}{n}\int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z\le\mu\}} \, \mathrm{d}\mathcal{N}_2^1(s,z) \\ + \frac{1}{n}\int_0^t \int_{\mathbb{R}+} \mathbf{1}_{\{z\le\mu\}}\mathbf{1}_{\left\{\overline{L}_n^\dagger(s^-)=0\right\}} \, \mathrm{d}\mathcal{N}_2^1(s,z), \tag{32}$$

where $\mathcal{N}_1^1$ and $\mathcal{N}_2^1$ denote two Poisson random measures of intensity $n \, \mathrm{d}s \otimes \mathrm{d}z$. The following result completes the classical diffusion limit presented e.g. in Proposition 5.16 in [31], and Theorem 1 in [3],

Proposition 5.5. *Let*

$$\Lambda(t) = (x + (\lambda - \mu)t)^+ , \quad t \geq 0, \tag{33}$$

and Z be the process defined by

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

for B a standard Brownian motion. Then the following convergence holds at the usual rate,

$$U_n = \sqrt{n} \left(\overline{L}_n^\dagger - \Lambda \right) \Longrightarrow \mathrm{Sko}(Z) \quad \text{in } \mathbb{D}([0,T], \mathbf{R}),$$

where the mapping Sko *is defined by (12).*

Proof. The relation (32) means that $\overline{L}_n^\dagger = \mathrm{Sko}(\overline{X}_n)$ for the process $\overline{X}_n$ define for all $t \geq 0$ by

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

It is then immediate that both Theorems 5.1 and 5.2 can be applied to the sequence of processes $\{\overline{X}_n\}$ 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 $\Theta_A := \mathbf{Id}$), that the convergence

$$V_n := \sqrt{n} \left(\overline{X}_n - \Gamma \right) \Longrightarrow Z \quad \text{in } \mathbb{D}([0,T], \mathbf{R})$$

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

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

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

5.4. **SIR epidemics.** We consider a population of constant size n, in which individuals can go through three states, susceptible, infectious ad 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 $t \geq 0$, $\overline{S}_n(t) = S^n(t)/n$, $\overline{I}_n(t) = I^n(t)/n$ and $\overline{R}^n(t) = R^n(t)/n$, which represent respectively the proportions of Susceptible, Infected and Recovered individuals in the whole population at time t. We also let $\overline{X}^n(t) = \begin{pmatrix} \overline{S}^n(t) \\ \overline{I}^n(t) \end{pmatrix}$, $t \geq 0$. A large-graph limit and a functional central limit theorem for the process $\overline{X}$ are given in Chapter 2 of [8], 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 [15], completing the result in [34]. The following result makes precise the speed of convergence in the functional CLT for the complete-graph case, given in [8]. We are confident that similar results hold for the other related models addressed in Chapter 2 of [8], however we only consider here the SIR case for brevity,

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

$$\begin{cases} s'(t) &= -\lambda s(t)i(t) \\ i'(t) &= \lambda s(t)i(t) - \gamma i(t) \end{cases}, \quad t \geq 0. \tag{34}$$

Let $\Theta(Y)$ *denote the unique solution in* $\mathbb{D}([0,T], \mathbf{R})$ *of the following SDE of unknown* $g = \begin{pmatrix} g_1 \\ g_2 \end{pmatrix}$,

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

where for all $t \geq 0$,

$$Y(t) = \int_0^t \sqrt{\lambda s(u)i(u)} \,\mathrm{d}B_1(u) \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \int_0^t \sqrt{\gamma i(u)} \,\mathrm{d}B_2(u) \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

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,*

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

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

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

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 $d = 2$, $m = 2$, $\alpha_1 = \alpha_2 = 0$, $\zeta_1 = \binom{-1}{1}$ and $\zeta_2 = \binom{0}{-1}$. 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 $y = \binom{y_1}{y_2} \in [0,1]^2$, $y' = \binom{y_1}{y_2} \in [0,1]^2$ for all t,

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

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

$$r_1(t,y') - r_1(t,y) = \lambda \left(y_2(y_1' - y_1) + y_1(y_2' - y_2) + (y_2' - y_2)(y_1' - y_1) \right),$$

entailing that for all n and t,

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

for

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

Then, from (23) we obtain that

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

Likewise, it is immediate that r_2 is Lipschitz continuous in its second variable, and we get for all n and t that

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

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

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 mutes independently to type B at rate ν_1 and every gene of type B mutes 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 $\overline{X}_n(t) = X_n(t)/n$ represents the proportion of individuals carrying the gene of type A at time t. A functional Stein method is applied to this model in [27]. 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*

$$\Lambda(t) = \Lambda(0) + \int_0^t (\nu_2 - (\nu_1 + \nu_2)\Lambda(s)) \, \mathrm{d}s, \quad t \geq 0.$$

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

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

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

$$y(t) = y(0) + (\nu_1 + \nu_2)\int_0^t y(s)\,\mathrm{d}s + f(t), \quad t \geq 0,$$

and where

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

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

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

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

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

where $\mathcal{N}_i^{-1}, i \in [\![1,4]\!]$, are four independent Poisson measures of intensity $n^{-1}\,\mathrm{d}s \otimes \mathrm{d}z$. 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 $\alpha_1 = ... = \alpha_4 = -1$. 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

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

and from (23) we obtain that

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

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

6. Limit theorem for Hawkes processes

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

$$\kappa := \int_0^\infty \varphi(t) \, \mathrm{d}t < 1 \text{ and } \int_0^\infty t^{1/2} \varphi(t) \, \mathrm{d}t < \infty.$$

We denote by Φ the first primitive of φ:

$$\Phi(t) = \int_0^t \varphi(s) \, \mathrm{d}s, \quad t \geq 0.$$

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

$$\int_0^\infty \psi(t) \, \mathrm{d}t = \sum_{k \geq 1} \kappa^k = \frac{\kappa}{1 - \kappa} \cdot$$

For $\mu > 0$, according to [24, 29], there exists a point process N (unique in distribution) such that N admits the compensator

$$t \longmapsto \mu t + \int_0^t \varphi(t - s) \, \mathrm{d}N(s).$$

In [1], it is proved that

$$\mathbf{E}\left[\sup_{v \in [0,1]} \left|\frac{1}{n} N(nv) - \rho v\right|^2\right] \lesssim \frac{1}{n}, \tag{35}$$

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

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

Our goal is to assess the rate of this convergence. To that end, we use a particular construction of N based on a Poisson measure M of intensity measure $\mathrm{d}s \otimes \mathrm{d}z$. For all $t \geq 0$, we know from [29] and references therein, that we can write

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

Denote also

$$W(t) = \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + \int_0^t \varphi(t-s)\ \mathrm{d}N(s)\}} \left(\mathrm{d}M(s,z) - \ \mathrm{d}s\ \mathrm{d}z\right),$$

the corresponding compensated integral, so that W is a local martingale with respect to the filtration induced by M:

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

For a process Z, we denote by

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

From [26, Chapter 13], we also know that

$$W^{(n)}(t) = \int_0^{nt} \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + n\int_0^t \varphi(nt-ns)\ \mathrm{d}\tilde{N}^{(n)}(s)\}} \left(\mathrm{d}M(s,z) - \ \mathrm{d}s\ \mathrm{d}z\right),\, t \geq 0.$$

These considerations result in the following lemma,

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

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

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

Proof. Fix $n \geq 1$. As $\tilde{N}^{(n)}$ is asymptotically close to γ we consider the process

$$R^{(n)} : t \longmapsto \int_0^{nt} \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + n\int_0^t \varphi(nt-ns)\rho\ \mathrm{d}s\}} \left(\mathrm{d}M(s,z) - \ \mathrm{d}s\ \mathrm{d}z\right).$$

It has the same distribution as

$$\hat{R}^{(n)} : t \longmapsto \int_0^t \int_{\mathbf{R}^+} \mathbf{1}_{\{z \leq \mu + \int_0^t \varphi(t-s)\rho\ \mathrm{d}s\}} \left(\mathrm{d}\mathcal{N}^1(s,z) - n\ \mathrm{d}s\ \mathrm{d}z\right),$$

where $\mathcal{N}^1$ is a Poisson measure of intensity $n\ \mathrm{d}s \otimes \mathrm{d}z$. According to Corollary 4.7,

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

which is equivalent to

$$\mathfrak{K}\left(\Xi_n(\overline{R}^{(n)}),\ \Xi_n(B \circ \gamma)\right) \lesssim n^{-1/6} \log(n). \tag{36}$$

We now want to estimate the error made by considering $\overline{R}_n$ instead of $\overline{W}_n$. Let for all $t \geq 0$,

$$r(t) = \mu + n \int_0^t \varphi(nt-ns)\ \mathrm{d}\tilde{N}^{(n)}(s)$$

$$r'(t) = \mu + n \int_0^t \varphi(nt-ns)\rho\ \mathrm{d}s.$$

We have

$$
\begin{aligned}
&\|\Xi_n(\overline{R}^{(n)}) - \Xi_n(\overline{W}^{(n)})\|_{\infty,1} \\
&\qquad \lesssim \sup_{i\in[\![0,n-1]\!]} \left| (\overline{R}^{(n)} - \overline{W}^{(n)})(\frac{i+1}{n}) - (\overline{R}^{(n)} - \overline{W}^{(n)})(\frac{i+1}{n}) \right| \\
&\qquad = \sup_{i\in[\![0,n-1]\!]} \left| \int_{i/n}^{(i+1)/n} \int_{\mathbf{R}^+} \left(\mathbf{1}_{\{r\le r(s)\}} - \mathbf{1}_{\{z\le r'(s)\}}\right) \mathrm{d}\widetilde{\mathcal{N}}(s,z) \right|.
\end{aligned}
$$

Apply the BDG inequality to the vector valued martingale

$$
t \longmapsto \left(\int_{t\vee i/n}^{t\wedge(i+1)/n} \int_{\mathbf{R}^+} \left(\mathbf{1}_{\{r\le r(s)\}} - \mathbf{1}_{\{z\le r'(s)\}}\right) \mathrm{d}\widetilde{\mathcal{N}}(s,z),\ i\in[\![0,n-1]\!] \right),
$$

to obtain that

$$
\begin{aligned}
\mathbf{E}\left[\|\Xi_n(\overline{R}^{(n)}) - \Xi_n(\overline{W}^{(n)})\|_{\infty,1}\right] &\lesssim \frac{1}{\sqrt{n}}\mathbf{E}\left[\sum_{i=0}^{n-1} \int_{i/n}^{(i+1)/n} |r(s)-r'(s)|\ \mathrm{d}s\right]^{1/2} \\
&= \frac{1}{\sqrt{n}}\mathbf{E}\left[\int_0^1 |r(s)-r'(s)|\ \mathrm{d}s\right].
\end{aligned}
$$

With the particular expression of r and r', we get that

$$
\begin{aligned}
&\mathbf{E}\left[\|\Xi_n(\overline{R}^{(n)}) - \Xi_n(\overline{W}^{(n)})\|_{\infty,1}\right] \\
&\qquad\lesssim \frac{1}{\sqrt{n}}\mathbf{E}\left[\left|\int_0^n n\int_0^t \varphi(nt-ns)(\,\mathrm{d}\tilde{N}^{(n)}(s) - \rho\ \mathrm{d}s)\right|\right]^{1/2} \\
&\qquad= \frac{1}{\sqrt{n}}\mathbf{E}\left[\left|n\int_0^1 \Phi(s)\left(\,\mathrm{d}\tilde{N}^{(n)}(s) - \rho\ \mathrm{d}s\right)\right|\right]^{1/2} \\
&\qquad= \mathbf{E}\left[\left|\int_0^1 (\tilde{N}^{(n)}(s) - \rho s)\varphi(s)\ \mathrm{d}s\right|\right]^{1/2},
\end{aligned}
$$

by integration by parts. Thus, we have that

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

In view of (35), we get

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

Thus, we can substitute $\overline{W}^{(n)}$ to $\overline{R}^{(n)}$ in (36), and the result follows. □

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

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

Then, we have for all $t\ge 0$,

$$
\overline{X}^{(n)}(t) = \overline{W}^{(n)}(t) + \int_0^t n\psi(ns)\,\overline{W}^{(n)}(t-s)\ \mathrm{d}s. \tag{37}
$$

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. *Assume that there exist $\epsilon \in (0,1/2)$ such that for all n,*

$$\int_{n^\epsilon}^{\infty} \psi(t)\ \mathrm{d}t \lesssim n^{-1/2}. \tag{38}$$

Then, for all $n \geq 1$ we have that

$$\mathfrak{K}\left(\Xi_n(\overline{X}^{(n)}),\, \Xi_n(B\circ\zeta)\right) \lesssim n^{-1/6}\log(n)$$

where

$$\zeta(t) = \frac{\rho}{\sqrt{1-\kappa}}\, t, \quad t \geq 0.$$

Remark 1. *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* (13)*, which is not valid for the other distances investigated above.*

Remark 2. *The classical choice of functions φ are sums of exponential functions. In this situation, the integral of the left-hand-side of* (38) *goes to zero exponentially fast, so that the hypothesis is truly satisfied.*

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

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

$$I_n := \mathbf{E}\left[\sup_{t\in[0,1]}\left|\int_0^t n\psi(ns)\overline{W}^{(n)}(t-s)\ \mathrm{d}s - \frac{\kappa}{1-\kappa}\overline{W}^{(n)}(v)\right|\right] \lesssim n^{-1/2}. \tag{39}$$

Then according to (37),

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

The result then follows from Lemma 6.1.

We now establish (39). According to the decomposition given in [1], for any $0<\delta<\eta$, we have

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

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

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

In view of Theorem 3.4, we also have that

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

for any $\epsilon > 0$. If we choose $\eta = n^{-1+\epsilon}$ and $\delta = \eta/2$, we get

$$I_n \lesssim \int_{n^\epsilon}^{\infty}\psi(t)\ \mathrm{d}t + n^{-1/2-\epsilon/2}.$$

The proof is thus complete. $\square$

7. Auxiliary result

Proposition 7.1. *Let $(X_i, i = 1, \cdots, n)$ be Poisson random variables of parameter ν. The Lambert W function is defined over $[-1/e, \infty]$ by the equation $W(x)e^{W(x)} = x$. Then*

$$\mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \leq \frac{\log n/e^\nu}{W(\log(n/e^\nu)/\nu e)} = \nu e \exp\left(W\left(\log(n/e^\nu)/\nu e\right)\right).$$

Proof. Consider $(Z_i, i = 1, \cdots, n)$ some independent centered Poisson variables (i.e., for all i, $Z_i = X_i - \nu$). By a straightforward calculation, for all $u \in \mathbf{R}$ and all i,

$$\mathbf{E}\left[e^{uZ_i}\right] = e^{-u\nu} \sum_{k=0}^{\infty} e^{uk} e^{-\nu} \frac{\nu^k}{k!} = e^{-u\nu-\nu} e^{\nu e^u}$$

Therefore the logarithm of the moment generating function of Z_i is $\Psi_{Z_i}(u) = \nu\left(e^u - u - 1\right)$.

By Jensen's inequality, we obtain

$$\exp\left(u\mathbf{E}\left[\max_{i=1,\cdots,n} Z_i\right]\right) \leq \mathbf{E}\left[\exp\left(u \max_{i=1,\cdots,n} Z_i\right)\right] = \mathbf{E}\left[\max_{i=1,\cdots,n} \exp(uZ_i)\right]$$

Because the maximum of a sequence of positive numbers is lower than its sum, the right hand side of the last equation is lower than $\mathbf{E}\left[\sum_{i=0}^{n} \exp(uZ_i)\right]$. Hence, by the definition of Ψ_{Z_i},

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

Taking the log, for any u in $\mathbf{R}$,

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

so that

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

By differentiation, it is easy to check that the infimum is reached when

$$\nu u e^u - \nu e^u + \nu = \log n. \tag{40}$$

Therefore, the infimum is equal to

$$\frac{\log n + \nu\left(e^{1+W(a)} - 1 - W(a) - 1\right)}{1 + W(a)} \tag{41}$$

But we know from (40) that $\nu(1 + W(a))e^{1+W(a)} - \log n = \nu e^{1+W(a)} - \nu$ so that (41) is equal to

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

Remembering that the Z_i are the centered X_i we thus obtain that

$$\mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \le \nu e \frac{a}{W(a)} - \nu + \nu = \frac{\log\left(n/e^{\nu}\right)}{W(\log(n/e^{\nu})/\nu e)}$$

which completes the proof.

□

We conclude by observing that $W(z) \ge \log(z) - \log\log(z)$ for all $z > e$. Therefore for $n \ge \exp\left(e^{\nu+1} + \nu\right)$, using the second expression for the bound of the expectation of the maximum in Proposition 7.1 we get that

$$\mathbf{E}\left[\max_{i=1,\cdots,n} X_i\right] \le \frac{\log(n/e^{\nu})}{\log\left(\frac{\log(n/e^{\nu})}{\nu e}\right)}.$$

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] E. Besançon, L. Decreusefond, and P. Moyal, *Stein's method for diffusive limit of Markov processes*, Queueing Systems **95** (2020), 173–201.

[4] 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.

[5] AA. Borovkov, *On limit laws for service processes in multi-channel systems*, Siberian Mathematical Journal **8** (1967), no. 5, 746–763.

[6] A. 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.

[7] A. Braverman, J. G. Dai, and J. 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.

[8] T. Britton and E. Pardoux, *Stochastic epidemics in a homogeneous community*, arXiv e-prints (2018), arXiv:1808.05350.

[9] P. Brémaud, *Point processes and queues, martingale dynamics*, Springer-Verlag, New York, 1981, Springer Series in Statistics.

[10] P. M. Chaikin and T. C. Lubensky, *Principles of condensed matter physics*, Cambridge University Press.

[11] L. Coutin and L. Decreusefond, *Stein's method for Brownian approximations*, (2012) (en).

[12] L. Coutin and L. Decreusefond, *Donsker's theorem in Wasserstein-1 distance*, Electron. Commun. Probab. **25** (2020), 1–13.

[13] L. Coutin and L. Decreusefond, *Stein's method for rough paths*, Potential Analysis **50** (2020), 387–406.

[14] L. Decreusefond, M. Etienne, G. Lang, S. Robin, and P. Vallois, *Longest excursion of the Ornstein Uhlenbeck Process: applications to genomics and telecom*, in preparation.

[15] 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.

[16] 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.

[17] L. Decreusefond and P. Moyal, *Stochastic modeling and analysis of telecom networks*, John Wiley & Sons, 2012.

[18] A. Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer, Heidelberg.
[19] S.N. Ethier and T.G. Kurtz, *Markov processes : Characterizations and convergence*, Wiley, 1986.
[20] P. K. Friz and N. B. Victoir, *Multidimensional Stochastic Processes as Rough Paths: Theory and Applications*, Cambridge University Press, 2010.
[21] R. E. Gaunt and N. Walton, *Stein's method for the single server queue in heavy traffic*, Statistics & Probability Letters **156** (2020), 108566.
[22] E. Haeusler, *On the rate of convergence in the invariance principle for real-valued functions of doeblin processes*, **15**, 73–90.
[23] J. M. Harrison and R. J. Williams, *Brownian models of open queueing networks with homogeneous customer populations*, **22**, no. 2, 77–115.
[24] 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.
[25] 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.
[26] J. Jacod, *Calcul stochastique et problèmes de martingales*, Springer-Verlag, 1979.
[27] M.J. Kasprzak, *Diffusion approximations via Stein's method and time changes*, arXiv:1701.07633 [math] (2017).
[28] Ł. Kruk, J. Lehoczky, K. Ramanan, and S. Shreve, *Heavy traffic analysis for EDF queues with reneging*, **21**, no. 2, 484–545.
[29] L. Massoulié and P. Brémaud, *Stability of nonlinear hawkes processes*, The Annals of Probability **24** (1996), no. 3, 1563–1588.
[30] M. Raič, *A multivariate central limit theorem for lipschitz and smooth test functions.*
[31] P. Robert, *Stochastic networks and queues*, french ed., Applications of Mathematics, vol. 52, Springer-Verlag, Berlin, 2003, Stochastic Modelling and Applied Probability.
[32] S. Sawyer, *Rates of convergence for some functionals in probability*, **43**, 273–284 (English).
[33] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.
[34] E. 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`