Spatial birth-death-move processes : basic properties and estimation of their intensity functions

Frédéric Lavancier, Ronan Le Guével

Spatial birth-death-move processes : basic properties and estimation of their intensity functions

Frédéric Lavancier[1] and Ronan Le Guével[2]

[1]LMJL, BP 92208, 2 Rue de la Houssinière, F-44322 Nantes Cedex 03, France,
frederic.lavancier@univ-nantes.fr
[2]Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France,
ronan.leguevel@univ-rennes2.fr

Abstract

Spatial birth-death processes are generalisations of simple birth-death processes, where the birth and death dynamics depend on the spatial locations of individuals. In this article, we further let individuals move during their life time according to a continuous Markov process. This generalisation, that we call a spatial birth-death-move process, finds natural applications in computer vision, bio-imaging and individual-based modelling in ecology. In a first part, we verify that birth-death-move processes are well-defined homogeneous Markov processes, we study their convergence to an invariant measure and we establish their underlying martingale properties. In a second part, we address the non-parametric estimation of their birth, death and total intensity functions, in presence of continuous-time or discrete-time observations. We introduce a kernel estimator that we prove to be consistent under fairly simple conditions, in both settings. We also discuss how we can take advantage of structural assumptions made on the intensity functions, and we explain how bandwidth selection by likelihood cross-validation can be conducted. A simulation study completes the theoretical results. We finally apply our model to the analysis of the spatio-temporal dynamics of proteins involved in exocytosis mechanisms in cells.

Keywords: birth-death process, growth-interaction model, individual-based model, kernel estimator.

1 Introduction

Simple birth-death processes have a long history, ever since at least Feller (1939) and Kendall (1949). They model the temporal evolution of the size of a population. Many generalisations have been considered. Among them, spatial birth-death processes, introduced by Preston (1975), take into account the spatial location of each individual of the population. For these processes, the birth rate (ruling the waiting time before a new birth) may depend on the spatial location of the individuals and, when a birth occurs, a new individual appears in space according to a distribution that may also depend on the locations of the already existing individuals. The dynamics for deaths is similarly a function of the spatial configurations. Spatial birth-death processes concern applications where the births and deaths in a population not only depend on the cardinality of the population but also on spatial competitions or/and spatial dispersions, for example. For this reason, they are adapted to population dynamics in ecology, but they have also been used in other contexts such as in Møller and Sørensen (1994) to describe the evolution of dunes and in Sadahiro (2019) to understand the opening and closure of shops and restaurants in a city. These processes are moreover at the heart of perfect simulation methods for spatial point process models, see (Møller and Waagepetersen, 2004, Chapter 11) and the references therein.

However, in a spatial birth-death process, the spatial location of each individual remains constant until it dies, which is unrealistic for many applications. In this article we rigorously introduce spatial birth-death-move processes, that are an extension where the individuals can move in space during their life time. This kind of dynamics appears for instance in forestry (Renshaw and Särkkä, 2001; Särkkä and Renshaw, 2006; Pommerening and Grabarnik, 2019), where new plants arrive, then grow (this is the move step), before dying. In computer vision, Wang and Zhu (2002) also consider this kind of processes for the appearance in a video of new objects (like flying birds or snowflakes), which follow some motion before disappearing from the video. Another application, that we will specifically consider later in Section 4 for illustration, concerns the dynamics of proteins involved in the exocytosis mechanisms in cells.

As we define it more formally in Section 2, a birth-death-move process $(X_t)_{t\geq 0}$ starts at $t = 0$ from an initial configuration in a polish space E_n, which can be thought of as the space of (possibly marked) point configurations in $\mathbb{R}^d$ with cardinality n. Then it evolves in E_n according to a continuous Markov process $(Y_t^{(n)})_{t\geq 0}$ during a random time (which is the move step), before jumping either in E_{n+1} (a birth) or in E_{n-1} (a death). After a jump, in $E_{n'}$ say, the process starts a new motion in $E_{n'}$ according to a continuous Markov process $(Y_t^{(n')})_{t\geq 0}$, and so on. Apart from the Markov processes $(Y_t^{(n)})_{t\geq 0}$, for $n \geq 0$, that drive the motion in the spaces E_n, this dynamics depends on a birth rate β and a death rate δ, that are in full generality two functions of the current state of $(X_t)_{t\geq 0}$, and on transition kernels for the births and the deaths.

A pure spatial birth-death process, the case where $(X_t)_{t\geq 0}$ does not move between two jumps, is a well defined pure-jump Markov process and its basic properties, including the study of convergence to an invariant measure, have been established in Preston (1975) using a coupling with a simple birth-death process. The rate of convergence have been further investigated in Møller (1989) using the same coupling. The introduction of a move step in the dynamics implies that the process is no more a pure-jump process and the coupling of Preston (1975) does not apply. Nevertheless, we carefully verify in Section 2 that birth-death-move processes are well-defined homogeneous Markov processes. We then confine ourselves to the case where the number of individuals admits an upper bound n^* (that is, $X_0 \in E_n$ with $n \leq n^*$ and if $X_t \in E_{n^*}$ for some t, then only a death can happen). This assumption is not a restriction for most statistical applications in practice and it allows us to state our main results under fairly simple and understandable conditions. Under this single hypothesis, we prove the convergence of the process to an invariant measure, uniformly over the initial configurations. This uniformity, that turns out to be an important technical argument in our statistical study, requires the existence of n^*, even for pure spatial birth-death processes. We moreover clarify in Section 2 the martingale properties of the counting process defined by the cumulative number of jumps, a cornerstone in most of our proofs.

From a statistical perspective, our main contribution developed in Section 3 concerns the non-parametric estimation of the birth and death intensity functions β and δ, and of the total intensity function $\alpha = \beta + \delta$. As far as we know, this is the first attempt in this direction, even for pure spatial birth-death processes. Yet for simple birth-death processes, where the intensities functions only depend on the size of the population, a non-parametric estimator is studied in Reynolds (1973). We consider two settings, whether the process $(X_t)_{t\geq 0}$ is observed continuously in the time interval $[0, T]$ or at discrete time points $t_0, \ldots, t_m$. In both settings we introduce a kernel estimator and prove that it is consistent under natural conditions, when T tends to infinity and the discretisation step tends to 0. The technical difficulties are mainly due to the facts that the intensity functions are defined on an infinite dimension and non-vectorial space (the state-space of the process), and that the invariant measure of $(X_t)_{t\geq 0}$ is unknown. While our estimator takes a quite general form, it typically involves a bandwidth h_T, which for consistency must tend to 0 but not too fast, as usual in non-parametric inference. We explain how h_T can be chosen in practice by likelihood cross-validation.

We also discuss several strategies of estimation, depending on structural assumptions made on the

intensity functions. Their performances are assessed in a simulation study carried out in Section 4. Consider for instance the estimation of $\alpha(x)$ in the continuous case. In a pure non-parametric approach, this estimation relies through the kernel on the distance between x and the observed configurations of (X_t) for $t \in [0,T]$. This approach is ambitious given the infinite dimension of the state space of the process. It is however consistent if α is regular enough, as a consequence of our theoretical results and confirmed in the simulation study, and this first approach may constitute in practice a first step towards more structural hypothesis on α. In a second approach, we assume that $\alpha(x)$ only depends on the cardinality of x, which is the common setting in standard birth-death processes. Under this assumption, we make our estimator depend only on the distance between the cardinality of x and the cardinalities of the observations (X_t) for $t \in [0,T]$. A particular case is the maximum likelihood estimator of the intensity of a simple birth-death process studied in Reynolds (1973), where only observations with exactly the same cardinality as x are used for the estimation of $\alpha(x)$. We instead allow to use all observations, which makes sense if α has some regularity properties, a situation where our estimator outperforms the previous one in our simulation study. This second approach is a way to question the classical assumption in birth-death processes, namely that each individual has constant birth rate and death rate, implying that $\alpha(x)$ is a linear function in the cardinality of x. Testing formally this hypothesis based on our non-parametric estimator is an interesting perspective for future investigation.

We finally conclude Section 4 by an application to the analysis of the spatio-temporal dynamics of proteins involved in exocytosis. For biological reasons, these proteins are visible at some time point (their birth) and disappear some time later (their death), their "life time" being closely related to their activity during the exocytosis process. Meanwhile, they may also follow some motion in the cell, which is the move step. We question several biological hypotheses, which imply that the birth intensity function should be constant whatever the current configuration of proteins is, and that the death intensity function could instead depend on the current activity. Our study confirms these hypotheses and further reveals the joint dynamics of two types of different proteins that seem to interact during the exocytosis mechanisms.

Section 5 and Appendix A contain the proofs of our theoretical results and technical lemmas. The codes and data used in Section 4 are available as supporting information in the GitHub repository at `https://github.com/lavancier-f/Birth-Death-Move-process`.

2 General definition and basic properties

2.1 Birth-death-move process

Let $(E_n)_{n\geq 0}$ be a sequence of disjoint polish spaces, each equipped with the Borel σ-algebra $\mathcal{E}_n$. We assume that E_0 consists of a single element, written $\varnothing$ for short, i.e. $E_0 = \{\varnothing\}$. Our state space is $E = \bigcup_{n=0}^{+\infty} E_n$, associated with the σ-field $\mathcal{E} = \sigma\left(\bigcup_{n=0}^{+\infty} \mathcal{E}_n\right)$. For $x \in E_n$, we denote $n(x) = n$, so that for any $x \in E$, $x \in E_{n(x)}$. We will sometimes call $n(x)$ the cardinality of x, in reference to the standard situation where E_n represents the space of point configurations in $\mathbb{R}^d$ with cardinality n.

For each n, we denote by $(Y_t^{(n)})_{t\geq 0}$ the process on E_n that drives the motion of the birth-death-move process $(X_t)_{t\geq 0}$ between two jumps. For each n, $(Y_t^{(n)})_{t\geq 0}$ is assumed to be a continuous homogeneous Markov process on E_n with transition kernel $(Q_t^{(n)})_{t\geq 0}$ on $E_n \times \mathcal{E}_n$, i.e. for any $x \in E_n$ and any $A \in \mathcal{E}_n$,

$$Q_t^{(n)}(x, A) = \mathbb{P}(Y_t^{(n)} \in A | Y_0^{(n)} = x).$$

To define the birth-death-move process $(X_t)_{t\geq 0}$, we need to introduce the birth intensity function $\beta : E \to \mathbb{R}^+$ and the death intensity function $\delta : E \to \mathbb{R}^+$, both assumed to be continuous on E. We prevent a death in E_0 by assuming that $\delta(\varnothing) = 0$. The intensity of the jumps of the process $(X_t)_{t\geq 0}$

is then $\alpha = \beta + \delta$. We will assume that α is bounded from below and above, i.e. there exist $\alpha_* > 0$ and $\alpha^* < \infty$ such that for every $x \in E$, $\alpha_* \leq \alpha(x) \leq \alpha^*$. Moreover, we need the probability transition kernel for a birth $K_\beta : E \times \mathcal{E} \to \mathbb{R}_+$ and for a death $K_\delta : E \times \mathcal{E} \to \mathbb{R}_+$. They satisfy, for all $x \in E$,

$$K_\beta(x, E_{n+1}) = \mathbf{1}_{x \in E_n} \quad \text{if } n \geq 0$$

and

$$K_\delta(x, E_{n-1}) = \mathbf{1}_{x \in E_n} \quad \text{if } n \geq 1.$$

The transition kernel for the jumps of $(X_t)_{t \geq 0}$ is then, for any $x \in E$ and any $A \in \mathcal{E}$,

$$K(x, A) = \frac{\beta(x)}{\alpha(x)} K_\beta(x, A) + \frac{\delta(x)}{\alpha(x)} K_\delta(x, A). \tag{1}$$

The intensity functions β, δ, α and the transition kernels K_β, K_δ, K play exactly the same role here than for a pure spatial birth-death process as introduced in Preston (1975).

Given an initial distribution μ_0 on E, the birth-death-move process $(X_t)_{t \geq 0}$ is then constructed as follows:

0) Generate $X_0 \sim \mu_0$.

1) Given $X_0 = z_0$, generate $(Y_t^{(n(z_0))})_{t \geq 0}$ conditional on $Y_0^{(n(z_0))} = z_0$ according to the kernel $(Q_t^{(n(z_0))}(z_0, .))_{t \geq 0}$. Then,

   - Given $X_0 = z_0$ and $(Y_t^{(n(z_0))})_{t \geq 0}$, generate the first inter-jump time τ_1 according to the cumulative distribution function

$$F_1(t) = 1 - \exp\left(- \int_0^t \alpha(Y_u^{(n(z_0))}) du \right).$$

   - Given $X_0 = z_0$, $(Y_t^{(n(z_0))})_{t \geq 0}$ and τ_1, generate the first post-jump location Z_1 according to the transition kernel $K\left(Y_{\tau_1}^{(n(z_0))}, \cdot\right)$.

   - Set $T_1 = \tau_1$, $X_t = Y_t^{(n(z_0))}$ for $t \in [0, T_1)$ and $X_{T_1} = Z_1$.

And iteratively, for $j \geq 1$,

j+1) Given $X_{T_j} = z_j$, generate $(Y_t^{(n(z_j))})_{t \geq 0}$ conditional on $Y_0^{(n(z_j))} = z_j$ according to $(Q_t^{(n(z_j))}(z_j, .))_{t \geq 0}$. Then

   - Given $X_{T_j} = z_j$ and $(Y_t^{(n(z_j))})_{t \geq 0}$, generate τ_{j+1} according to the cumulative distribution function

$$F_{j+1}(t) = 1 - \exp\left(- \int_0^t \alpha(Y_u^{(n(z_j))}) du \right). \tag{2}$$

   - Given $X_{T_j} = z_j$, $(Y_t^{(n(z_j))})_{t \geq 0}$ and τ_{j+1}, generate Z_{j+1} according to $K\left(Y_{\tau_{j+1}}^{(n(z_j))}, \cdot\right)$.

   - Set $T_{j+1} = T_j + \tau_{j+1}$, $X_t = Y_{t-T_j}^{(n(z_j))}$ for $t \in [T_j, T_{j+1})$ and $X_{T_{j+1}} = Z_{j+1}$.

In this construction $(T_j)_{j \geq 1}$ is the sequence of jump times of the process $(X_t)_{t \geq 0}$ and we set $T_0 = 0$. Note that by definition of the jump transition kernel K in (1), these jumps can only be a birth (a transition from E_n to E_{n+1}) or a death (a transition from E_n to E_{n-1} for $n \geq 1$).

We will adopt the following notation. The transition kernel of $(X_t)_{t \geq 0}$ is, for any $x \in E$ and $A \in \mathcal{E}$,

$$Q_t(x, A) = \mathbb{P}(X_t \in A | X_0 = x).$$

The number of jumps before t is denoted by $(N_t)_{t\geq 0}$, that is $N_t = Card\{j \geq 1 : T_j \leq t\}$. Similarly we denote by $(N_t^\beta)_{t\geq 0}$ and $(N_t^\delta)_{t\geq 0}$ the number of births and of deaths respectively, before t. For consistency, we shall sometimes denote N_t^α for N_t. Finally, we denote by $\mathcal{F}_t = \cap_{s>t}\sigma(X_u, u \leq s)$, $t > 0$, the natural right-continuous filtration of the process $(X_t)_{t\geq 0}$. To avoid any measurability issues, we further make the filtration $(\mathcal{F}_t)_{t>0}$ complete (Bass, 2011) and still abusively denote it by $(\mathcal{F}_t)_{t>0}$.

We verifies in Section 2.3 that $(X_t)_{t\geq 0}$ is a Markov process, along with additional basic properties.

2.2 Examples

The general definition above includes spatial birth-death processes, as introduced by Preston (1975). They correspond to $(Y_t^{(n)})_{t\geq 0}$ being the constant random variable $Y_t^{(n)} = Y_0^{(n)}$ for any $t \geq 0$ and any n, so that $(X_t)_{t\geq 0}$ does not move between two jumps.

A simple birth-death process is the particular case of a spatial birth-death process where $E_n = \{n\}$, so that $E = \mathbb{N}$ and $(X_t)_{t\geq 0}$ is interpreted as the evolution of a population size, each jump corresponding to the birth of a new individual or to the death of an existing one. In this case the intensity functions β and δ are just sequences, i.e. functions of n. The standard historical simple birth-death process, as introduced in Feller (1939) and Kendall (1949), corresponds to a constant birth rate and a constant death rate for each individual of the population, leading to linear functions of n for β and δ. The case of general sequences β and δ is considered in Reynolds (1973), who studied their estimation by maximum likelihood. This approach will be a particular case of our procedure, see Example 3 (i) in Section 3.

A spatial birth-death process in $\mathbb{R}^d$ corresponds to E_n being the set of point configurations in $\mathbb{R}^d$ with cardinality n. In this case, a birth consists of the emergence of a new point and a death to the disappearance of an existing point. This important special case is treated in details in Preston (1975) and further studied in Møller (1989). Special instances are discussed in Comas and Mateu (2008) and some applications to real data sets are considered in Møller and Sørensen (1994) and Sadahiro (2019) for example. Perfect simulation of spatial point processes moreover relies on these processes (Møller and Waagepetersen, 2004, Chapter 11).

Allowing each point of a spatial birth-death process in $\mathbb{R}^d$ to move according to a continuous Markov process leads to a spatial birth-death-move process in $\mathbb{R}^d$. A simple example is to assume that each point independently follows a Brownian motion in $\mathbb{R}^d$, which means that $(Y_t^{(n)})_{t\geq 0}$ is a vector of n independent Brownian motions. This is the dynamics (with $d = 2$) we will consider in our simulation study in Section 4. In Wang and Zhu (2002), a spatial birth-death-move dynamics in the plane has been adopted to model and track the joint trajectories of elements in a video, like snowflakes or flying birds. The motion in this application follows some autoregressive discrete-time process. In our real-data application in Section 4.2, we observe the location of proteins in a planar projection of a cell during some time interval. The motion can be different from a protein to another. Previous studies (Briane et al., 2019) have identified three main regimes: Brownian motion, superdiffusive motion (like a Brownian process with drift or a fractional Brownian motion with Hurst parameter greater than 1/2) and subdiffusive motion (like a Ornstein-Uhlenbeck process or a fractional Brownian motion with Hurst parameter less than 1/2). The process $(Y_t^{(n)})_{t\geq 0}$ in this application could then be a vector of n such processes, and some interactions between these n processes may further been introduced. In Section 4.2, we do not actually address the choice of a model for $(Y_t^{(n)})_{t\geq 0}$, but rather focus on the estimation of the intensity functions β and δ by the procedures developed in Section 3. Fortunately, no knowledge about the process $(Y_t^{(n)})_{t\geq 0}$ is needed for these estimations, except that it is a continuous Markov process.

Spatial birth-death-move processes also include spatio-temporal growth interaction models used in individual-based modelling in ecology. In this case E_n is the set of point configurations in $\mathbb{R}^2 \times \mathbb{R}_+$ with cardinality n, where $\mathbb{R}^2$ represents the space of location of the points (the plants in ecology) and $\mathbb{R}_+$ the space of their associated mark (the height of plants, say). Each birth in the process corresponds

to the emergence of a new plant in $\mathbb{R}^2$ associated with some positive mark. Through the birth kernel K_β, we may favor a new plant to arrive nearby existing ones (or not in case of competitions) and the new mark may be set to zero or be generated according to some specific distribution (Renshaw and Särkkä (2001) chose for instance a uniform distribution on $[0, \epsilon]$ for some small $\epsilon > 0$). The growth process only concerns the marks. Let us denote by $(U_i(t), m_i(t))_{t\geq 0}$, for $i = 1, \ldots, n$, each component of the process $(Y_t^{(n)})_{t\geq 0}$, to distinguish the location $U_i(t) \in \mathbb{R}^2$ to the mark $m_i(t) \in \mathbb{R}_+$ of a plant i. We thus have $U_i(t) = U_i(0)$ for all i, and some continuous Markov dynamics can be chosen for $(m_1(t), \ldots, m_n(t))$. In Renshaw and Särkkä (2001); Renshaw et al. (2009); Comas (2009); Häbel et al. (2019) several choices for this so-called growth interaction process are considered. Furthermore, while in the previous references the birth rate of each plant is constant, the death rate may depend on the location and size of the other plants, leading to a non trivial death intensity function δ.

2.3 Some theoretical properties

The next theorem proves that a birth-death-move process is a well-defined Markov process.

Theorem 1. *Let $(X_t)_{t\geq 0}$ be a spatial birth-death-move process as defined in Section 2.1. Then $(X_t)_{t\geq 0}$ is a homogeneous Markov process.*

Let $\beta_n = \sup_{x\in E_n} \beta(x)$ and $\delta_n = \inf_{x\in E_n} \delta(x)$. We verify in the next proposition that $(X_t)_{t\geq 0}$ converges to an invariant measure, uniformly over the initial configurations, under the following assumption.

(H1) For all $n \geq 1$, $\delta_n > 0$ and there exists n^* such that $\beta_n = 0$ for all $n \geq n^*$.

In light of this hypothesis, assumed in the rest of the paper, we henceforth redefine $E = \bigcup_{n=0}^{n^*} E_n$.

Proposition 2. *Under* (H1)*, $(X_t)_{t\geq 0}$ admits a unique invariant probability measure μ_∞ and there exist $a > 0$ and $c > 0$ such that for any measurable bounded function g and any $t > 0$,*

$$\sup_{y\in E} \left| \int_E g(z) Q_t(y, dz) - \int_E g(z)\mu_\infty(dz) \right| \leq a e^{-ct} \|g\|_\infty. \tag{3}$$

For a pure spatial birth-death process, this proposition is a consequence of Preston (1975) and Møller (1989). In these references, geometric ergodicity is also proven under a less restrictive setting than (H1), provided the sequence (δ_n) compensates in a proper way the sequence (β_n) to avoid explosion. This generalisation seems more difficult to prove for a general birth-death-move process. Moreover, the uniformity in (3) only holds under (H1), even for pure spatial birth-death processes, and this is a crucial property needed to establish (5) in the next corollary.

Corollary 3. *Under* (H1)*, for any measurable bounded non-negative function g and any $t \geq 0$,*

$$\left| \mathbb{E}\left(\int_0^t g(X_s) ds \right) - t \int_E g(z)\mu_\infty(dz) \right| \leq \frac{a}{c} \|g\|_\infty (1 - e^{-ct}) \tag{4}$$

where a and c are the same positive constants as in (3)*. Moreover,*

$$\mathbb{V}\left(\int_0^t g(X_s) ds \right) \leq c_0 \|g\|_\infty \mathbb{E}\left(\int_0^t g(X_s) ds \right) \tag{5}$$

where c_0 is some positive constant independent of t and g.

Finally, we clarify the martingale properties of the counting processes N_t^α, N_t^β and N_t^δ. They will be at the heart of the statistical applications of the next section. We set $X_{s-} := \lim_{t\to s, t<s} X_t$.

Proposition 4. *Let γ be either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$. Then a left-continuous version of the intensity of N_t^γ with respect to $\mathcal{F}_t$ is $\gamma(X_{t^-})$. Moreover, for any measurable bounded function g, the process $(M_t)_{t\geq 0}$ defined by*

$$M_t = \int_0^t g(X_{s^-})[dN_s^\gamma - \gamma(X_s)ds]$$

is a martingale with respect to $\mathcal{F}_t$ and for all $t \geq 0$

$$\mathbb{E}(M_t^2) = \mathbb{E}\left(\int_0^t g^2(X_s)\gamma(X_s)ds\right). \tag{6}$$

The proof of Proposition 4 relies in particular on the next lemma that is of independent interest.

Lemma 5. *Let $N^* \sim \mathcal{P}(\alpha^* T)$ and $N_* \sim \mathcal{P}(\alpha_* T)$, where $\mathcal{P}(a)$ denotes the Poisson distribution with rate $a > 0$. Then for any $n \in \mathbb{N}$,*

$$\mathbb{P}(N^* \leq n) \leq \mathbb{P}(N_T \leq n) \leq \mathbb{P}(N_* \leq n).$$

3 Estimation of the intensity functions

3.1 Continuous time observations

Assume that we observe continuously the process (X_t) in the time interval $[0, T]$ for some $T > 0$. Let $(k_t)_{t\geq 0}$ be a family of non-negative functions on $E \times E$ such that $k^* := \sup_{x,y\in E} \sup_{t\geq 0} |k_t(x, y)| < \infty$. Some typical choices for (k_t) are discussed in the examples below.

Using the convention $0/0 = 0$, a natural estimator of $\alpha(x)$ for a given $x \in E$ is

$$\hat{\alpha}(x) = \frac{1}{\hat{T}(x)} \int_0^T k_T(x, X_{s^-})dN_s = \frac{1}{\hat{T}(x)} \sum_{j=1}^{N_T} k_T(x, X_{T_j^-}), \tag{7}$$

where

$$\hat{T}(x) = \int_0^T k_T(x, X_s)ds \tag{8}$$

is an estimation of the time spent by $(X_s)_{0\leq s\leq T}$ in configurations similar to x. In words, $\hat{\alpha}(x)$ counts the number of times $(X_s)_{0\leq s\leq T}$ has jumped when it was in configurations similar to x over the time spent in these configurations. Similarly, we consider the following estimators of $\beta(x)$ and $\delta(x)$:

$$\hat{\beta}(x) = \frac{1}{\hat{T}(x)} \int_0^T k_T(x, X_{s^-})dN_s^\beta = \frac{1}{\hat{T}(x)} \sum_{j=1}^{N_T} k_T(x, X_{T_j^-})\mathbf{1}_{\{\text{a birth occurs at } T_j\}},$$

$$\hat{\delta}(x) = \frac{1}{\hat{T}(x)} \int_0^T k_T(x, X_{s^-})dN_s^\delta = \frac{1}{\hat{T}(x)} \sum_{j=1}^{N_T} k_T(x, X_{T_j^-})\mathbf{1}_{\{\text{a death occurs at } T_j\}}.$$

The next theorem establishes the consistency of these estimators under the following assumptions.

(H2) Setting $v_T(x) = \int_E k_T(x, z)\mu_\infty(dz)$,

$$\lim_{T\to\infty} Tv_T(x) = \infty.$$

(H3) Let γ be either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$. Setting $w_T(x) = \frac{1}{v_T(x)} \int_E (\gamma(z) - \gamma(x)) k_T(x,z) \mu_\infty(dz)$,

$$\lim_{T \to \infty} w_T(x) = 0.$$

Before stating this theorem, let us consider some typical examples, where γ stands for either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$.

Example 1: If $(X_t)_{t \geq 0}$ is a pure birth-death process, corresponding to the case where there are no motions between its jumps, i.e. for any n the process $(Y_t^{(n)})_{t \geq 0}$ is a constant random variable, then $X_{T_j^-} = X_{T_{j-1}}$ and $\hat{T}(x)$ becomes a discrete sum, so that

$$\hat{\alpha}(x) = \frac{\sum_{j=0}^{N_T - 1} k_T(x, X_{T_j})}{\sum_{j=0}^{N_T - 1} (T_{j+1} - T_j) k_T(x, X_{T_j}) + (T - T_{N_T}) k_T(x, X_{T_{N_T}})}.$$

Similar simplifications occur in this case for $\hat{\beta}(x)$ and $\hat{\delta}(x)$.

Example 2: A standard choice for k_T is

$$k_T(x,y) = k\left(\frac{d(x,y)}{h_T}\right) \tag{9}$$

where k is a bounded kernel function on $\mathbb{R}$, d is a pseudo-distance on E and $h_T > 0$ is a bandwidth parameter. In this case, (H2) and (H3) can be understood as hypotheses on the bandwidth h_T, demanding that h_T tends to 0 as $T \to \infty$ but not too fast, as usual in non-parametric estimation. To make this interpretation clear, assume that $k(u) = \mathbf{1}_{|u|<1}$ and that γ is Lipschitz with constant ℓ. Then denoting $B(x, h_T) := \{y \in E, d(x,y) < h_T\}$, we have that $v_T(x) = \mu_\infty(B(x, h_T))$ and

$$|w_T| \leq \frac{1}{v_T(x)} \int_{B(x,h_T)} |\gamma(z) - \gamma(x)| \, \mu_\infty(dz) \leq \frac{\ell}{v_T(x)} \int_{B(x,h_T)} d(x,z) \mu_\infty(dz) \leq \frac{\ell}{v_T(x)} h_T v_T(x) = \ell h_T.$$

In this setting, (H2) and (H3) are satisfied whenever $h_T \to 0$ and $T \mu_\infty(B(x, h_T)) \to \infty$.

Example 3: If we assume that $\gamma(x) = \gamma_0(n(x))$ only depends on the cardinality of x through some function γ_0 defined on $\mathbb{N}$, the setting becomes similar to simple birth-death processes, except that we allow continuous motions between jumps. We may consider two strategies in this case:

(i) We recover the standard non-parametric likelihood estimator of the intensity studied in Reynolds (1973) by choosing $k_T(x,y) = 1$ if $n(x) = n(y)$ and $k_T(x,y) = 0$ otherwise. Then $\hat{\alpha}(x)$ (resp. $\hat{\beta}(x)$, $\hat{\delta}(x)$) just counts the number of jumps (resp. of births, of deaths) of the process $(X_s)_{0 \leq s \leq T}$ when it is in $E_{n(x)}$ divided by the time spent by the process in $E_{n(x)}$. We get in this case that $v_T(x) = \mu_\infty(E_{n(x)})$ does not depend on T and $v_T(x) w_T(x) = 0$, so (H2) and (H3) are satisfied whenever $\mu_\infty(E_{n(x)}) \neq 0$.

(ii) Alternatively, we may choose k_T as in (9) with $d(x,y) = |n(x) - n(y)|$, in which case $\hat{\gamma}(x)$ differs from the previous estimator in that not only configurations in $E_{n(x)}$ are taken into account in $\hat{\gamma}(x)$ but all configurations in E_n provided n is close to $n(x)$. For this reason this new estimator can be seen as a smoothing version of the previous one and is less variable (see the simulation study of Section 4). It makes sense if we assume some regularity properties on $n \mapsto \gamma_0(n)$, as demanded by (H3).

Example 4: When E_n is the space of point configurations in $\mathbb{R}^d$ with cardinality n, we can take k_T as in (9) where $d(x,y)$ is a distance on the space E of finite point configurations in $\mathbb{R}^d$. Several choices for this distance are possible. A first standard option is the Hausdorff distance

$$d_H(x,y) = \max\{\max_{u\in x}\min_{v\in y}\|u-v\|, \max_{v\in y}\min_{u\in x}\|v-u\|\}$$

if $x \neq \varnothing$ and $y \neq \varnothing$, while $k_T(x,y) = \mathbf{1}_{x=y}$ if $x = \varnothing$ or $y = \varnothing$. Some alternatives are discussed in Mateu et al. (2015). Another option, that will prove to be more appropriate for our applications in Section 4, has been introduced by Schuhmacher and Xia (2008). Letting $x = \{u_1, \ldots, u_{n(x)}\}$, $y = \{v_1, \ldots, v_{n(y)}\}$ and assuming that $n(x) \leq n(y)$, this distance is defined for some $\kappa > 0$ by

$$d_\kappa(x,y) = \frac{1}{n(y)}\left(\min_{\pi\in\mathfrak{S}_{n(y)}}\sum_{i=1}^{n(x)}(\|u_i - v_{\pi(i)}\| \wedge \kappa) + \kappa(n(y)-n(x))\right),$$

where $\mathfrak{S}_{n(y)}$ denotes the set of permutations of $\{1,\ldots,n(y)\}$. In words, d_κ calculates the total (truncated) distance between x and its best match with a sub-pattern of y with cardinality $n(x)$, and then it adds a penalty κ for the difference of cardinalities between x and y. If all point configurations belong to a bounded subset W of $\mathbb{R}^d$, a natural choice for κ is to take the diameter of W, in which case the distance between two point patterns with the same cardinalities corresponds to the average distance between their optimal matching. The definition of d_κ when $n(x) \geq n(y)$ is similar by inverting the role played by x and y. Contrary to Example 3, the choice of d_H or d_κ does not exploit any particular structural form of $\gamma(x)$, allowing for a pure non-parametric estimation. Note however that some regularities are implicitly demanded on $\gamma(x)$ because of (H3), as illustrated in Example 2 where γ is assumed to be Lipschitz.

Proposition 6. *Let γ be either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$. Assume* (H1), (H2) *and* (H3), *then*

$$\hat{\gamma}(x) - \gamma(x) = O_p\left(\frac{1}{Tv_T(x)} + w_T^2(x)\right)$$

as $T \to \infty$, whereby $\hat{\gamma}(x)$ is a consistent estimator of $\gamma(x)$.

From (4) and (5) applied to $g(X_s) = k_T(x, X_s)$ and $t = T$, we deduce using (H2) that for any $T \geq 0$

$$\mathbb{E}(\hat{T}(x)) \sim Tv_T(x) \quad \text{and} \quad \mathbb{V}\left(\hat{T}(x)\right) \leq c_0\mathbb{E}(\hat{T}(x)). \tag{10}$$

Even if $\mathbb{E}(\hat{T}(x)) \to \infty$, $\hat{T}(x)$ may take infinitely small values for some x, in which case $\hat{\gamma}(x)$ may be arbitrarily large. This implies that $\hat{\gamma}(x)$ is not integrable in general, as illustrated in the next lemma, which explains why Proposition 6 considers convergence in probability and not the second order properties of $\hat{\gamma}(x)$.

Lemma 7. *Assume that $\alpha(.) = \alpha_0(n(.))$ for some function α_0 and let $k_T(x,y) = \mathbf{1}_{n(x)=n(y)}$ (which is the situation of Example 3 (i) with $\gamma = \alpha$). Let $x \in E$ be such that $\mu_\infty(E_{n(x)}) \neq 0$, so that* (H2) *and* (H3) *are satisfied. If moreover $\mu_0(E_{n(x)}) \neq 0$, then $\mathbb{E}(\hat{\alpha}(x)) = \infty$.*

Arguably, a statistician would not trust any estimation of $\gamma(x)$ if $\hat{T}(x)$ is very small and in our opinion the result of this lemma does not rule out using $\hat{\gamma}(x)$ for reasonable configurations x, that are configurations for which a minimum time has been spent by the process in configurations similar to x. To reflect this idea, let us consider the following modified estimator, for a given small $\epsilon > 0$,

$$\hat{\gamma}_\epsilon(x) = \hat{\gamma}(x)\mathbf{1}_{\hat{T}(x)>\epsilon}.$$

This alternative estimator obviously satisfies Proposition 6 but it has the pleasant additional property to be mean-square consistent, as stated in the following proposition.

Proposition 8. *Let γ be either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$. Assume* (H1), (H2) *and* (H3), *then for all $\epsilon > 0$ and all $0 < \eta < 1/3$,*

$$\mathbb{E}\left[(\hat{\gamma}_\epsilon(x) - \gamma(x))^2\right] = O\left(\frac{1}{(Tv_T(x))^{1/3-\eta}} + w_T^2(x)\right)$$

as $T \to \infty$, whereby $\hat{\gamma}_\epsilon(x)$ is mean-square consistent.

3.2 Discrete time observations

Assume now that we observe the process at $m+1$ time points $t_0, \ldots, t_m$ where $t_0 = 0$ and $t_m = T$. We denote $\Delta t_j = t_j - t_{j-1}$, for $j = 1, \ldots, m$, and $\Delta_m = \max_{j=1\ldots m} \Delta t_j$ the maximal discretization step. We thus have $T = \sum_{j=1}^m \Delta t_j$ and $T \leq m\Delta_m$. We assume further that $m\Delta_m/T$ is uniformly bounded. The asymptotic properties of this section will hold when both $T \to \infty$ and $\Delta_m \to 0$, implying $m \to \infty$.

To consider a discrete version of the estimator (7) of $\alpha(x)$, a natural idea is to use the number of jumps between two observations. We denote this number by $\Delta N_{t_j}^\alpha = N_{t_j} - N_{t_{j-1}}$, for $j = 1, \ldots, m$. However, $\Delta N_{t_j}^\alpha$ is not necessarily observed because we can miss some jumps in the interval $(t_{j-1}, t_j]$ (for instance the birth of an individual can be immediately followed by its death). For this reason we introduce an approximation D_j^α of $\Delta N_{t_j}^\alpha$ which is observable. Similarly we need an approximation D_j^β and D_j^δ of $\Delta N_{t_j}^\beta$ and $\Delta N_{t_j}^\delta$, using obvious notation. For the asymptotic validity of our estimator, we specifically require that for either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$, $(D_j^\gamma)_{j\geq 1}$ is a sequence of random variables taking values in $\mathbb{N}$ and satisfying:

(H4) For all $j \geq 1$, $D_j^\gamma = \Delta N_{t_j}^\gamma$ if $\Delta N_{t_j}^\gamma \leq 1$ and $D_j^\gamma \leq \Delta N_{t_j}^\gamma$ if $\Delta N_{t_j}^\gamma \geq 2$.

When $\Delta_m \to 0$, the case $\Delta N_{t_j}^\gamma \geq 2$ becomes unlikely and this is the reason why $\Delta N_{t_j}^\gamma$ can be poorly approximated in this case when considering asymptotic properties. When $\gamma = \alpha$, an elementary example is to take $D_j^\alpha = \mathbf{1}_{n(X_{t_{j-1}}) \neq n(X_{t_j})}$. This choice satisfies (H4) but some better approximations may be available. For instance, when E_n is the space of point configurations in $\mathbb{R}^d$ with cardinality n, assuming that we can track the points between times t_{j-1} and t_j, we can choose for D_j^α the number of new points in X_{t_j} (that is a lower estimation of the number of births), plus the number of points that have disappeared between t_{j-1} and t_j (that is a lower estimation of the number of deaths). Obvious adaptations lead to the same remark for the choice of D_j^β and D_j^δ.

Our estimator of $\gamma(x)$ in the discrete case, where γ is either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$, is then defined as follows:

$$\hat{\gamma}_{(d)}(x) = \frac{\sum_{j=0}^{m-1} D_{j+1}^\gamma k_T(x, X_{t_j})}{\sum_{j=0}^{m-1} \Delta t_{j+1} k_T(x, X_{t_j})}. \tag{11}$$

To get its consistency, we assume that the discretization step Δ_m asymptotically vanishes at the following rate.

(H5) Let $v_T(x)$ be as in (H2),

$$\lim_{T\to\infty} \frac{\Delta_m}{v_T^2(x)} \to 0.$$

We also need some smoothness assumptions on the paths of $Y^{(n)}$ for all $n \in \mathbb{N}$. This is necessary to control the difference between $\hat{T}(x)$ defined in (8) and its discretized version $\hat{T}_{(d)}(x) = \sum_{j=0}^{m-1} \Delta t_{j+1} k_T(x, X_{t_j})$ appearing in the denominator of $\hat{\gamma}_{(d)}(x)$.

(H6) For all n, the function $s \mapsto k_T(x, Y_s^{(n)})$ is a-Hölderian with constant $\ell_T(x) \geq 0$, i.e. $|k_T(x, Y_s^{(n)}) - k_T(x, Y_t^{(n)})| \leq \ell_T(x)|s-t|^a$, such that

$$\lim_{T\to\infty} \frac{\Delta_m^a \ell_T(x)}{v_T^2(x)} \to 0.$$

Example 1 (continued): If $(X_t)_{t\geq 0}$ is a pure birth-death process, then (H6) is obviously satisfied with $\ell_T(x) = 0$.

Example 2 (continued): Assume k_T takes the general form (9) and that both $u \mapsto k(u)$ and $s \mapsto Y_s^{(n)}$ are Hölderian with respective exponent a_k, a_Y and respective constant c_k, c_Y. Then we get in (H6) $\ell_T(x) = c_k c_Y / h_T^{a_k}$ and $a = a_k + a_Y$. For instance, if k is moreover assumed to be supported on $[-1,1]$ and to satisfy $k(u) \geq k_* \mathbf{1}_{|u|<c}$ for some $k_* > 0$ and $0 < c < 1$, the same interpretation of (H2) and (H3) than for the choice $k(u) = \mathbf{1}_{|u|<1}$ remains valid for a Lipschitz function γ. We obtain in this case that (H2)-(H3) and (H5)-(H6) hold true whenever $h_T \to 0$, $T\mu_\infty(B(x, ch_T)) \to \infty$, $\Delta_m = o(\mu_\infty^2(B(x, ch_T))$ and $\Delta_m^a = o(h_T^{a_k} \mu_\infty^2(B(x, ch_T))$.

Example 3 (i) (continued): If we assume that $\gamma(x) = \gamma_0(n(x))$ and choose $k_T(x,y) = \mathbf{1}_{n(x)=n(y)}$ to recover the standard non-parametric likelihood estimator of Reynolds (1973), then we can take $\ell_T(x) = 0$ in (H6) so that (H2)-(H3) and (H5)-(H6) are satisfied whenever $\mu_\infty(E_{n(x)}) \neq 0$ and $\Delta_m \to 0$.

Proposition 9. *Let γ be either $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$. Assume* (H1)-(H6)*, then*

$$\hat{\gamma}_{(d)}(x) - \gamma(x) = O_p\left(\frac{1}{T v_T(x)} + w_T^2(x) + \frac{\Delta_m}{v_T^2(x)} + \frac{\Delta_m^a \ell_T(x)}{v_T^2(x)}\right)$$

whereby $\hat{\gamma}_{(d)}(x)$ is a consistent estimator of $\gamma(x)$.

As in Proposition 8, it would be possible to extend this convergence in probability to a mean-square convergence, provided we consider the modified estimator $\hat{\gamma}_{(d)}(x)\mathbf{1}_{\hat{T}_{(d)}(x)>\epsilon}$ for some $\epsilon > 0$. We do not provide the details.

3.3 Bandwidth selection by likelihood cross-validation

We assume in this section that k_T takes the form (9), in which case we have to choose in practice a value of the bandwidth h_T to implement our estimators, whether for $\hat{\gamma}(x)$ in the continuous case or for $\hat{\gamma}_{(d)}(x)$ in the discrete case. We explain in the following how to select h_T by likelihood cross-validation, a widely used procedure in kernel density estimation of a probability distribution and intensity kernel estimation of a point process, see for instance (Loader, 2006, Chapter 5). Remark that the alternative popular plug-in method and least-squares cross-validation method, that are both based on second order properties of the estimator, do not seem adapted to our setting because $\hat{\gamma}(x)$ is not necessarily integrable as showed in Lemma 7. Furthermore, even for the square-integrable modified estimator $\hat{\gamma}_\epsilon(x)$, Proposition 8 only provides a bound for the rate of convergence, and this one depends on unknown quantities that appear difficult to estimate. We focus for simplicity on the estimation of $\alpha(x)$ but the procedure adapts straightforwardly to the estimation of $\beta(x)$ or $\delta(x)$.

By Proposition 4, the intensity of N_t with respect to $\mathcal{F}_t$ is $\alpha(X_{t^-})$. By Girsanov theorem (Brémaud, 1981, Chapter 6.2), the log-likelihood of $(N_t)_{0\leq t\leq T}$ with respect to the unit rate Poisson counting process on $[0,T]$ is therefore

$$\int_0^T (1 - \alpha(X_{s^-}))ds + \int_0^T \log\alpha(X_{s^-})dN_s = T - \int_0^T \alpha(X_s)ds + \sum_{j=1}^{N_T} \log\alpha(X_{T_j^-}).$$

For continuous time observations, bandwidth selection by likelihood cross-validation amounts to choose h_T as

$$\hat{h} = \underset{h}{\operatorname{argmax}} \sum_{j=1}^{N_T} \log \hat{\alpha}_h^{(-)}(X_{T_j^-}) - \int_0^T \hat{\alpha}_h^{(-)}(X_s)ds$$

where $\hat{\alpha}_h^{(-)}(X_s)$ is the estimator (7) of $\alpha(x)$ for $x = X_s$, associated to the choice of bandwidth $h_T = h$, but without using the observation X_s. To carry out this removal, we suggest to discard all observations in the time interval $[T_{N_s}, T_{N_s+1}]$, which gives

$$\hat{\alpha}_h^{(-)}(X_s) = \frac{\sum_{i=1, i\neq N_s+1}^{N_T} k(d(X_s, X_{T_i^-})/h)}{\int_{[0,T]\setminus[T_{N_s}, T_{N_s+1}]} k(d(X_s, X_u)/h)du}.$$

In particular

$$\hat{\alpha}_h^{(-)}(X_{T_j^-}) = \frac{\sum_{i=1, i\neq j}^{N_T} k(d(X_{T_j^-}, X_{T_i^-})/h)}{\int_{[0,T]\setminus[T_{j-1}, T_j]} k(d(X_{T_j^-}, X_u)/h)du}.$$

For discrete time observations, this cross-validation procedure becomes

$$\hat{h}_{(d)} = \underset{h}{\operatorname{argmax}} \sum_{j=0}^{m-1} D_{j+1}^{\alpha} \log \hat{\alpha}_{(d),h}^{(-)}(X_{t_j}) - \sum_{j=0}^{m-1} \Delta t_{j+1} \hat{\alpha}_{(d),h}^{(-)}(X_{t_j}),$$

using the same notation as in Section 3.2 and where

$$\hat{\alpha}_{(d),h}^{(-)}(X_{t_j}) = \frac{\sum_{i=0, i\neq j}^{m-1} D_{i+1}^{\alpha} k(d(X_{t_j}, X_{t_i})/h)}{\sum_{i=0, i\neq j}^{m-1} \Delta t_{i+1} k(d(X_{t_j}, X_{t_i})/h)}.$$

4 Applications

4.1 Simulations

In order to assess the performances of our intensity estimator, we simulate a birth-death-move process in the square window $W = [0,1]^2$ during the time interval $[0,T]$ with $T = 1000$ (the time unit does not matter). The initial configuration at $t = 0$ consists of 108 points uniformly distributed in W. For the jump intensity function α, we choose $\alpha(x) = \exp(5(n(x)/100 - 1))$, i.e. $\alpha(x)$ only depends on the cardinality of x, as in the setting of Example 3, and this dependence is exponential. We fix a truncation value $n^* = 1000$. If $0 < n(x) < n^*$, each jump is a birth or a death with equal probability. If $n(x) = 0$ the jump is a birth, and it is a death if $n(x) = n^*$. Each birth consists of the addition of one point which is drawn uniformly on $[0,1]^2$, and each death consists of the removal of an existing point uniformly over the points of x. Finally, between each jump, each point of x independently evolves according to a planar Brownian motion with standard deviation 2.10^{-3}. Many other dynamics could have been simulated. Our choice is motivated by the real-data dynamics treated in the next section, that we try to roughly mimic.

Figure 1 shows, for one simulated trajectory, the state of the process at some jump times. For this simulation, $N_T = 1530$ jumps have been observed in the time interval $[0,T]$. The top-left plot of Figure 1 is the initial configuration and the state after the first jump T_1 is visible next to it. This first jump was a birth and the location of the new point is indicated by a red dot, pointed out by a red arrow. The locations of the other points at time T_1 are indicated by black dots, while their initial locations are recalled in gray, illustrating the motions between T_0 and T_1.

For this simulation, we want to estimate the intensity function $\alpha(x)$ for each $x = X_{T_i}$, $i = 0, \ldots, N_T$. This objective is again motivated by the real-data application of the next section, where the main

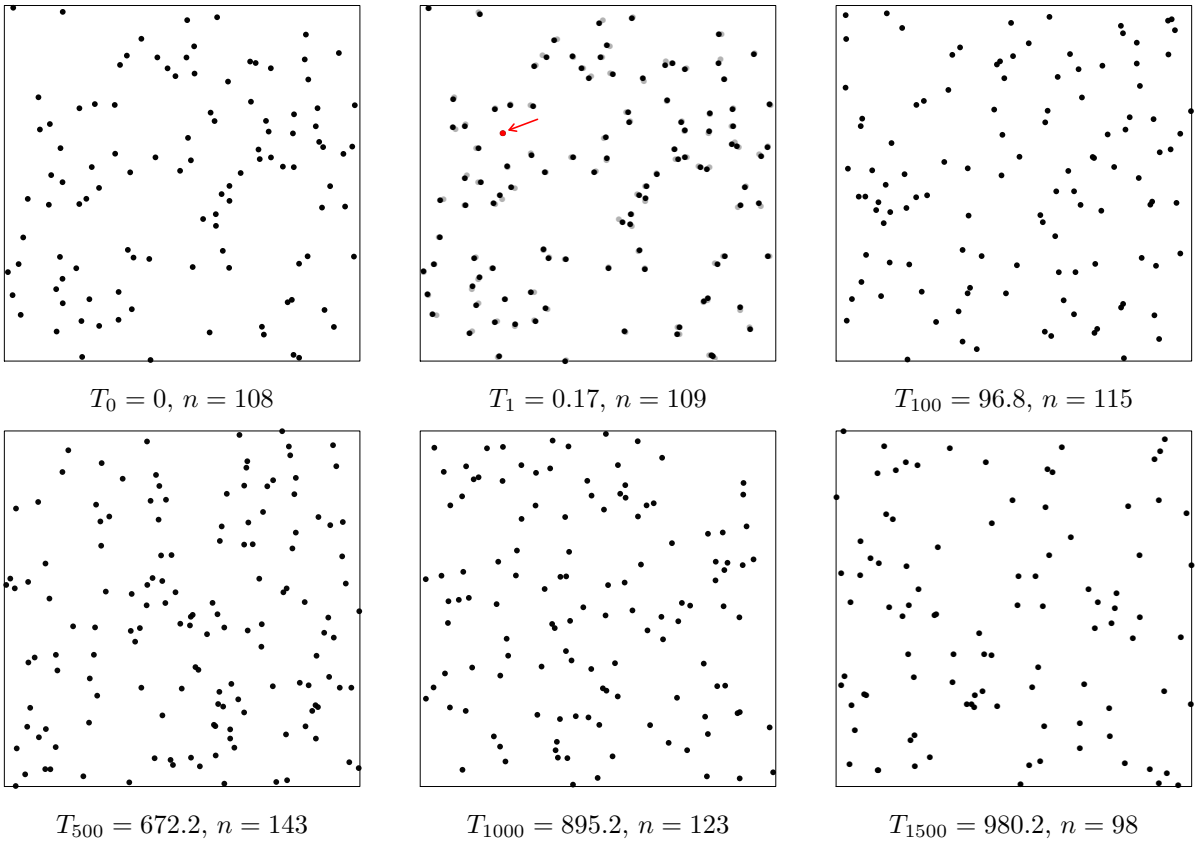


Figure 1: Realisations of the simulated process during the time interval $[0, 1000]$ at jump times T_0, T_1, T_{100}, T_{500}, T_{1000} and T_{1500} with the indication of their number of points n. For this simulation 1530 jumps have been observed. The top-middle plot at the first jump time T_1 shows the new born point in red (pointed out by a red arrow); the other points show as black dots while their initial locations at time T_0 are recalled in gray.

question will be to analyse the temporal evolution of the intensity. Assuming for the moment that we observe the process continuously in $[0, T]$, we consider the estimator (7) and the following four different choices for k_T. The first two strategies are fully non-parametric and follow Example 4: the first one depends on the Hausdorff distance d_H, while the second one depends on the distance d_κ where we have chosen $\kappa = \sqrt{2}$ to be the diameter of W. The two other strategies assume (rightly) that $\alpha(x)$ only depends on the cardinality of x. They respectively correspond to the cases (i) and (ii) of Example 3. Each time needed, we select the bandwidth by likelihood cross-validation as explained in Section 3.3. The result of these four estimations are showed in Figure 2. The first row depicts the evolution of the true value of $\alpha(X_{T_i})$ in black, for $i = 0, \dots, N_T$, along with its estimation in red. The second row shows the scatterplot $(n(X_{T_i}), \hat{\alpha}(X_{T_i}))$ along with the ground-truth curve $\exp(5(n(X_{T_i})/100 - 1))$.

From the two left scatterplots of Figure 2, we see that the two non-parametric estimators based on d_H and d_κ are able to detect a dependence between $\alpha(x)$ and $n(x)$. Between them, the estimator based on d_κ seems by far more accurate. Concerning the two other estimators that assume a dependence in $n(x)$, the second one is much less variable. As indicated in Example 3, this is because this estimator exploits the underlying regularity properties of $\alpha(x)$ in $n(x)$, unlike the other one. It is remarkable that the non-parametric estimator based on d_κ, because it takes advantage of the smoothness of $\alpha(x)$, achieves better performances than the third estimator, although this one assumes the dependence in $n(x)$. All in all, the best results come from the last estimator, defined in Example 3-(ii), that takes

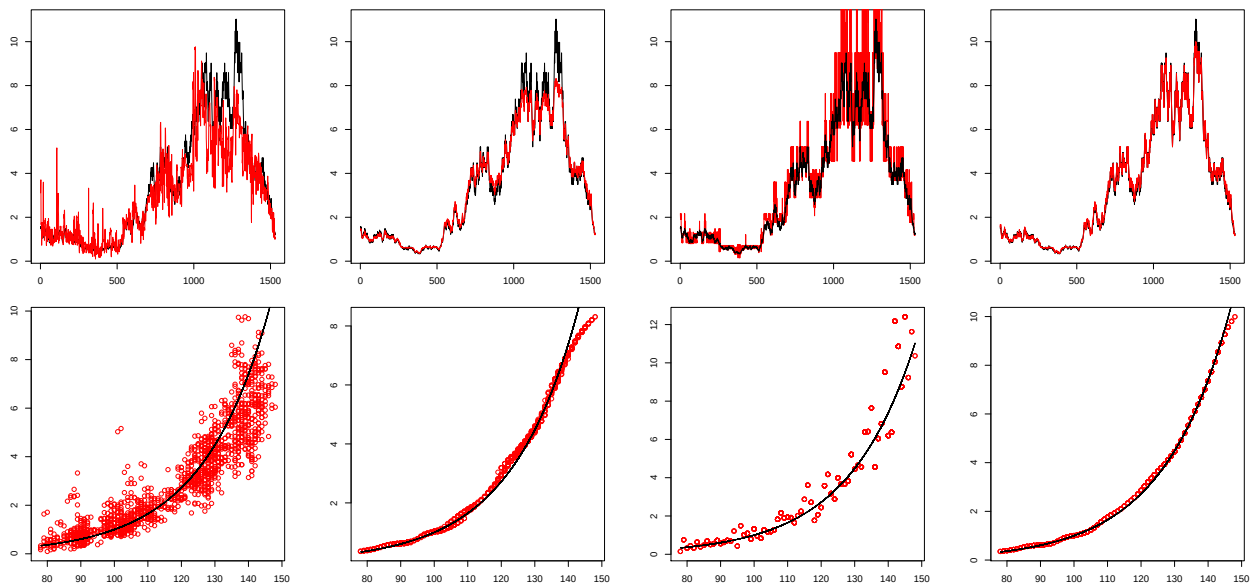


Figure 2: First row: true value of $\alpha(X_{T_i})$ in black, for $i = 0, \dots, N_T$, along with its estimation in red. The estimator is (7) with, from left to right, the two choices of k_T specified in Example 4, i.e. based on the distances d_H and d_κ, and the two choices of k_T proposed in Example 3 (i) and (ii). Second row: scatterplots $(n(X_{T_i}), \hat{\alpha}(X_{T_i}))$ along with the ground-truth curve $\exp(5(n(X_{T_i})/100 - 1))$ for the same 4 estimators.

advantage of both the dependence in $n(x)$ and the smoothness of $\alpha(x)$. These conclusions are confirmed by the mean squared errors reported in the first row of Table 1, computed from the estimation of $\alpha(x_j)$ at 100 point configurations $x_j = X_{T_j}$ for j regularly sequenced from 1 to $N_T = 1530$.

In order to assess the effect of discretisation, we consider $m + 1$ observations X_{t_j} taken from the above simulated continuous trajectory, regularly spaced from $t_0 = 0$ to $t_m = T = 1000$. This implies an average of $N_T/m = 1530/m$ jumps between two observations. We then apply the estimator (11) of $\alpha(x)$ at the same 100 point configurations x_j as above and the same four choices of k_T, based on the m observations. For the approximation D_j^α of the number of jumps between two observations $X_{t_{j-1}}$ and X_{t_j}, we take the number of new points observed in X_{t_j} plus the number of points having disappeared from $X_{t_{j-1}}$. The mean square errors of the results, for different values of m, are reported in Table 1, along with their standard deviations. Note that for some configurations x_j, there was no observation with cardinality $n(x_j)$ making impossible the computation of the third estimator, which explains the

	d_H	d_κ	Ex. 3 (i)	Ex. 3 (ii)
Continuous time obs.	151 (34)	18 (8)	93 (21)	1.8 (1.1)
Discrete time, $m + 1 = 5000$	266 (62)	18 (8)	141 (44)	3.0 (1.9)
Discrete time, $m + 1 = 1000$	226 (56)	20 (9)	NA	4.1 (1.8)
Discrete time, $m + 1 = 100$	376 (94)	36 (16)	NA	36 (19)
Discrete time, $m + 1 = 30$	767 (131)	182 (48)	NA	128 (37)

Table 1: Mean square errors of the estimation of $\alpha(x_j)$ at 100 different point configurations x_j, along with their standard deviations in parenthesis. The same four estimators as in Figure 2 are considered. The estimation is based on the observation of the full trajectory (continuous time observations) in $[0, T]$ with $T = 1000$, or on $m+1$ observations regularly spaced in this interval (discrete time observations), implying $N_T/m = 1530/m$ jumps in average between each obervation.

presence of some NA's in the table. As seen from Table 1, the comparison between the four estimators are in line with the continuous case and their performances increase with m, as could be expected. For illustration, Figure 3 compares the true values of $\alpha(x_j)$ for $j = 1, \ldots, 100$ with their estimation by the second estimator based on d_κ (in blue) and the last estimator (in red), for the same different values of m as in Table 1. While the quality of estimation degrades when m decreases, it remains quite decent even for small values of m, in particular when $m + 1 = 100$ implying more than 15 jumps in average between each observation.

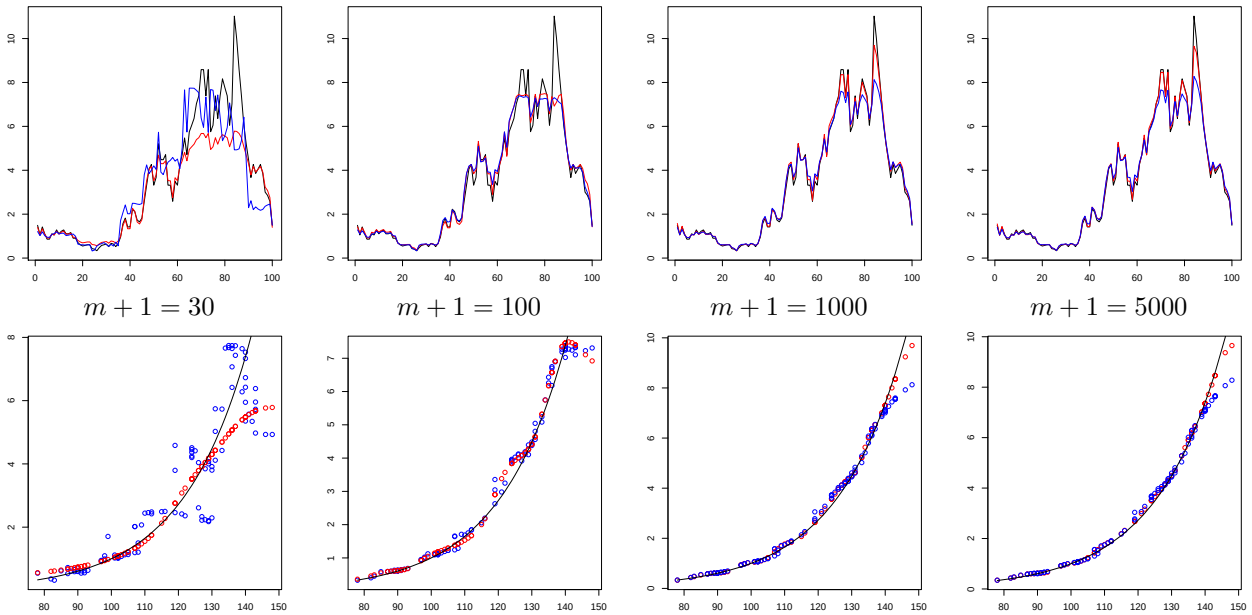


Figure 3: First row: true value of $\alpha(x_j)$ in black, for 100 different x_j extracted from the simulated continuous trajectory, along with, in blue: the discrete time estimation given by (11) based on d_κ; and in red: the discrete time estimation given by (11) with the choice of k_T as in Example 3-(ii). From left to right: the number of observations is $m + 1 = 30$, $m + 1 = 100$, $m + 1 = 1000$ and $m + 1 = 5000$, implying $N_T/m = 1530/m$ jumps in average between each observation. Second row: scatterplots $(n(x_j), \hat{\alpha}(x_j))$ for the same two estimators and the same values of m, along with the ground-truth curve $\exp(5(n(x_j)/100 - 1))$ in black.

4.2 Data analysis

Our data consist of a sequence of 1199 frames showing the locations of two types of proteins inside a living cell, namely Langerin and Rab11 proteins, both being involved in exocytosis mechanisms in cells. The total length of the sequence is 171 seconds for a 140 ms time interval between each frame. These images have been acquired by 3D multi-angle TIRF (total internal reflection fluorescence) microscopy technique (Boulanger et al., 2014), and we observe projections along the z-axis on a 2D plane close to the plasma membrane of the cell. The raw sequence can be seen online as part of the supporting information. As a result of this acquisition, we observe hundreds of proteins of each type on each frame following some random motions, while some new proteins appear at some time point and others disappear. The reason why a protein becomes visible can be simply due to its appearance into the axial resolution of the microscope, or because it becomes fluorescent only at the last step of the exocytosis process due to the pH change close to the plasma membrane. Similarly, a protein disappears from the image when it exits the axial resolution or when it disaggregates after the exocytosis process. Between its appearance and disappearance, the dynamics of a protein depends on its function and

its environment. The whole spatio-temporal process at hand is thus composed of multiple fluorescent spots appearing, moving and disappearing over time, all of them in interaction with each other. The underlying biological challenge is to be able to decipher this complex spatio-temporel dynamics, and in particular to understand the interaction between the different types of involved proteins, in the present case Langerin and Rab11 proteins (Gidon et al., 2012; Boulanger et al., 2014). Existing works either study the trajectories of each individual protein, independently to the other proteins (Briane et al., 2019; Pécot et al., 2018), or investigate the interaction between different types of proteins frame by frame (which is the co-localization problem), without temporal insight (Costes et al., 2004; Bolte and Cordelieres, 2006; Lagache et al., 2015; Lavancier et al., 2019). As far as we know, the present approach is the first attempt to tackle the joint spatio-temporal dynamics of two types of proteins involved in exocytosis mechanisms.

To analyse the data, we do not consider the raw sequence but the post-processed sequence introduced in Pécot et al. (2008, 2014), leading to more valuable data, where the most relevant regions of the cell, corresponding to the locations of the most dynamical proteins, have been enhanced thanks to a specific filtering procedure. The post-processed sequence for each type of protein is available online, see the supporting information. We then apply the U-track algorithm developed in Jaqaman et al. (2008) in order to track over time the locations of proteins. For both types of them, the result is a sequence of 1199 point patterns that follow a certain birth-death-move dynamics for the reasons explained earlier. Figure 4 shows the repartition of Langerin (resp. Rab11) proteins in the first (resp. second) row, for a few frames extracted from these two sequences. The two leftmost plots correspond to the two first frames: we observe that a new Langerin protein and two new Rab11 proteins, visible in red, appeared between times $t_0 = 0$ and $t_1 = 0.14s$. In the second plot, we also recalled the initial positions of proteins as gray dots. Close inspection reveals that the proteins have slightly moved between the two frames. This motion is more apparent on the full sequence available in the supporting information. For the Langerin sequence, we observe 21 to 76 proteins per frame (36.4 in average) and 1.26 jumps in average between each frame, 50.7% of which being deaths and 49.3% being births. For the Rab11 sequence, there are 10 to 52 proteins per frame (22.3 in average) and 0.85 jumps in average between each frame, 50.6% being deaths and 49.4% being births.

Based on these observations, we would like to question several biological hypotheses:

i) The first one assumes that each protein may appear at any time independently on the configuration and the number of proteins already involved in the exocytosis process. This would imply a constant birth intensity function over the whole sequence, for both types of proteins.

ii) The second hypothesis is that each protein may disappear independently of the others after its exocytosis process. Accordingly, the death intensity function at a configuration x should then depend on the cardinality of x, that is on the number of currently active proteins.

iii) The third hypothesis is that Langerin and Rab11 proteins interact during the exocytosis process, which should imply a correlation between their respective intensities.

We estimate separately the birth intensity function and the death intensity function of both sequences thanks to our estimator (11), where for D_j^β (resp. D_j^δ) we take the observed number of new proteins having appeared (resp. the number of proteins having disappeared) between frames $j-1$ and j. Motivated by the simulation study conducted in the previous section, we consider the non-parametric estimator based on the distance d_κ, where κ is the diameter of the observed cell, and the estimator from Example 3-(ii) that assumes the intensities only depend on the cardinality.

For the birth intensities, both estimators agree on a constant value of 4.45 births per second for the Langerin sequence and 2.98 births per second for the Rab11 sequence. This constant values result from the choice of a large value of the bandwidths by cross-validation, and they are in agreement with the first biological hypothesis.

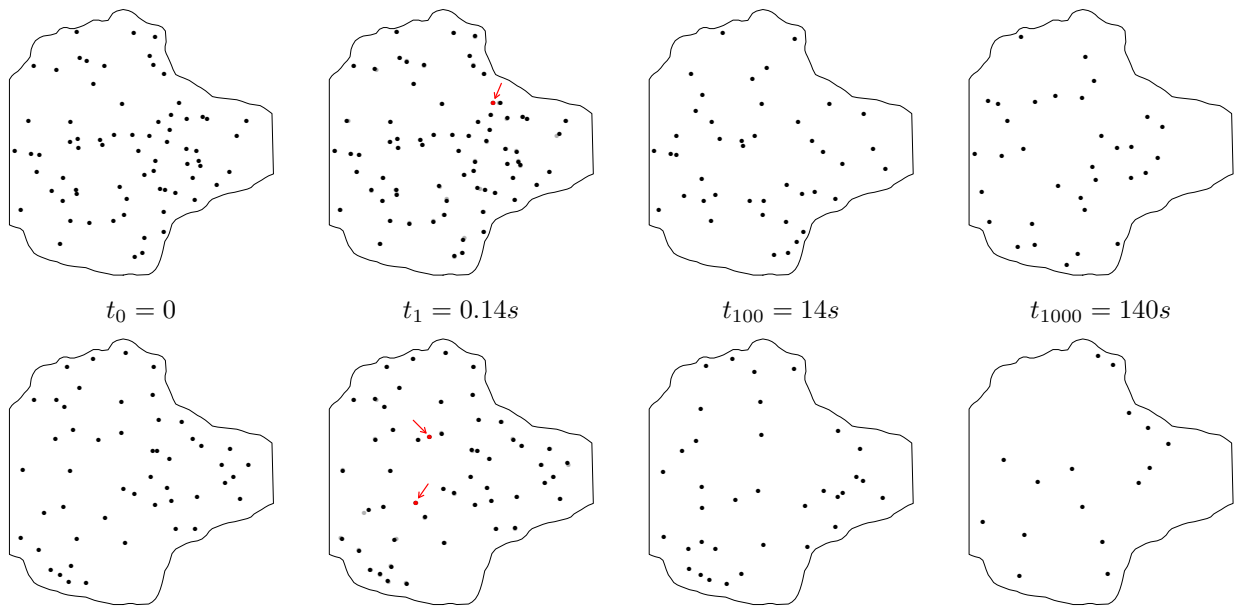
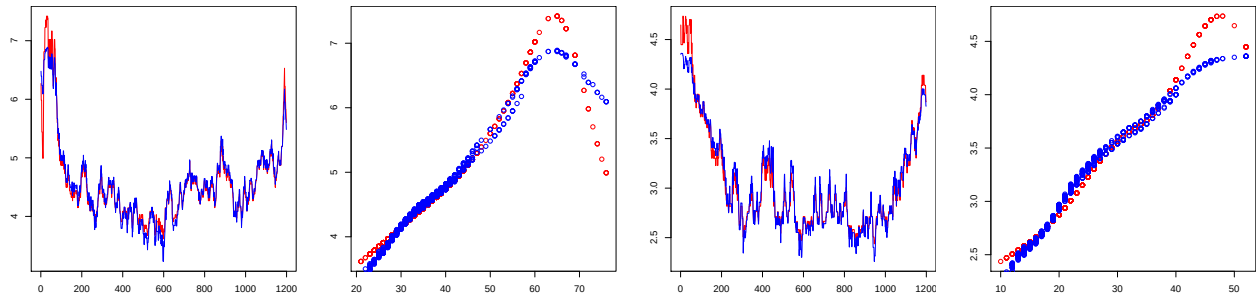


Figure 4: Locations of Langerin proteins (first row) and Rab11 proteins (second row) at different time points t_j, corresponding to the j-th frame extracted from two sequences of 1199 images, observed simultaneously in the same living cell in TIRF microscopy. In the second plot corresponding to time t_1, the new proteins that appeared during the first time interval are represented as red dots (pointed out by arrows), while the initial positions of the other proteins are recalled in gray.

For the death intensities, we represent their estimations in Figure 5-(a) and Figure 5-(c) for the Langerin and Rab11 sequences, respectively. For each sequence, both considered estimators provide similar results. Figure 5-(b) and Figure 5-(d) show the evolution of the estimated death intensities with respect to the number of observed proteins. From these plots we deduce first, that the death intensities seem to depend only on the number of proteins and second, that this dependence seems to be linear, up to some value where the death intensities decrease. This last observation could indicate the existence of periods when the exocytosis is very active, meaning that many proteins are involved and spend more time than usual in the exocytosis process. However, further experiments on new cells must be made to investigate this new hypothesis and be sure that this is not due to an estimation artefact or to experimental conditions.

Finally, we reproduce in Figure 6 the estimated death intensities for both types of proteins, based on the estimators from Example 3-(ii) (that are the red curves in Figure 5), along with the estimated cross-correlation function (ccf) between these two estimated death intensities. This last plot represents for each lag $h = -20, \dots, 20$ the empirical correlation between the death intensity of the Rab11 sequence at frame j and the death intensity of the Langerin sequence at frame $j+h$, for $j = 1, \dots, m-h$. The left plot of Figure 6 provides evidence that the death intensities of the two types of proteins follow the same trend, which is confirmed by the global high values of the empirical ccf. This observation is consistent with the biological hypothesis that both proteins interact. Interestingly, the ccf is asymmetric, showing higher values for positive lags than for negative lags. This tends to confirm previous studies (Gidon et al., 2012; Boulanger et al., 2014; Lavancier et al., 2019), where it has been concluded that Rab11 seems to be active before Langerin.



(a) (b) (c) (d)

Figure 5: (a) Estimation of the death intensity $\delta_L(x_j)$ at each configuration x_j of the Langerin sequence, for all frames j from 1 to 1199, by the discrete time estimator given by (11) based on d_κ (in blue) and the discrete time estimator given by (11) with the choice of k_T as in Example 3-(ii) (in red). (b) Scatterplot of $(n(x_j), \hat{\delta}_L(x_j))$ for the same two estimators and $j = 1, \ldots, 1199$. (c)-(d) Same plots as (a)-(b) but for the Rab11 sequence.

5 Proofs

5.1 Proof of Theorem 1

We must prove that for $A \in \mathcal{E}$ and $0 < s \leq t$,

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{P}(X_t \in A | X_s). \tag{12}$$

For any event $B \in \sigma(X_s)$, we have that $B \in \mathcal{F}_s$ so

$$\mathbb{E}[\mathbf{1}_{\{X_t \in A\}} \mathbf{1}_B] = \mathbb{E}[\mathbb{E}[\mathbf{1}_{\{X_t \in A\}} | \mathcal{F}_s] \mathbf{1}_B].$$

By definition of conditional expectation, it remains to show that $\mathbb{E}[\mathbf{1}_{\{X_t \in A\}} | \mathcal{F}_s]$ belongs to $\sigma(X_s)$ to prove (12). Let $T^{(s)}$ be the next jump after s, i.e. $T^{(s)} = \inf_{T_j > s}\{T_j\}$. We use the decomposition

$$\mathbb{P}(X_t \in A | \mathcal{F}_s) = \mathbb{E}[\mathbf{1}_{X_t \in A} \mathbf{1}_{T^{(s)} \leq t} | \mathcal{F}_s] + \mathbb{E}[\mathbf{1}_{X_t \in A} \mathbf{1}_{T^{(s)} > t} | \mathcal{F}_s]. \tag{13}$$

Then

$$\begin{aligned}
\mathbb{E}[\mathbf{1}_{X_t \in A} \mathbf{1}_{T^{(s)} \leq t} | \mathcal{F}_s] &= \sum_{j \geq 1} \mathbb{E}[\mathbf{1}_{X_t \in A} \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j} \\
&= \sum_{j \geq 1} \mathbb{E}[\mathbb{E}[\mathbf{1}_{X_t \in A} | \mathcal{F}_{T_j}] \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j}.
\end{aligned}$$

By construction of the process $(X_t)_{t \geq 0}$, given $\mathcal{F}_{T_j}$, the process $(X_t)_{t \geq T_j}$ has the same distribution as the process $(X_{t-T_j})_{t \geq T_j}$ given $X_0 = X_{T_j}$, so for $t \geq T_j$, $\mathbb{E}[\mathbf{1}_{X_t \in A} | \mathcal{F}_{T_j}]$ is a function of X_{T_j}, $t - T_j$ and A. We thus may write $\mathbb{E}[\mathbf{1}_{X_t \in A} | \mathcal{F}_{T_j}] = h(X_{T_j}, t - T_j, A)$ where h is a borelian function, so that

$$\mathbb{E}[\mathbf{1}_{X_t \in A} \mathbf{1}_{T^{(s)} \leq t} | \mathcal{F}_s] = \sum_{j \geq 1} \mathbb{E}[h(X_{T_j}, t - T_j, A) \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j}. \tag{14}$$

We have

$$\begin{aligned}
&\mathbb{E}[h(X_{T_j}, t - T_j, A) \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j} \\
&\qquad = \mathbb{E}[\int_E h(z, t - T_j, A) K(Y^{(n(X_s))}_{T_j - T_{j-1}}, dz) \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j} \\
&\qquad = \mathbb{E}[g(Y^{(n(X_{T_{j-1}}))}_{T_j - T_{j-1}}, t - T_j, A) \mathbf{1}_{T_j \leq t} | \mathcal{F}_s] \mathbf{1}_{T_{j-1} \leq s < T_j},
\end{aligned} \tag{15}$$

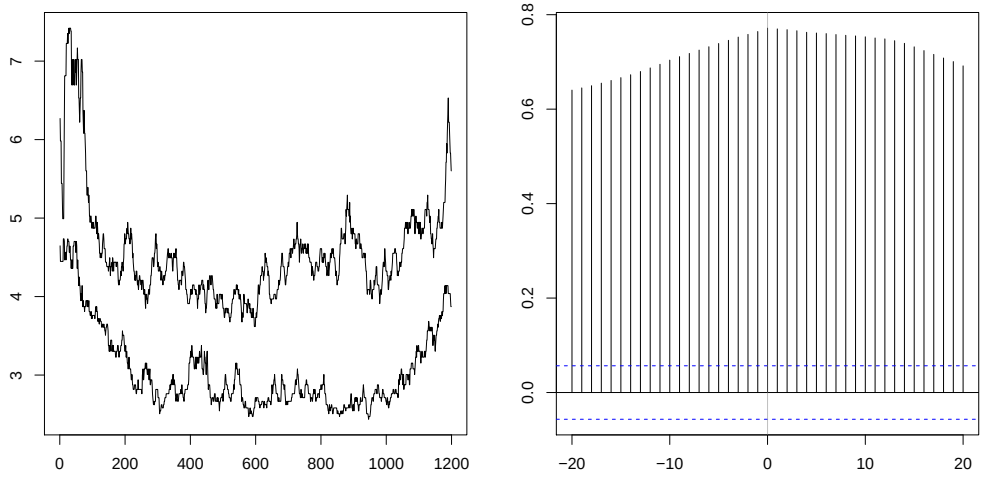


Figure 6: Left: estimation of the death intensity for the Langerin sequence (top curve) and the Rab11 sequence (bottom curve) along the 1199 observed frames. These correspond to the red curves in Figure 5. Right: empirical cross-correlation function between these two estimated death intensities, where the reference is Rab11 and the lag is applied to Langerin.

where $g(y,t,A) = \int_E h(z,t,A)K(y,dz)$.

In the sequel, for a σ-field $\mathcal{G}$ and an event B satisfying $\mathbb{P}(B) > 0$, $\mathbb{E}[X|\mathcal{G},B]$ stands for the conditional expectation of X given $\mathcal{G}$, under the conditional probability measure $\mathbb{P}_{|B}(.) = \mathbb{P}(B \cap .)/\mathbb{P}(B)$, that is $\mathbb{E}[X|\mathcal{G},B] = \mathbb{E}_{|B}[X|\mathcal{G}]$. Specifically, $Z = \mathbb{E}_{|B}[X|\mathcal{G}]$ if and only if Z is $\mathcal{G}$-measurable, integrable and for any $\mathcal{G}$-measurable bounded random variable U, $\mathbb{E}(XU\mathbf{1}_B) = \mathbb{E}(ZU\mathbf{1}_B)$. In particular, for any events B and C,

$$\mathbb{P}(B \cap C|\mathcal{G}) = \mathbb{P}(C|\mathcal{G},B)\mathbb{P}(B|\mathcal{G}), \tag{16}$$

and if B is $\mathcal{G}$-measurable, this gives $\mathbb{P}(B \cap C|\mathcal{G}) = \mathbb{P}(C|\mathcal{G},B)\mathbf{1}_B$. Accordingly, for $x \geq 0$, since the event $\{T_j > s\} = \{N_s \leq j-1\}$ is $\mathcal{F}_s$-measurable, we get

$$\begin{aligned}
&\mathbb{P}(T_j - T_{j-1} > x|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0})\mathbf{1}_{T_{j-1}\leq s<T_j} \\
&= \mathbb{P}(T_j - T_{j-1} > x, T_{j-1} \leq s < T_j|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}) \\
&= \mathbb{P}\left(T_j - T_{j-1} > x|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}, T_{j-1} \leq s < T_j\right)\mathbf{1}_{T_{j-1}\leq s<T_j}.
\end{aligned} \tag{17}$$

By construction, the distribution of $\tau_j = T_j - T_{j-1}$ only depends on $X_{T_{j-1}}$ and $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}$, which implies using (16)

$$\begin{aligned}
&\mathbb{P}\left(T_j - T_{j-1} > x|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}, T_{j-1} \leq s < T_j\right)\mathbf{1}_{T_{j-1}\leq s<T_j} \\
&= \mathbb{P}\left(T_j - T_{j-1} > x|\mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}, T_{j-1} \leq s < T_j\right)\mathbf{1}_{T_{j-1}\leq s<T_j}.
\end{aligned}$$

This probability can be computed using (16) and (2), so that (17) simplifies into

$$\begin{aligned}
&\mathbb{P}(T_j - T_{j-1} > x|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0})\mathbf{1}_{T_{j-1}\leq s<T_j} \\
&= \left(\mathbf{1}_{x\leq s-T_{j-1}} + \mathbf{1}_{x>s-T_{j-1}} e^{-\int_{s-T_{j-1}}^{x} \alpha(Y_u^{(n(X_{T_{j-1}}))})du}\right)\mathbf{1}_{T_{j-1}\leq s<T_j}.
\end{aligned}$$

By continuity of $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}$ and α, this proves that

$$B \mapsto \mathbb{E}[\mathbf{1}_{T_j - T_{j-1}\in B}|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}]\mathbf{1}_{T_{j-1}\leq s<T_j}$$

defines a measure with density $\alpha(Y_x^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x}\alpha(Y_u^{n(X_{T_{j-1}})})du}\mathbf{1}_{x>s-T_{j-1}}\mathbf{1}_{T_{j-1}\leq s<T_j}$. Therefore,

$$\begin{aligned}
&\mathbb{E}[g(Y_{T_j-T_{j-1}}^{(n(X_{T_{j-1}}))}, t-T_j, A)\mathbf{1}_{T_j\leq t}|\mathcal{F}_s, \mathcal{F}_{T_{j-1}}, (Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}]\mathbf{1}_{T_{j-1}\leq s<T_j}\\
&= \int_{s-T_{j-1}}^{t-T_{j-1}} g(Y_x^{(n(X_{T_{j-1}}))}, t-T_{j-1}-x, A)\alpha(Y_x^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x}\alpha(Y_u^{n(X_{T_{j-1}})})du}dx\mathbf{1}_{T_{j-1}\leq s<T_j}\\
&= \int_{s}^{t} g(Y_{x-T_{j-1}}^{(n(X_{T_{j-1}}))}, t-x, A)\alpha(Y_{x-T_{j-1}}^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x-T_{j-1}}\alpha(Y_u^{n(X_{T_{j-1}})})du}dx\mathbf{1}_{T_{j-1}\leq s<T_j}.
\end{aligned}$$

Coming back to (15), this gives

$$\begin{aligned}
&\mathbb{E}[h(X_{T_j}, t-T_j, A)\mathbf{1}_{T_j\leq t}|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\
&= \mathbb{E}[\int_s^t g(Y_{x-T_{j-1}}^{(n(X_{T_{j-1}}))}, t-x, A)\alpha(Y_{x-T_{j-1}}^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x-T_{j-1}}\alpha(Y_u^{n(X_{T_{j-1}})})du}dx|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\
&= \mathbb{E}[\mathbb{E}[\int_s^t g(Y_{x-T_{j-1}}^{(n(X_{T_{j-1}}))}, t-x, A)\alpha(Y_{x-T_{j-1}}^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x-T_{j-1}}\alpha(Y_u^{n(X_{T_{j-1}})})du}dx|\mathcal{F}_s, X_{T_{j-1}}, T_{j-1}]\mathbf{1}_{T_{j-1}\leq s<T_j}|\mathcal{F}_s].
\end{aligned}$$

Since the process $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}$ is independent of T_{j-1}, conditional on $X_{T_{j-1}}$ and T_{j-1}, the process $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq s-T_{j-1}}$ is a Markov process whose conditional distribution with respect to $\mathcal{F}_s$ only depends on $Y_{s-T_{j-1}}^{n(X_{T_{j-1}})}$. Since by definition $Y_{s-T_{j-1}}^{n(X_{T_{j-1}})} = X_s$ almost surely, the expression above then becomes

$$\mathbb{E}[\int_s^t g(Y_{x-T_{j-1}}^{(n(X_{T_{j-1}}))}, t-x, A)\alpha(Y_{x-T_{j-1}}^{n(X_{T_{j-1}})})e^{-\int_{s-T_{j-1}}^{x-T_{j-1}}\alpha(Y_u^{n(X_{T_{j-1}})})du}dx|X_{T_{j-1}}, T_{j-1}, Y_{s-T_{j-1}}^{n(X_{T_{j-1}})} = X_s]\mathbf{1}_{T_{j-1}\leq s<T_j}$$

and by homogeneity of the Markov process $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq s-T_{j-1}}$ (given $X_{T_{j-1}}$ and T_{j-1}), we obtain

$$\mathbb{E}[\int_0^{t-s} g(Y_v^{(n(X_{T_{j-1}}))}, t-s-v, A)\alpha(Y_v^{n(X_{T_{j-1}})})e^{-\int_0^v\alpha(Y_u^{n(X_{T_{j-1}})})du}dv|X_{T_{j-1}}, T_{j-1}, Y_0^{n(X_{T_{j-1}})} = X_s]\mathbf{1}_{T_{j-1}\leq s<T_j}.$$

Given $X_{T_{j-1}}$, the process $(Y_u^{(n(X_{T_{j-1}}))})_{u\geq 0}$ is independent of T_{j-1}. Further $X_{T_{j-1}} = Y_0^{n(X_{T_{j-1}})}$ almost surely and $n(X_{T_{j-1}}) = n(X_s)$ for $T_{j-1}\leq s<T_j$, therefore we get

$$\begin{aligned}
&\mathbb{E}[h(X_{T_j}, t-T_j, A)\mathbf{1}_{T_j\leq t}|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\
&= \mathbb{E}[\int_0^{t-s} g(Y_v^{(n(X_s))}, t-s-v, A)\alpha(Y_v^{n(X_s)})e^{-\int_0^v\alpha(Y_u^{n(X_s)})du}dv|Y_0^{n(X_s)} = X_s]\mathbf{1}_{T_{j-1}\leq s<T_j}.
\end{aligned}$$

Finally, from (14),

$$\mathbb{E}[\mathbf{1}_{X_t\in A}\mathbf{1}_{T^{(s)}\leq t}|\mathcal{F}_s] = \mathbb{E}[\int_0^{t-s} g(Y_v^{(n(X_s))}, t-s-v, A)\alpha(Y_v^{n(X_s)})e^{-\int_0^v\alpha(Y_u^{n(X_s)})du}dv|Y_0^{n(X_s)} = X_s], \quad (18)$$

proving that the first term in (13) belongs to $\sigma(X_s)$. For the second term in (13),

$$\begin{aligned}\mathbb{E}[\mathbf{1}_{X_t\in A}\mathbf{1}_{T^{(s)}>t}|\mathcal{F}_s] &= \sum_{j\geq 1}\mathbb{E}[\mathbf{1}_{Y^{(n(X_s))}_{t-T_{j-1}}\in A}\mathbf{1}_{T^{(s)}>t}|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\ &= \sum_{j\geq 1}\mathbb{E}[\mathbb{P}(T_j>t|\mathcal{F}_s,(Y^{(n(X_s)}_u)_{u\geq 0})\mathbf{1}_{Y^{n(X_s)}_{t-T_{j-1}}\in A}|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\ &= \sum_{j\geq 1}\mathbb{E}[e^{-\int_s^t\alpha(Y^{(n(X_s))}_{v-T_{j-1}})dv}\mathbf{1}_{Y^{(n(X_s))}_{t-T_{j-1}}\in A}|\mathcal{F}_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\end{aligned}$$

Using the same arguments as for the first term in (13), we obtain

$$\begin{aligned}\mathbb{E}[\mathbf{1}_{X_t\in A}\mathbf{1}_{T^{(s)}>t}|\mathcal{F}_s] &= \sum_{j\geq 1}\mathbb{E}[e^{-\int_0^{t-s}\alpha(Y^{(n(X_s))}_v)dv}\mathbf{1}_{Y^{(n(X_s))}_{t-s}\in A}|Y^{(n(X_s))}_0=X_s]\mathbf{1}_{T_{j-1}\leq s<T_j}\\ &= \mathbb{E}[e^{-\int_0^{t-s}\alpha(Y^{(n(X_s))}_v)dv}\mathbf{1}_{Y^{(n(X_s))}_{t-s}\in A}|Y^{(n(X_s))}_0=X_s],\end{aligned}\tag{19}$$

which concludes the proof that $(X_t)_{t\geq 0}$ is a Markov process. Notice finally that the two expressions (18) and (19) imply for $t\geq s\geq 0$

$$\mathbb{E}[\mathbf{1}_{X_t\in A}|\mathcal{F}_s]=\mathbb{E}[\mathbf{1}_{X_{t-s}\in A}|X_0=X_s],$$

showing that $(X_t)_{t\geq 0}$ is homogeneous.

5.2 Proof of Proposition 2

To prove that $(X_t)_{t\geq 0}$ admits an invariant measure, we can view the process as a classical regenerative process with regeneration times $\{t, X_t=\varnothing\}$, see for instance (Thorisson, 2000, Chapter 10) for a definition. Under (H1), it is not difficult to verify that the expected time between two regenerations is finite, which implies the existence of an invariant measure (Thorisson, 2000, Chapter 10, Theorem 3.1).

It remains to establish the uniform geometric ergodicity (3). The proof uses a standard coupling argument as carried out for pure spatial birth-death processes in (Lotwick and Silverman, 1981, Theorem A) and (Møller, 1989, Theorem 3.1 and Corollary 3.1). Let $(X^{(1)}_t)_{t\geq 0}$ and $(X^{(2)}_t)_{t\geq 0}$ be two birth-death-move processes with the same transition kernel Q_t as $(X_t)_{t\geq 0}$ and with respective initial distribution ϕ_1 and ϕ_2 on E. Consider the stopping time

$$\tau=\inf\{t>0:(X^{(1)}_t,X^{(2)}_t)=(\varnothing,\varnothing)\}.$$

We verify in Lemma 13 that for any $A\in\mathcal{E}$ and any $t\geq 0$,

$$\mathbb{P}_{\phi_1\times\phi_2}(X^{(1)}_t\in A,\tau\leq t)=\mathbb{P}_{\phi_1\times\phi_2}(X^{(2)}_t\in A,\tau\leq t).$$

Therefore by the coupling argument,

$$\left|\int_E Q_t(x,A)\phi_1(dx)-\int_E Q_t(x,A)\phi_2(dx)\right|\leq 2P_{\phi_1\times\phi_2}(\tau>t).\tag{20}$$

In order to control $P_{\phi_1\times\phi_2}(\tau>t)$, let us denote by $(Y^{(n)}_{k,u})_{u\geq 0}$, for $k=1,2$, the Markov process driving the motion of $X^{(k)}_t$ in E_n and by $(T^{(k)}_j)_{j\geq 0}$ the jump times of $X^{(k)}_t$.

Let $t_0>0$, we have

$$\begin{aligned}&\mathbb{P}_{\phi_1\times\phi_2}(\tau\leq t_0|X^{(1)}_0,X^{(2)}_0)\\ &\qquad=\mathbf{1}_{\{(X^{(1)}_0,X^{(2)}_0)=(\varnothing,\varnothing)\}}+\sum_{(n_1,n_2)\neq(0,0)}\mathbb{P}_{\phi_1\times\phi_2}(\tau\leq t_0|X^{(1)}_0,X^{(2)}_0)\mathbf{1}_{\{n(X^{(1)}_0)=n_1,n(X^{(2)}_0)=n_2\}}.\end{aligned}\tag{21}$$

On the event $\{n(X_0^{(1)})=n_1, n(X_0^{(2)})=n_2\}$ with $n_1,n_2\geq 1$,

$$\begin{aligned}\mathbb{P}_{\phi_1\times\phi_2}(\tau\leq t_0|X_0^{(1)},X_0^{(2)}) &\geq \mathbb{P}_{\phi_1\times\phi_2}\Big(\bigcap_{j=1}^{n_1}\{T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n_1-j}\},\ T^{(1)}_{n_1+1}>t_0,\\ &\qquad\bigcap_{j=1}^{n_2}\{T_j^{(2)}-T_{j-1}^{(2)}\leq\frac{t_0}{n_2},X^{(2)}_{T_j^{(2)}}\in E_{n_2-j}\},\ T^{(2)}_{n_2+1}>t_0|X_0^{(1)},X_0^{(2)}\Big)\\ &\geq \mathbb{P}_{\phi_1}\Big(\bigcap_{j=1}^{n_1}\{T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n_1-j}\},\ T^{(1)}_{n_1+1}-T^{(1)}_{n_1}>t_0|X_0^{(1)}\Big)\\ &\quad\times\mathbb{P}_{\phi_2}\Big(\bigcap_{j=1}^{n_2}\{T_j^{(2)}-T_{j-1}^{(2)}\leq\frac{t_0}{n_2},X^{(2)}_{T_j^{(2)}}\in E_{n_2-j}\},\ T^{(2)}_{n_2+1}-T^{(2)}_{n_2}>t_0|X_0^{(2)}\Big),\end{aligned}\tag{22}$$

where we have used the independence between ϕ_1 and ϕ_2. Each term of this product is treated similarly and we only detail the first one. On the event $n(X_0^{(1)})=n_1$ with $n_1\geq 1$,

$$\begin{aligned}&\mathbb{P}_{\phi_1}\Big(\bigcap_{j=1}^{n_1}\{T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n_1-j}\},\ T^{(1)}_{n_1+1}-T^{(1)}_{n_1}>t_0|X_0^{(1)}\Big)\\ =&\mathbb{P}\left(T^{(1)}_{n_1+1}-T^{(1)}_{n_1}>t_0|X^{(1)}_{T^{(1)}_{n_1}}=\varnothing\right)\mathbb{E}_{\phi_1}\left[\mathbb{P}_{\phi_1}\left(T_1^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_1^{(1)}}\in E_{n_1-1}|X_0^{(1)},(Y^{(n_1)}_{1,u})_{u\geq 0}\right)|X_0^{(1)}\right]\\ &\prod_{j=2}^{n_1}\mathbb{E}_{\phi_1}\left[\mathbb{P}_{\phi_1}\left(T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n_1-j}|X^{(1)}_{T^{(1)}_{j-1}},(Y^{(n_1-j+1)}_{1,u})_{u\geq 0}\right)\Big|A_{j-1},X_0^{(1)}\right]\end{aligned}\tag{23}$$

where for $j\geq 1$, $A_j=\bigcap_{i=1}^{j}\{T_i^{(1)}-T_{i-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_i^{(1)}}\in E_{n_1-i}\}$. For any $j\geq 2$ and any $1\leq n\leq n^*$, on the event $n(X^{(1)}_{T^{(1)}_{j-1}})=n$, by definition of the process $(X_t)_{t\geq 0}$,

$$\begin{aligned}&\mathbb{P}_{\phi_1}\left(T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n-1}|X^{(1)}_{T^{(1)}_{j-1}},(Y^{(n)}_{1,u})_{u\geq 0}\right)\\ &\qquad=\mathbb{P}_{\phi_1}\left(T_1^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_1^{(1)}}\in E_{n-1}|X_0^{(1)},(Y^{(n)}_{1,u})_{u\geq 0}\right)\\ &\qquad=\frac{\delta(Y^{(n)}_{1,T_1^{(1)}})}{\alpha(Y^{(n)}_{1,T_1^{(1)}})}(1-e^{-\int_0^{t_0/n_1}\alpha(Y^{(n)}_{1,u})du})\\ &\qquad\geq\frac{\delta_*}{\alpha^*}(1-e^{-\alpha_* t_0/n_1})\\ &\qquad\geq\frac{\delta_*}{\alpha^*}(1-e^{-\alpha_* t_0/n^*}).\end{aligned}$$

This inequality is uniform in n. Applying it in (23) along with $\mathbb{P}\left(T^{(1)}_{n_1+1}-T^{(1)}_{n_1}>t_0|X^{(1)}_{T^{(1)}_{n_1}}=\varnothing\right)=e^{-\alpha(\varnothing)t_0}\geq e^{-\alpha^* t_0}$, we get on the event $n(X_0^{(1)})=n_1$ with $n_1\geq 1$ that

$$\begin{aligned}\mathbb{P}_{\phi_1}\Big(\bigcap_{j=1}^{n_1}\{T_j^{(1)}-T_{j-1}^{(1)}\leq\frac{t_0}{n_1},X^{(1)}_{T_j^{(1)}}\in E_{n_1-j}\},\ T^{(1)}_{n_1+1}-T^{(1)}_{n_1}>t_0|X_0^{(1)}\Big)&\geq e^{-\alpha^* t_0}(\frac{\delta_*}{\alpha^*}(1-e^{-\alpha_* t_0/n^*}))^{n_1}\\ &\geq e^{-\alpha^* t_0}(\frac{\delta_*}{\alpha^*}(1-e^{-\alpha_* t_0/n^*}))^{n^*}.\end{aligned}$$

Coming back to (22), we deduce that on the event $\{n(X_0^{(1)}) = n_1, n(X_0^{(2)}) = n_2\}$ with $n_1, n_2 \geq 1$

$$\mathbb{P}_{\phi_1 \times \phi_2}(\tau \leq t_0 | X_0^{(1)}, X_0^{(2)}) \geq \rho_{t_0}$$

for some $\rho_{t_0} > 0$ that depends on t_0 but not on n_1, n_2, $X_0^{(1)}$ and $X_0^{(2)}$. By a similar argument we obtain the same result on the event $\{n(X_0^{(1)}) = n_1, n(X_0^{(2)}) = n_2\}$ with $n_1 \geq 1, n_2 = 0$ or $n_1 = 0, n_2 \geq 1$. We conclude from (21) that for any $t_0 > 0$,

$$\mathbb{P}_{\phi_1 \times \phi_2}(\tau > t_0) \leq 1 - \rho_{t_0}$$

where ρ_{t_0} does not depend on ϕ_1 and ϕ_2. By a standard argument, see for instance Lotwick and Silverman (1981), we then deduce that

$$\mathbb{P}_{\phi_1 \times \phi_2}(\tau > t) \leq a e^{-ct}$$

for some $a > 0$ and $c > 0$ that do not depend on ϕ_1 and ϕ_2.

In view of (20), the latter inequality for the choice $\phi_1(.) = \mathbf{1}_{y \in .}$ and $\phi_2 = \mu_\infty$ implies (3) when $g = \mathbf{1}_A$, for some $A \in \mathcal{E}$. To extend it to any measurable bounded function g, first consider the case where g takes its values in $[0, 1]$. Then g can be approximated by the step function $g_n(z) = 2^{-n} \lfloor 2^n g(z) \rfloor$, where $\lfloor . \rfloor$ denotes the integer part function, so that $\|g - g_n\|_\infty \leq 2^{-n}$. Let $A_{j,n} = \{j 2^{-n} \leq g(z) < (j+1) 2^{-n}\}$. Using (3) for indicator functions, we deduce that for any $y \in E$ and any n

$$\left| \int_E g_n(z) Q_t(y, dz) - \int_E g_n(z) \mu_\infty(dz) \right| = 2^{-n} \sum_{j=0}^{2^n} j \, |Q_t(y, A_{j,n}) - \mu_\infty(A_{j,n})| \leq 2^n a e^{-ct}.$$

Therefore for any $y \in E$ and any n

$$\left| \int_E g(z) Q_t(y, dz) - \int_E g(z) \mu_\infty(dz) \right| \leq 2\|g - g_n\|_\infty + 2^n a e^{-ct} \leq 2^{-n+1} + 2^n a e^{-ct}.$$

Choosing $n = \lfloor ct/(2 \log 2) \rfloor$, we get (3) (for new constants $a > 0$ and $c > 0$) in the case where g takes its values in $[0, 1]$. Applying this result to $g/\|g\|_\infty$ proves (3) for positive bounded functions g. The extension to any measurable bounded function g is obtained by considering the decomposition $g = g_+ - g_-$ where g_+ and g_- respectively denote the positive and negative part of g.

5.3 Proof of Corollary 3

To prove (4), note that

$$\mathbb{E}\left(\int_0^t g(X_s) ds \right) - t \int_E g(z) \mu_\infty(dz) = \int_0^t \int_E \left(\int_E g(z) Q_s(y, dz) - \int_E g(z) \mu_\infty(dz) \right) \mu_0(dy) ds$$

where μ_0 denotes the distribution of X_0. Using (3), we get

$$\begin{aligned} \left| \mathbb{E}\left(\int_0^t g(X_s) ds \right) - t \int_E g(z) \mu_\infty(dz) \right| &\leq \int_0^t \sup_{y \in E} \left| \int_E g(z) Q_s(y, dz) - \int_E g(z) \mu_\infty(dz) \right| ds \\ &\leq a \|g\|_\infty \int_0^t e^{-cs} ds \end{aligned}$$

hence the result.

In order to prove (5) let us write E_s for

$$E_s := \mathbb{E}[g(X_s)] = \int_E \int_E g(y) Q_s(x, dy) \mu_0(dx).$$

For any $v \geq s$ we have $\mathbb{E}[g(X_v)|\mathcal{F}_s] = \int_E g(z)Q_{v-s}(X_s, dz)$, so

$$\mathbb{V}\left(\int_0^t g(X_s)ds\right) = 2\int_{s=0}^t \int_{v=s}^t \mathbb{E}\left[(g(X_s) - E_s)(\int_E g(z)Q_{v-s}(X_s, dz) - E_v)\right] dvds$$
$$= 2\int_{s=0}^t \int_{v=0}^{t-s} \iint_E (g(y) - E_s)\left(\int_E g(z)Q_v(y, dz) - E_{v+s}\right) Q_s(x, dy)\mu_0(dx)dvds$$
$$= 2\int_{s=0}^t \iint_E (g(y) - E_s)\left[\int_{v=0}^{t-s}\left(\int_E g(z)Q_v(y, dz) - \int_{z\in E} g(z)\mu_\infty(dz)\right)dv\right] Q_s(x, dy)\mu_0(dx)ds$$
$$+ 2\int_{s=0}^t \iint_E (g(y) - E_s)\left[\int_{v=s}^t \left(\int_{z\in E} g(z)\mu_\infty(dz) - \iint_E g(z)Q_v(u, dz)\mu_0(du)\right)dv\right] Q_s(x, dy)\mu_0(dx)ds.$$

Thanks to (3), each term in the square brackets above is uniformly bounded in s, t and y, so there exists a positive constant c_0 such that

$$\mathbb{V}\left(\int_0^t g(X_s)ds\right) \leq c_0\|g\|_\infty \int_{s=0}^t \iint_E (g(y) + E_s)Q_s(x, dy)\mu_0(dx)ds = 2c_0\|g\|_\infty \int_{s=0}^t E_s ds.$$

5.4 Proof of Proposition 4

The counting process N_t^γ is clearly adapted to $\mathcal{F}_t$ and its $\mathcal{F}_t$-intensity is obtained by $\lambda^\gamma(t) = \lim_{h\to 0^+} \lambda_h^\gamma(t)$, almost surely, where

$$\lambda_h^\gamma(t) = \frac{1}{h}\mathbb{E}\left(N_{t+h}^\gamma - N_t^\gamma|\mathcal{F}_t\right).$$

This makes sense if for instance $\lambda_h^\gamma(t)$ is uniformly bounded for any $t \geq 0$ and any $0 \leq h \leq 1$ (see formula (3.5) in Chapter 2 of Brémaud (1981)), which is our case as shown below. It is indeed not difficult to deduce that under these assumptions, by applications of Fubini and the dominated convergence theorem,

$$\mathbb{E}(N_{t+s}^\gamma - N_t^\gamma|\mathcal{F}_t) = \mathbb{E}\left(\lim_{h\to 0^+} \int_t^{t+s} \frac{1}{h}(N_{u+h}^\gamma - N_u^\gamma)du\Big|\mathcal{F}_t\right)$$
$$= \lim_{h\to 0^+} \mathbb{E}\left(\int_t^{t+s} \lambda_h^\gamma(u)du\Big|\mathcal{F}_t\right) = \mathbb{E}\left(\int_t^{t+s} \lambda^\gamma(u)du\Big|\mathcal{F}_t\right),$$

which shows that $\int_0^t \lambda^\gamma(u)du$ is the $\mathcal{F}_t$-compensator of N_t^γ, i.e. $N_t^\gamma - \int_0^t \lambda^\gamma(u)du$ is a $\mathcal{F}_t$-martingale.

Let us prove that $\lambda_h^\gamma(t)$ is uniformly bounded and $\lim_{h\to 0^+} \lambda_h^\gamma(t) = \gamma(X_t)$. By the Markov property, for any $y \in E$,

$$\frac{1}{h}\mathbb{E}\left(N_{t+h}^\gamma - N_t^\gamma|X_t = y\right) = \frac{1}{h}\mathbb{E}\left(N_h^\gamma|X_0 = y\right)$$
$$= \frac{1}{h}\mathbb{P}\left(N_h^\gamma = 1|X_0 = y\right) + \frac{1}{h}\sum_{j\geq 2} j\mathbb{P}\left(N_h^\gamma = j|X_0 = y\right). \tag{24}$$

On one hand, for any $j \geq 2$,

$$\mathbb{P}\left(N_h^\gamma = j|X_0 = y\right) \leq \mathbb{P}\left(N_h^\alpha = j|X_0 = y\right) \leq \mathbb{P}\left(N_h^\alpha \geq j|X_0 = y\right).$$

It is easy to verify that Lemma 5 remains true by replacing the probabilities in its statement by conditional probabilities. We deduce that the above conditional probability is lower than $\mathbb{P}(N^* \geq j)$ with $N^* \sim \mathcal{P}(\alpha^* h)$, which in turn is lower than $(\alpha^* h)^j/j!$. Consequently

$$\frac{1}{h}\sum_{j\geq 2} j\mathbb{P}\left(N_h^\gamma = j|X_0 = y\right) \leq \frac{1}{h}\sum_{j\geq 2} j\frac{(\alpha^* h)^j}{j!} \leq h(\alpha^*)^2 e^{\alpha^* h}.$$

On the other hand,

$$\frac{1}{h}\mathbb{P}\left(N_h^\gamma = 1 | X_0 = y\right) = \frac{1}{h}\mathbb{P}\left(N_h^\gamma = 1, N_h^\alpha = 1 | X_0 = y\right) + \frac{1}{h}\mathbb{P}\left(N_h^\gamma = 1, N_h^\alpha \geq 2 | X_0 = y\right).$$

The last term is lower than $\frac{1}{h}\mathbb{P}\left(N_h^\alpha \geq 2 | X_0 = y\right)$ which is less than $\frac{1}{h}(\alpha^* h)^2/2$ by the same arguments as above. We shall finally prove that $\left(\frac{1}{h}\mathbb{P}(N_h^\gamma = 1, N_h^\alpha = 1 | X_0 = y)\right)_{h \leq 1}$ is uniformly bounded and that $\lim_{h\to 0} \frac{1}{h}\mathbb{P}(N_h^\gamma = 1, N_h^\alpha = 1 | X_0 = y) = \gamma(y)$. Let us first consider the case where $\gamma = \beta$ and recall that $n(y)$ is the index n such that $y \in E_n$. By (2) along with the continuity of α and $(Y_u^{(n(y))})_{u \geq 0}$,

$$\begin{aligned}
&\mathbb{P}\left(N_h^\beta = 1, N_h^\alpha = 1 | X_0 = y\right) = \mathbb{P}\left(T_1 \leq h, \text{first jump is a birth}, T_2 > h | X_0 = y\right) \\
&= \int_{z \in E_{n(y)+1}} \int_0^h \mathbb{E}\left[\alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) e^{-\int_0^s \alpha(Y_u^{(n(y))})du} e^{-\int_0^{h-s} \alpha(Y_v^{(n(z))})dv} | X_0 = y\right] ds \qquad (25) \\
&\leq h\alpha^*.
\end{aligned}$$

We obtain similarly for $\gamma = \delta$ and $\gamma = \alpha$

$$\begin{aligned}
&\mathbb{P}\left(N_h^\delta = 1, N_h^\alpha = 1 | X_0 = y\right) = \mathbb{P}\left(T_1 \leq h, \text{first jump is a death}, T_2 > h | X_0 = y\right) \\
&= \int_{z \in E_{n(y)-1}} \int_0^h \mathbb{E}\left[\alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) e^{-\int_0^s \alpha(Y_u^{(n(y))})du} e^{-\int_0^{h-s} \alpha(Y_v^{(n(z))})dv} | X_0 = y\right] ds \\
&\leq h\alpha^*
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\left(N_h^\alpha = 1 | X_0 = y\right) &= \int_{z \in E} \int_0^h \mathbb{E}\left[\alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) e^{-\int_0^s \alpha(Y_u^{(n(y))})du} e^{-\int_0^{h-s} \alpha(Y_v^{(n(z))})dv} | X_0 = y\right] ds \\
&\leq h\alpha^*.
\end{aligned}$$

So $\left(\frac{1}{h}\mathbb{P}(N_h^\gamma = 1, N_h^\alpha = 1 | X_0 = y)\right)_{h \leq 1}$ is uniformly bounded whatever $\gamma = \beta$ or $\gamma = \delta$ or $\gamma = \alpha$.

Let us show that $\lim_{h\to 0} \frac{1}{h}\mathbb{P}(N_h^\gamma = 1, N_h^\alpha = 1 | X_0 = y) = \gamma(y)$ for $\gamma = \beta$, the other cases $\gamma = \delta$ and $\gamma = \alpha$ being treated similarly. Using (25), the inequalities $|1 - e^{-x}| \leq x$ for $x \geq 0$, $\int_0^{h-s} \alpha(Y_v^{(n(z))})dv \leq (h-s)\alpha^*$, $\int_0^s \alpha(Y_u^{(n(y))})du \leq \alpha^* s$ and $\int_{z \in E_{n(y)+1}} \alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) = \beta(Y_s^{(n(y))})$, we obtain

$$\begin{aligned}
&\left|\frac{1}{h}\mathbb{P}(N_h^\beta = 1, N_h^\alpha = 1 | X_0 = y) - \beta(y)\right| \\
&\leq \int_{z \in E_{n(y)+1}} \frac{1}{h} \int_0^h \mathbb{E}\left[\alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) e^{-\int_0^s \alpha(Y_u^{(n(y))})du} |1 - e^{-\int_0^{h-s} \alpha(Y_v^{(n(z))})dv}| | X_0 = y\right] ds \\
&+ \int_{z \in E_{n(y)+1}} \frac{1}{h} \int_0^h \mathbb{E}\left[\alpha(Y_s^{(n(y))}) K(Y_s^{(n(y))}, dz) |1 - e^{-\int_0^s \alpha(Y_u^{(n(y))})du}| | X_0 = y\right] ds \\
&+ \frac{1}{h} \int_0^h \mathbb{E}\left[|\beta(Y_s^{(n(y))}) - \beta(y)| | X_0 = y\right] ds \\
&\leq \frac{(\alpha^*)^2 h}{2} + \frac{(\alpha^*)^2 h}{2} + \frac{1}{h} \int_0^h \mathbb{E}\left[|\beta(Y_s^{(n(y))}) - \beta(y)| | X_0 = y\right] ds,
\end{aligned}$$

and the continuity of β and $(Y_s^{n(y)})_{s \geq 0}$ entail the result.

This concludes the proof that $\gamma(X_t)$ is the $\mathcal{F}_t$-intensity of N_t^γ, meaning that $N_t^\gamma - \int_0^t \gamma(X_s)ds$ is a martingale with respect to $\mathcal{F}_t$. A left-continuous (predictable) version of the intensity is $\gamma(X_{t^-})$. Since

for any bounded measurable function g, the function $s \mapsto g(X_{s^-})$ is predictable with respect to $\mathcal{F}_s$, we deduce that $M_t = \int_0^t g(X_{s^-})[dN_s^\gamma - \gamma(X_s)ds]$ is also a $\mathcal{F}_t$-martingale (see Section 17.2 in Bass (2011)), proving the second statement of the lemma. Finally, writing $M_t(g)$ to stress the dependence of M_t on g, we have that for any bounded measurable function g, $\mathbb{E}(M_t(g)) = 0$ and by (17.8) in Bass (2011),

$$\mathbb{E}(M_t^2(g)) = \mathbb{E}\left(\int_0^t g^2(X_{s^-})dN_s^\gamma\right) = \mathbb{E}\left(M_t(g^2)\right) + \mathbb{E}\left(\int_0^t g^2(X_s)\gamma(X_s)ds\right) = \mathbb{E}\left(\int_0^t g^2(X_s)\gamma(X_s)ds\right).$$

5.5 Proof of Proposition 6

We consider the decomposition

$$\hat{\gamma}(x) - \gamma(x) = \frac{M_T}{\hat{T}(x)} + R_T,$$

where

$$M_T = \int_0^T k_T(x, X_{s^-})[dN_s^\gamma - \gamma(X_s)ds] \quad \text{and} \quad R_T = \frac{1}{\hat{T}(x)}\int_0^T (\gamma(X_s) - \gamma(x))k_T(x, X_s)ds.$$

Here, when $\gamma = \alpha$, we use the notation $N_s^\alpha = N_s$. For all $\epsilon > 0$,

$$\begin{aligned}
\mathbb{P}(|\hat{\gamma}(x) - \gamma(x)| > \epsilon) &\leq \mathbb{P}\left(\hat{T}(x) < \frac{\mathbb{E}(\hat{T}(x))}{2}\right) + \mathbb{P}\left(|\hat{\gamma}(x) - \gamma(x)| > \epsilon, \hat{T}(x) \geq \frac{\mathbb{E}(\hat{T}(x))}{2}\right) \\
&\leq \mathbb{P}\left(\hat{T}(x) < \frac{\mathbb{E}(\hat{T}(x))}{2}\right) + \mathbb{P}\left(\frac{|M_T|}{\hat{T}(x)} > \frac{\epsilon}{2}, \hat{T}(x) \geq \frac{\mathbb{E}(\hat{T}(x))}{2}\right) + \mathbb{P}\left(|R_T| > \frac{\epsilon}{2}\right) \\
&\leq \mathbb{P}\left(|\hat{T}(x) - \mathbb{E}(\hat{T}(x))| > \frac{\mathbb{E}(\hat{T}(x))}{2}\right) + \mathbb{P}\left(|M_T| > \frac{\epsilon\mathbb{E}(\hat{T}(x))}{4}\right) + \mathbb{P}\left(|R_T| > \frac{\epsilon}{2}\right) \\
&\leq \frac{4\mathbb{V}(\hat{T}(x))}{\mathbb{E}(\hat{T}(x))^2} + \frac{16\mathbb{E}(M_T^2)}{\epsilon^2\mathbb{E}(\hat{T}(x))^2} + \frac{4\mathbb{E}(R_T^2)}{\epsilon^2}, \qquad (26)
\end{aligned}$$

using the Markov inequality in the last line. By (6) applied with $g(X_s) = k_T(x, X_s)$ and $t = T$, and since k_T is uniformly bounded, we get that $\mathbb{E}\left(M_T^2\right) \leq c_0\mathbb{E}(\hat{T}(x))$. By (10), we deduce that

$$\mathbb{P}(|\hat{\gamma}(x) - \gamma(x)| > \epsilon) \leq \frac{c_0}{Tv_T(x)} + \frac{4\mathbb{E}(R_T^2)}{\epsilon^2}.$$

The proof is then complete once we show that

$$E\left(R_T^2\right) \leq c_0\left(\frac{1}{Tv_T(x)} + w_T^2(x)\right). \qquad (27)$$

To this end, notice that almost surely $|R_T| \leq 2\alpha^*$ because $\gamma \leq \alpha$, so

$$E\left(R_T^2\right) \leq 4(\alpha^*)^2\mathbb{P}\left(\hat{T}(x) \leq \frac{\mathbb{E}(\hat{T}(x))}{2}\right) + \frac{4}{\mathbb{E}(\hat{T}(x))^2}\mathbb{E}\left[\left(\int_0^T (\gamma(X_s) - \gamma(x))k_T(d(x, X_s))ds\right)^2\right].$$

Denoting $I_T = \int_0^T \gamma(X_s)k_T(x, X_s)ds$ and using the Chebyshev inequality, we obtain

$$\begin{aligned}
E\left(R_T^2\right) &\leq \frac{c_0}{\mathbb{E}(\hat{T}(x))^2}\left(\mathbb{V}\left(\hat{T}(x)\right) + \mathbb{E}\left[\left(I_T - \gamma(x)\hat{T}(x)\right)^2\right]\right) \\
&\leq \frac{c_0}{\mathbb{E}(\hat{T}(x))^2}\left(\mathbb{V}\left(\hat{T}(x)\right) + 3\mathbb{V}\left(I_T\right) + 3\mathbb{V}\left(\gamma(x)\hat{T}(x)\right) + 3\left(\mathbb{E}(I_T) - \gamma(x)\mathbb{E}(\hat{T}(x))\right)^2\right).
\end{aligned}$$

By (4), we know that

$$\left|\mathbb{E}(I_T) - T\int_E \gamma(z)k_T(x,z)\mu_\infty(dz)\right| \le c_0\alpha^* k^*$$

and

$$\left|\gamma(x)\mathbb{E}(\hat{T}(x)) - T\gamma(x)\int_E k_T(x,z)\mu_\infty(dz)\right| \le c_0\alpha^* k^*$$

so, using the notation from (H2)-(H3),

$$\left|\mathbb{E}(I_T) - \gamma(x)\mathbb{E}(\hat{T}(x))\right| \le 2c_0\alpha^* k^* + T\left|\int_E (\gamma(z)-\gamma(x))k_T(x,z)\mu_\infty(dz)\right| = c_1 + Tv_T(x)|w_T(x)|$$

where $c_1 = 2c_0\alpha^* k^*$. Moreover, from (5), $\mathbb{V}(I_T) \le c_0\mathbb{E}(I_T) \le c_0\alpha^*\mathbb{E}(\hat{T}(x))$ and by (10), we obtain

$$E\left(R_T^2\right) \le \frac{c_0}{(Tv_T(x))^2}\left(Tv_T(x) + (c_1 + Tv_T(x)|w_T(x)|)^2\right) \le \frac{2c_0}{(Tv_T(x))^2}\left(Tv_T(x) + c_1^2 + T^2 v_T^2(x)w_T^2(x)\right)$$

which implies (27) under (H2).

5.6 Proof of Proposition 8

Using the same decomposition and the same notation as in the proof of Proposition 6, we get for T large enough

$$\begin{aligned}\mathbb{E}((\hat{\gamma}_\epsilon(x) - \gamma(x))^2) &\le 3\,\mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}M_T^2(k)\right) + 3\,\mathbb{E}(R_T^2) + 3(\alpha^*)^2\mathbb{P}\left(\hat{T}(x)\le\epsilon\right)\\ &\le 3\,\mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}M_T^2(k)\right) + c_0\left(\frac{1}{Tv_T(x)} + w_T^2(x)\right), \end{aligned} \tag{28}$$

where we write $M_T(k)$ for M_T to highlight the dependence in k. With the latter convention, let us introduce $\tilde{M}_T = M_T^2(k) - \int_0^T k_T^2(x, X_{s^-})dN_s^\gamma$ and set $I_T(k^2) = \int_0^T k_T^2(x,X_s)\gamma(X_s)ds$. We have

$$\begin{aligned}\mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}M_T^2(k)\right) &= \mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}\tilde{M}_T\right) + \mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}M_T(k^2)\right) + \mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)^2}I_T(k^2)\right)\\ &=: A_1 + A_2 + A_3. \end{aligned} \tag{29}$$

For the third term A_3, since $\gamma(x) \le \alpha(x)$,

$$A_3 \le k^*\alpha^*\mathbb{E}\left(\frac{\mathbf{1}_{\hat{T}(x)>\epsilon}}{\hat{T}(x)}\right) \le \frac{k^*\alpha^*}{\mathbb{E}[\hat{T}(x)]} + k^*\alpha^*\mathbb{E}\left(\left|\frac{1}{\hat{T}(x)} - \frac{1}{\mathbb{E}[\hat{T}(x)]}\right|\mathbf{1}_{\hat{T}(x)>\epsilon}\right) \le \frac{k^*\alpha^*}{\mathbb{E}[\hat{T}(x)]} + \frac{k^*\alpha^*}{\varepsilon}\frac{\sqrt{\mathbb{V}(\hat{T}(x))}}{\mathbb{E}[\hat{T}(x)]}.$$

So for T large enough, by (10),

$$A_3 \le \frac{c_0}{\sqrt{Tv_T(x)}}. \tag{30}$$

For the second term A_2, we have

$$\begin{aligned}A_2 =& \mathbb{E}\left[\frac{M_T(k^2)}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)>Tv_T(x)} + \mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)\le Tv_T(x)}M_T(k^2)\left(\frac{1}{\hat{T}(x)} - \frac{1}{\mathbb{E}[\hat{T}(x)]} + \frac{1}{\mathbb{E}[\hat{T}(x)]}\right)^2\right]\\ \le& \mathbb{E}\left[\frac{|M_T(k^2)|}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)>Tv_T(x)}\right] + 2\mathbb{E}\left[\frac{|M_T(k^2)|}{\mathbb{E}[\hat{T}(x)]^2}\frac{(\hat{T}(x)-\mathbb{E}[\hat{T}(x)])^2}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)\le Tv_T(x)}\right]\\ &+ 2\mathbb{E}\left[\frac{|M_T(k^2)|}{\mathbb{E}[\hat{T}(x)]^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)\le Tv_T(x)}\right]\\ =:& A_{2,1} + A_{2,2} + A_{2,3}. \end{aligned}$$

Let us control each term. First note that by Jensen inequality and Proposition 4,

$$\mathbb{E}[|M_T(k^2)|]^2 \leq \mathbb{E}[|M_T(k^2)|^2] = \mathbb{E}\left(\int_0^T k_T^4(x, X_s)\gamma(X_s)ds\right) \leq c_0\mathbb{E}[\hat{T}(x)] \leq c_0 T v_T(x).$$

Therefore, since $I_T(k^2) \leq k^*\alpha^*\hat{T}(x)$,

$$A_{2,1} = \mathbb{E}\left[\frac{|M_T(k^2)|}{I_T(k^2)^2}\frac{I_T(k^2)^2}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)>Tv_T(x)}\right] \leq \frac{(k^*\alpha^*)^2}{(Tv_T(x)^2}\mathbb{E}[|M_T(k^2)|] \leq \frac{c_0}{(Tv_T(x))^{\frac{3}{2}}} \tag{31}$$

and for T large enough,

$$A_{2,3} \leq 2\frac{\mathbb{E}[|M_T(k^2)|]}{\mathbb{E}[\hat{T}(x)]^2} \leq \frac{c_0}{(Tv_T(x))^{\frac{3}{2}}}. \tag{32}$$

For the term $A_{2,2}$, let $\theta \in (\frac{1}{2}, 1)$ and consider the decomposition

$$\begin{aligned}
A_{2,2} =& \frac{2}{\mathbb{E}[\hat{T}(x)]^2}\mathbb{E}\left[|M_T(k^2)|\frac{(\hat{T}(x)-\mathbb{E}[\hat{T}(x)])^2}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}\mathbf{1}_{|M_T(k^2)|\leq(Tv_T(x))^\theta\hat{T}(x)^2}\right] \\
&+ \frac{2}{\mathbb{E}[\hat{T}(x)]^2}\mathbb{E}\left[|M_T(k^2)|\frac{(\hat{T}(x)-\mathbb{E}[\hat{T}(x)])^2}{\hat{T}(x)^2}\mathbf{1}_{\hat{T}(x)>\varepsilon}\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}\mathbf{1}_{|M_T(k^2)|>(Tv_T(x))^\theta\hat{T}(x)^2}\right] \\
\leq& \frac{2(Tv_T(x))^\theta\mathbb{V}(\hat{T}(x))}{\mathbb{E}[\hat{T}(x)]^2} + c_0\mathbb{E}[|M_T(k^2)|\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}\mathbf{1}_{|M_T(k^2)|>(Tv_T(x))^\theta\varepsilon^2}] \\
\leq& c_0\left((Tv_T(x))^{\theta-1} + \mathbb{E}[|M_T(k^2)|\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}\mathbf{1}_{|M_T(k^2)|>(Tv_T(x))^\theta\varepsilon^2}]\right).
\end{aligned} \tag{33}$$

Letting $Z = M_T(k^2)\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}$,

$$\begin{aligned}
\mathbb{E}[|M_T(k^2)|\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}\mathbf{1}_{|M_T(k^2)|>(Tv_T(x))^\theta\varepsilon^2}] &= \mathbb{E}[|Z|\mathbf{1}_{|Z|>(Tv_T(x))^\theta\varepsilon^2}] \\
&= \int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty}\mathbb{P}(|Z|\geq u)du + (Tv_T(x))^\theta\varepsilon^2\mathbb{P}(|Z|>(Tv_T(x))^\theta\varepsilon^2).
\end{aligned} \tag{34}$$

Since $Z = \int_0^T k_T^2(x, X_{s^-})\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}[dN_s^\gamma - \gamma(X_s)ds]$, Theorem 1 of Le Guével (2019) with $H_s = k_T^2(x, X_{s^-})\mathbf{1}_{I_T(k^2)\leq Tv_T(x)}$, $\|H\|_\infty = (k^*)^2$ and $\|H\|_2^2 = (k^*)^2 Tv_T(x)$ leads for $u \geq 0$ to

$$\mathbb{P}(|Z|\geq u) \leq 2\exp\left(-\frac{Tv_T(x)}{(k^*)^2}I\left(\frac{u}{Tv_T(x)}\right)\right) \tag{35}$$

where $I(u) = (1+u)\log(1+u) - u$. Using the fact that $I(u) \sim_0 u^2/2$, (35) leads to

$$(Tv_T(x))^\theta\varepsilon^2\mathbb{P}(|Z|\geq(Tv_T(x))^\theta\varepsilon^2) \leq 2(Tv_T(x))^\theta\varepsilon^2\exp\left(-\frac{Tv_T(x)}{(k^*)^2}I\left(\frac{\varepsilon^2}{(Tv_T(x))^{1-\theta}}\right)\right) \leq \frac{c_0}{Tv_T(x)}. \tag{36}$$

Moreover, thanks again to (35) we have

$$\int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty}\mathbb{P}(|Z|\geq u)du \leq 2\int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty}\exp\left(-\frac{Tv_T(x)}{(k^*)^2}I\left(\frac{u}{Tv_T(x)}\right)\right)du$$

and by the inequality $I(u) \geq (u^2/3)\mathbf{1}_{0\leq u<1} + (u/3)\mathbf{1}_{u\geq 1}$ we get

$$\begin{aligned}\int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty} \mathbb{P}(|Z|\geq u)du &\leq 2\int_{(Tv_T(x))^\theta\varepsilon^2}^{Tv_T(x)} \exp\left(-\frac{u^2}{3(k^*)^2 Tv_T(x)}\right)du + 2\int_{Tv_T(x)}^{+\infty}\exp\left(-\frac{u}{3(k^*)^2}\right)du \\ &\leq c_0\left(\sqrt{Tv_T(x)}\int_{(Tv_T(x))^{2\theta-1}\varepsilon^4}^{Tv_T(x)} e^{-\frac{\omega}{3(k^*)^2}}\frac{d\omega}{\sqrt{\omega}} + e^{-\frac{Tv_T(x)}{3(k^*)^2}}\right) \\ &\leq c_0\left((Tv_T(x))^{1-\theta}e^{-\frac{\varepsilon^4(Tv_T(x))^{2\theta-1}}{3(k^*)^2}} + e^{-\frac{Tv_T(x)}{3(k^*)^2}}\right) \leq \frac{c_0}{Tv_T(x)}. \end{aligned} \tag{37}$$

Gathering (33), (34), (36) and (37), we obtain $A_{2,2} \leq c_0(Tv_T(x))^{\theta-1}$, which combined with (31) and (32) provides

$$|A_2| \leq \frac{c_0}{(Tv_T(x))^{1-\theta}}. \tag{38}$$

It remains to control the term A_1 in (29). Using exactly the same decomposition and arguments than for A_2, along with the inequality $\mathbb{E}[|\tilde{M}_T|] \leq c_0\mathbb{E}[\hat{T}(x)]$, we obtain for $\theta \in (0,1)$ and $\rho \in (\frac{1}{2},1)$ such that $\theta/2 < \rho < \theta$

$$\mathbb{E}[|A_1] \leq \frac{c_0}{(Tv_T(x))^{2\rho-1}} + \frac{c_0}{(Tv_T(x))^{1-\theta}} + c_0\mathbb{E}\left[|\tilde{M}_T\mathbf{1}_{I_T(k^2)\leq(Tv_T(x))^\rho}\mathbf{1}_{|\tilde{M}|>(Tv_T(x))^\theta\varepsilon^2}\right]. \tag{39}$$

To control the last term in (39), we can use (34) where now $Z = \tilde{M}_T\mathbf{1}_{I_T(k^2)\leq(Tv_T(x))^\rho}$. Theorem 2 of Le Guével (2019) with $H_s = k_T(x, X_{s^-})\mathbf{1}_{I_T(k^2)\leq(Tv_T(x))^\rho}$, $\|H\|_\infty = k^*$ and $\|H\|_2^2 = (Tv_T(x))^\rho$ leads for $u \geq 0$ to

$$\mathbb{P}(|Z|\geq u) \leq 6\exp\left(-\frac{(Tv_T(x))^\rho}{(k^*)^2} I\left(\frac{k^*}{(Tv_T(x))^\rho}\sqrt{\frac{u}{2}}\right)\right), \tag{40}$$

which gives for $u = (Tv_T(x))^\theta\varepsilon^2$, since $I(u) \sim_0 \frac{u^2}{2}$ and $\theta/2 < \rho < \theta$,

$$(Tv_T(x))^\theta\varepsilon^2\mathbb{P}\left(|Z|\geq (Tv_T(x))^\theta\varepsilon^2\right) \leq \frac{c_0}{Tv_T(x)}. \tag{41}$$

Moreover, using again (40) and the same arguments as to get (37), we obtain

$$\int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty}\mathbb{P}(|Z|\geq u)du \leq 6\int_{(Tv_T(x))^\theta\varepsilon^2}^{+\infty}\exp\left(-\frac{(Tv_T(x))^\rho}{(k^*)^2}I\left(\frac{k^*}{(Tv_T(x))^\rho}\sqrt{\frac{u}{2}}\right)\right)du \leq \frac{c_0}{Tv_T(x)}.$$

Coming back to (39), we deduce that

$$|A_1| \leq \frac{c_0}{(Tv_T(x))^{2\rho-1}} + \frac{c_0}{(Tv_T(x))^{1-\theta}}. \tag{42}$$

The best rate is clearly achieve for the highest values of ρ, or writing $\rho = \theta - \eta/2$ with $0 < \eta < 2\theta - 1$, for the smallest values of η. The best trade-off in θ between $2\rho - 1 = 2\theta - 1 - \eta$ and $1 - \theta$ is obtained for $\theta = 2/3$. This result combined with (30) and (38) in (28) concludes the proof.

5.7 Proof of Proposition 9

We detail the proof for $\gamma = \alpha$, the other cases being similar, and we write D_j for D_j^α to save notation. Recall that $\hat{T}_{(d)}(x) = \sum_{j=0}^{m-1}\Delta t_{j+1}k_T(x, X_{t_j})$. To save space, we write $\hat{T}_{(d)}$ for $\hat{T}_{(d)}(x)$, $\hat{T}$ for $\hat{T}(x)$, ℓ_T

for $\ell_T(x)$ and v_T for $v_T(x)$. We have

$$\begin{aligned}|\hat{\alpha}_{(d)}(x)-\alpha(x)| \leq & \frac{1}{\hat{T}_{(d)}}\sum_{j=0}^{m-1} k_T(x,X_{t_j})|D_{j+1}-\Delta N_{t_{j+1}}| + \frac{1}{\hat{T}_{(d)}}\left|\sum_{j=0}^{m-1} k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_0^T k_T(x,X_{s^-})dN_s\right| \\ & + \frac{\int_0^T k_T(x,X_{s^-})dN_s}{\hat{T}_{(d)}\hat{T}}|\hat{T}-\hat{T}_{(d)}| + |\hat{\alpha}(x)-\alpha(x)| \\ \leq & \frac{k^*}{\hat{T}_{(d)}}\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2} + \frac{1}{\hat{T}_{(d)}}\sum_{j=0}^{m-1}|k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_{t_j}^{t_{j+1}} k_T(x,X_{s^-})dN_s| \\ & + \frac{k^*N_T}{\hat{T}_{(d)}\hat{T}}|\hat{T}-\hat{T}_{(d)}| + |\hat{\alpha}(x)-\alpha(x)|.\end{aligned}$$

Concerning the second term in this sum we have

$$\left|k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_{t_j}^{t_{j+1}} k_T(x,X_{s^-})dN_s\right|\mathbf{1}_{\Delta N_{t_{j+1}}=0} = 0,$$

$$\begin{aligned}&\left|k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_{t_j}^{t_{j+1}} k_T(x,X_{s^-})dN_s\right|\mathbf{1}_{\Delta N_{t_{j+1}}=1} \\ &\qquad = \left|k_T(x,Y^{(n(X_{t_j}))}_{t_j-T_{N_{t_j}}}) - k_T(x,Y^{(n(X_{t_j}))}_{T^-_{N_{t_{j+1}}}-T_{N_{t_j}}})\right|\mathbf{1}_{\Delta N_{t_{j+1}}=1} \leq \ell_T\Delta_m^a\mathbf{1}_{\Delta N_{t_{j+1}}=1}\end{aligned}$$

and

$$\left|k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_{t_j}^{t_{j+1}} k_T(x,X_{s^-})dN_s\right|\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2} \leq 2k^*\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2},$$

so that

$$\begin{aligned}\sum_{j=0}^{m-1}|k_T(x,X_{t_j})\Delta N_{t_{j+1}} - \int_{t_j}^{t_{j+1}} k_T(x,X_{s^-})dN_s| &\leq \sum_{j=0}^{m-1}\ell_T\Delta_m^a\mathbf{1}_{\Delta N_{t_{j+1}}=1} + 2k^*\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2} \\ &\leq \sum_{j=0}^{m-1}\ell_T\Delta_m^a\Delta N_{t_{j+1}} + 2k^*\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2} \\ &\leq \ell_T\Delta_m^a N_T + 2k^*\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2}.\end{aligned}$$

On the other hand,

$$\begin{aligned}|\hat{T}-\hat{T}_{(d)}| &\leq \sum_{j=0}^{m-1}\int_{t_j}^{t_{j+1}}\left|k_T(x,X_s)-k_T(x,X_{t_j})\right|ds \\ &\leq \sum_{j=0}^{m-1}\int_{t_j}^{t_{j+1}}\left|k_T(x,X_s)-k_T(x,X_{t_j})\right|ds\mathbf{1}_{\Delta N_{t_{j+1}}=0} + 2k^*\Delta_m\sum_{j=0}^{m-1}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 1} \\ &\leq \sum_{j=0}^{m-1}\Delta t_{j+1}\ell_T\Delta_m^a + 2k^*\Delta_m\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 1} \\ &\leq \ell_T\Delta_m^a T + 2k^*\Delta_m N_T.\end{aligned}$$

We thus obtain

$$|\hat{\alpha}_{(d)}(x)-\alpha(x)| \leq \frac{3k^*}{\hat{T}_{(d)}}\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2}+\frac{\ell_T\Delta_m^a N_T}{\hat{T}_{(d)}}+\frac{k^*N_T}{\hat{T}_{(d)}\hat{T}}(\ell_T\Delta_m^a T+2k^*\Delta_m N_T)+|\hat{\alpha}(x)-\alpha(x)|. \tag{43}$$

The last term has been treated in Proposition 6. Let us consider the other terms. First note that

$$\begin{aligned}\mathbb{P}(\hat{T}_{(d)}<Tv_T/4) &\leq \mathbb{P}(|\hat{T}-\hat{T}_{(d)}|>Tv_T/4)+\mathbb{P}(\hat{T}<Tv_T/2)\\ &\leq \mathbb{P}(N_T>\frac{Tv_T/4-\ell_T\Delta_m^a T}{2k^*\Delta_m})+\frac{c_0}{Tv_T}\\ &\leq \mathbb{E}(N_T)\frac{2k^*\Delta_m}{Tv_T/4-\ell_T\Delta_m^a T}+\frac{c_0}{Tv_T}.\end{aligned}$$

By Proposition 4 applied with $g=1$, we get that $\mathbb{E}(N_T)=\mathbb{E}\int_0^T\alpha(X_s)ds\leq\alpha^*T$. The first term above is thus lower (up to a constant) to $\Delta_m/(v_T-4\ell_T\Delta_m^a)$ which can be written

$$v_T\frac{\Delta_m}{v_T^2}\left(1-4v_T\frac{\ell_T\Delta_m^a}{v_T^2}\right)^{-1}\leq c_0\frac{\Delta_m}{v_T^2}$$

in view of (H6) and the fact that v_T is bounded. We deduce from (H5) that $\mathbb{P}(\hat{T}_{(d)}<Tv_T/4)$ tends to 0 and for the rest of the proof we place ourselves on the event $\{\hat{T}_{(d)}\geq Tv_T/4\}$. The following lemma is verified in Section 5.8.

Lemma 10. *For any $j=1,\dots,m$ and any $i\in\mathbb{N}$, $\mathbb{P}(\Delta N_{t_j}=i)\leq(\alpha^*\Delta t_j)^i/(i!)$.*

For any $\epsilon>0$, using this lemma,

$$\begin{aligned}\mathbb{P}\left(\frac{3k^*}{\hat{T}_{(d)}}\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2}>\epsilon,\hat{T}_{(d)}\geq\frac{Tv_T}{4}\right) &\leq \mathbb{P}\left(\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2}>\frac{\epsilon Tv_T}{12k^*}\right)\\ &\leq \frac{12k^*}{\epsilon Tv_T}\sum_{j=0}^{m-1}\mathbb{E}(\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2})\\ &\leq \frac{12k^*}{\epsilon Tv_T}\sum_{j=0}^{m-1}\sum_{i\geq 2}i\mathbb{P}(\Delta N_{t_{j+1}}=i)\\ &\leq \frac{12k^*}{\epsilon Tv_T}m\sum_{i\geq 2}i\frac{(\alpha^*\Delta_m)^i}{i!}=\frac{12k^*m\alpha^*\Delta_m}{\epsilon Tv_T}(e^{\alpha^*\Delta_m}-1).\end{aligned}$$

This last expression is lower than $m\Delta_m^2/(Tv_T)$, up to a positive constant, because $\Delta_m\to 0$ as a consequence of (H5). Since by assumption $m\Delta_m/T$ is uniformly bounded and v_T is also bounded, we deduce that

$$\mathbb{P}\left(\frac{3k^*}{\hat{T}_{(d)}}\sum_{j=0}^{m-1}\Delta N_{t_{j+1}}\mathbf{1}_{\Delta N_{t_{j+1}}\geq 2}>\epsilon\right)\leq c_0\frac{\Delta_m}{v_T^2}.$$

Next,

$$\mathbb{P}\left(\frac{\ell_T\Delta_m^a N_T}{\hat{T}_{(d)}}>\epsilon,\hat{T}_{(d)}\geq\frac{Tv_T}{4}\right)\leq\mathbb{P}\left(N_T>\frac{\epsilon Tv_T}{4\ell_T\Delta_m^a}\right)\leq\mathbb{E}(N_T)\frac{4\ell_T\Delta_m^a}{\epsilon Tv_T}\leq\frac{4\alpha^*\ell_T\Delta_m^a}{\epsilon v_T}$$

and finally

$$\begin{aligned}
&\mathbb{P}\left(\frac{k^*N_T}{\hat{T}_{(d)}\hat{T}}(\ell_T\Delta_m^a T + 2k^*\Delta_m N_T) > \epsilon, \hat{T}_{(d)} \geq \frac{Tv_T}{4}\right) \\
&\leq \mathbb{P}\left(\frac{k^*N_T}{\hat{T}}(\ell_T\Delta_m^a T + 2k^*\Delta_m N_T) > \frac{\epsilon T v_T}{4}\right) \\
&\leq \mathbb{P}\left(\frac{k^*N_T}{\hat{T}}(\ell_T\Delta_m^a T + 2k^*\Delta_m N_T) > \frac{\epsilon T v_T}{4}, \hat{T} \geq \frac{Tv_T}{2}\right) + \mathbb{P}\left(\hat{T} < \frac{Tv_T}{2}\right) \\
&\leq \mathbb{P}\left(k^*N_T(\ell_T\Delta_m^a T + 2k^*\Delta_m N_T) > \frac{\epsilon(Tv_T)^2}{8}\right) + \frac{c_0}{Tv_T} \\
&\leq \frac{8k^*}{\epsilon(Tv_T)^2}\left(\ell_T\Delta_m^a T\,\mathbb{E}(N_T) + 2k^*\Delta_m\mathbb{E}(N_T^2)\right) + \frac{c_0}{Tv_T} \\
&\leq \frac{c_0\ell_T\Delta_m^a}{v_T^2} + \frac{c_0\Delta_m}{v_T^2} + \frac{c_0}{Tv_T}
\end{aligned}$$

where we have used the fact that $\mathbb{E}(N_T^2) < c_0T^2$ (that can be proven as a consequence of Proposition 4).

5.8 Proof of Lemmas

5.8.1 Proof of Lemma 5

For any $t \geq 0$ and any n,

$$\begin{aligned}
\mathbb{P}(N_t \geq n) = \mathbb{P}(T_n \leq t) &= \mathbb{P}(0 \leq T_n - T_{n-1} \leq t - T_{n-1}) \\
&= \mathbb{E}\left(\mathbf{1}_{t\geq T_{n-1}}\mathbb{P}\left(T_n - T_{n-1} \leq t - T_{n-1} | T_{n-1}, X_{T_{n-1}}, Y^{(n(X_{T_{n-1}}))}\right)\right) \\
&= \mathbb{E}\left(\mathbf{1}_{t\geq T_{n-1}}\left(1 - e^{-\int_0^{t-T_{n-1}}\alpha(Y_u^{(n(X_{T_{n-1}}))})du}\right)\right).
\end{aligned}$$

We deduce that

$$\mathbb{E}\left(\mathbf{1}_{t\geq T_{n-1}}\left(1 - e^{-\alpha_*(t-T_{n-1})}\right)\right) \leq \mathbb{P}(N_t \geq n) \leq \mathbb{E}\left(\mathbf{1}_{t\geq T_{n-1}}\left(1 - e^{-\alpha^*(t-T_{n-1})}\right)\right).$$

Let e_* and e^* be two random variables independent of T_{n-1} distributed according to an exponential distribution with rate α_* and α^*, respectively. The bounds above are nothing else than $\mathbb{P}(e_* \leq t - T_{n-1})$ and $\mathbb{P}(e^* \leq t - T_{n-1})$ so that

$$\mathbb{P}(e_* \leq t - T_{n-1}) \leq \mathbb{P}(N_t \geq n) \leq \mathbb{P}(e^* \leq t - T_{n-1}). \tag{44}$$

We start from the lower bound to prove the first inequality of the lemma. Denote by γ_j the probability density function of a Gamma distribution with parameters j and α_*. We have that

$$\mathbb{P}(e_* \leq t - T_{n-1}) = \int_0^t \mathbb{P}(T_{n-1} \leq t - s)\gamma_1(s)ds,$$

so for any n,

$$\mathbb{P}(T_n \leq t) = \mathbb{P}(N_t \geq n) \geq \int_0^t \mathbb{P}(T_{n-1} \leq t - s)\gamma_1(s)ds.$$

Iterating this inequality, we obtain

$$\begin{aligned}
\int_0^t \mathbb{P}(T_{n-1} \leq t - s)\gamma_1(s)ds &\geq \int_0^t\int_0^{t-s} \mathbb{P}(T_{n-2} \leq t - s - u)\gamma_1(u)du\gamma_1(s)ds \\
&= \int_0^t \mathbb{P}(T_{n-2} \leq t - z)\left(\int_0^z \gamma_1(z-s)\gamma_1(s)ds\right)dz \\
&= \int_0^t \mathbb{P}(T_{n-2} \leq t - z)\gamma_2(z)dz
\end{aligned}$$

and recursively

$$\mathbb{P}(N_t \geq n) \geq \int_0^t \mathbb{P}(T_{n-1} \leq t-s)\gamma_1(s)ds \geq \int_0^t \gamma_n(z)dz.$$

This lower bound equals $\mathbb{P}(N_* \geq n)$ where $N_* \sim \mathcal{P}(\alpha_* t)$, which proves one inequality in Lemma 5. The other inequality is obtained similarly by starting from the upper-bound in (44) and by replacing γ_j with a Gamma distribution with parameters j and α^*.

5.8.2 Proof of Lemma 7

We have

$$\begin{aligned}\mathbb{E}(\hat{\alpha}(x)) &\geq \mathbb{E}(\hat{\alpha}(x)\mathbf{1}_{N_T=1}\mathbf{1}_{X_0\in E_{n(x)}}) \\ &= \mathbb{E}\left(\frac{k_T(x, X_{T_1^-})}{\int_0^{T_1} k_T(x, Y_s^{(n(X_0))})ds + \int_{T_1}^T k_T(x, Y_{s-T_1}^{(n(X_{T_1}))})ds}\mathbf{1}_{T_1\leq T}\mathbf{1}_{T_2-T_1>T-T_1}\mathbf{1}_{X_0\in E_{n(x)}}\right).\end{aligned}$$

Since by assumption $k_T(x,y) = \mathbf{1}_{n(x)=n(y)}$, we have $k_T(x, X_{T_1^-})\mathbf{1}_{X_0\in E_{n(x)}} = k_T(x, X_0)\mathbf{1}_{X_0\in E_{n(x)}} = 1$. Similarly for all $s>0$, $k_T(x, Y_s^{(n(X_0))})\mathbf{1}_{X_0\in E_{n(x)}} = 1$ and $k_T(x, Y_{s-T_1}^{(n(X_{T_1}))})\mathbf{1}_{X_0\in E_{n(x)}} = 0$. Therefore

$$\begin{aligned}\mathbb{E}(\hat{\alpha}(x)) &\geq \mathbb{E}\left(\frac{1}{T_1}\mathbf{1}_{T_1\leq T}\mathbf{1}_{T_2-T_1>T-T_1}\mathbf{1}_{X_0\in E_{n(x)}}\right) \\ &= \mathbb{E}\left(\mathbf{1}_{X_0\in E_{n(x)}}\int_0^T \frac{1}{s}\alpha(Y_s^{(n(x))})e^{-\int_0^s \alpha(Y_u^{(n(x))})du}e^{-\int_0^{T-s}\alpha(Y_u^{(n(X_s))})du}ds\right) \\ &\geq \alpha_* e^{-\alpha^* T}\mu_0(E_{n(x)})\int_0^T \frac{1}{s}e^{-\alpha^* s}ds = \infty.\end{aligned}$$

5.8.3 Proof of Lemma 10

By the Markov property, $\mathbb{P}(\Delta N_{t_j} = i) = \mathbb{E}[\mathbb{P}(\Delta N_{t_j} = i|X_{t_j})] = \mathbb{E}[\mathbb{P}(N_{\Delta t_j} = i|X_0 = X_{t_j})]$. The lemma is proved once we verify that

$$\forall t \geq 0, \forall i \in \mathbb{N}, \quad \sup_{x\in E}\mathbb{P}(N_t = i|X_0 = x) \leq \frac{(\alpha^* t)^i}{i!}.$$

This inequality is obvious if $i = 0$. Let $i \geq 1$,

$$\mathbb{P}(N_t = i|X_0 = x) = \mathbb{E}\left[\mathbb{P}\left(T_1 < T_2 < ... < T_i < t < T_{i+1}\Big|X_0 = x, n(X_{T_j}), (Y_t^{(n(X_{T_j}))})_{t\geq 0}, \forall j \geq 0\right)\Big|X_0 = x\right].$$

Let $\tau_j = T_j - T_{j-1}$. Conditional on $n(X_{T_j})$ and $(Y_t^{(n(X_{T_j}))})_{t\geq 0}$, the random variable τ_j has density

$$t \mapsto \alpha(Y_t^{(n(X_{T_j}))})e^{-\int_0^t \alpha(Y_u^{(n(X_{T_j}))})du},$$

whereby

$$\mathbb{P}\left(T_1 < T_2 < ... < T_i < t < T_{i+1} \Big| X_0 = x, n(X_{T_j}), (Y_t^{(n(X_{T_j}))})_{t\geq 0}, \forall j \geq 0\right)$$
$$= \mathbb{P}\left(\tau_1 < t, \tau_2 < t - \tau_1, \ldots, \tau_i < t - \sum_{j=1}^{i-1} \tau_j, \tau_{i+1} > t - \sum_{j=1}^{i} \tau_j \Big| X_0 = x, n(X_{T_j}), (Y_t^{(n(X_{T_j}))})_{t\geq 0}, \forall j \geq 0\right)$$
$$= \int_0^t \alpha(Y_{s_1}^{(n(X_0))}) e^{-\int_0^{s_1} \alpha(Y_u^{(n(X_0))}) du} ds_1 \int_0^{t-s_1} \alpha(Y_{s_2}^{(n(X_{T_1}))}) e^{-\int_0^{s_2} \alpha(Y_u^{(n(X_{T_1}))}) du} ds_2$$
$$\cdots \int_0^{t-\sum_{j=1}^{i-1} s_j} \alpha(Y_{s_i}^{(n(X_{T_{i-1}}))}) e^{-\int_0^{s_i} \alpha(Y_u^{(n(X_{T_{i-1}}))}) du} ds_i \, e^{-\int_0^{t-\sum_{j=1}^{i} s_j} \alpha(Y_u^{(n(X_{T_i}))}) du}$$
$$\leq (\alpha^*)^i \int_0^t ds_1 \int_0^{t-s_1} ds_2 ... \int_0^{t-\sum_{j=1}^{i-1} s_j} ds_k = (\alpha^*)^i \int_{\mathbb{R}^i} \mathbf{1}_{t_1<t_2<...<t_i<t} dt_1 \ldots dt_i = \frac{(\alpha^* t)^i}{i!}.$$

A Technical lemmas

Lemma 11. *Let $\mathcal{F} = (\mathcal{F}_t)_{t\geq 0}$ be a filtration, complete and right-continuous, Z be a random variable in $L^1(\Omega)$, and T a $\mathcal{F}$-stopping time. Set $T_n = \sum_{k\geq 1} \frac{k}{2^n} \mathbf{1}_{\frac{k-1}{2^n} \leq T < \frac{k}{2^n}}$. Then*

$$\mathbb{E}[Z|\mathcal{F}_T] = \lim_{n\to+\infty} \mathbb{E}[Z|\mathcal{F}_{T_n}].$$

Proof. Set $M_t = \mathbb{E}[Z|\mathcal{F}_t]$. M is a $\mathcal{F}$-martingale. Set $\mathcal{F}_{t^+} = \bigcap_{r>t} \mathcal{F}_r$, then $M_{t^+} := \lim_{s\to t, s>t, s\in\mathbb{Q}} M_s$ exists and is $\mathcal{F}_{t^+}$-measurable. Let $A \in \mathcal{F}_{t^+}$ and $s > t$, $s \in \mathbb{Q}$, then since $A \in \mathcal{F}_s$,

$$\mathbb{E}[M_s \mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_{t^+}]\mathbf{1}_A].$$

Taking the limit leads to $\mathbb{E}[M_{t^+}\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[Z|\mathcal{F}_{t^+}]\mathbf{1}_A]$.

Since M_{t^+} is $\mathcal{F}_{t^+}$-measurable, $M_{t^+} = \mathbb{E}[Z|\mathcal{F}_{t^+}]$ almost surely. The right-continuity of $\mathcal{F}$ ($\mathcal{F}_t = \mathcal{F}_{t^+}$) gives $M_{t^+} = \mathbb{E}[Z|\mathcal{F}_t] = M_t$, meaning that M is right-continuous for sequences taking values in $\mathbb{Q}$. Now if T is a stopping time, T_n is a sequence of stopping times decreasing to T and taking values in $\mathbb{Q}$ so

$$M_T = \lim_{n\to+\infty} M_{T_n}.$$

□

Let $(X_t^{(1)})_{t\geq 0}$ and $(X_t^{(2)})_{t\geq 0}$ be two independent copies of the process X, with respective initial distributions ϕ_1 and ϕ_2, and jumping times $(T_j^{(1)})_{j\geq 0}$ and $(T_j^{(2)})_{j\geq 0}$. We shall write $(\mathcal{F}_t^{(k)})_{t\geq 0}$ for the completed right-continuous natural filtration of $X^{(k)}$, $k = 1, 2$, and $\mathcal{G}_t = \sigma(\mathcal{F}_t^{(1)}, \mathcal{F}_t^{(2)})$. Consider

$$\tau = \inf\{t > 0 : (X_t^{(1)}, X_t^{(2)}) = (\varnothing, \varnothing)\}$$

and for $k = 1, 2$,

$$T_+^{(k)} = \inf\{t > \tau : X_t^{(k)} \neq \varnothing\}.$$

Lemma 12. *For every positive t,*

$$\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t + \tau|\mathcal{G}_\tau) = \mathbb{P}(T_1^{(1)} > t|X_0^{(1)} = \varnothing)$$
$$\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(2)} > t + \tau|\mathcal{G}_\tau) = \mathbb{P}(T_1^{(2)} > t|X_0^{(2)} = \varnothing)$$

thereby we have the following equalities in distribution, denoted by $\stackrel{\mathcal{L}}{=}$,

$$\left[(T_+^{(1)} - \tau)|\mathcal{G}_\tau\right] \stackrel{\mathcal{L}}{=} \left[(T_+^{(2)} - \tau)|\mathcal{G}_\tau\right] \stackrel{\mathcal{L}}{=} \left[T_1^{(1)}|(X_0^{(1)} = \varnothing)\right] \stackrel{\mathcal{L}}{=} \left[T_1^{(2)}|(X_0^{(2)} = \varnothing)\right].$$

Proof. We only prove the first equality since the second one is obtained similarly. Given the two first identities, the last statement of the lemma is straightforward. For any $t > 0$,

$$\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau) = \sum_{j\geq 1}\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(1)}} + \mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(2)}}. \quad (45)$$

On one hand, for any $j \geq 1$,

$$\begin{aligned}\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(1)}} &= \mathbb{P}_{\phi_1\times\phi_2}(T_{j+1}^{(1)} > t+T_j^{(1)}|\mathcal{G}_{T_j^{(1)}})\mathbf{1}_{\tau=T_j^{(1)}}\\ &= \mathbb{P}_{\phi_1}(T_1^{(1)} > t|X_0^{(1)} = X_{T_j^{(1)}}^{(1)})\mathbf{1}_{\tau=T_j^{(1)}}\end{aligned}$$

by construction of the process $X^{(1)}$ and so

$$\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(1)}} = \mathbb{P}(T_1^{(1)} > t|X_0^{(1)} = \varnothing)\mathbf{1}_{\tau=T_j^{(1)}}. \quad (46)$$

On the other hand

$$\begin{aligned}\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(2)}} &= \sum_{k\geq 1}\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t+\tau|\mathcal{G}_\tau)\mathbf{1}_{\tau=T_j^{(2)}}\mathbf{1}_{T_{k-1}^{(1)}\leq\tau<T_k^{(1)}}\\ &= \sum_{k\geq 1}\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+T_j^{(2)}|\mathcal{G}_{T_j^{(2)}})\mathbf{1}_{\tau=T_j^{(2)}}\mathbf{1}_{T_{k-1}^{(1)}\leq\tau<T_k^{(1)}}. \quad (47)\end{aligned}$$

Consider now $(T_{n,j}^{(2)})_{n\geq 1}$ a sequence of stopping times, decreasing to $T_j^{(2)}$ and taking values in $\{\frac{k}{2^n}, k\in\mathbb{N}\}$. We get from Lemma 11 that

$$\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+T_j^{(2)}|\mathcal{G}_{T_j^{(2)}}) = \lim_{n\to+\infty}\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+T_j^{(2)}|\mathcal{G}_{T_{n,j}^{(2)}}). \quad (48)$$

Letting $\epsilon_n = \mathbb{P}_{\phi_1\times\phi_2}(t+T_j^{(2)} < T_k^{(1)} \leq t+T_{n,j}^{(2)}|\mathcal{G}_{T_{n,j}^{(2)}})$, we have

$$\begin{aligned}\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+T_j^{(2)}|\mathcal{G}_{T_{n,j}^{(2)}}) &= \mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+T_{n,j}^{(2)}|\mathcal{G}_{T_{n,j}^{(2)}}) + \epsilon_n\\ &= \sum_{l\geq 0}\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t+\frac{l}{2^n}|\mathcal{G}_{\frac{l}{2^n}})\mathbf{1}_{T_{n,j}^{(2)}=\frac{l}{2^n}} + \epsilon_n\\ &= \sum_{l\geq 0}\mathbb{P}_{\phi_1}(T_k^{(1)} > t+\frac{l}{2^n}|\mathcal{F}_{\frac{l}{2^n}}^{(1)})\mathbf{1}_{T_{n,j}^{(2)}=\frac{l}{2^n}} + \epsilon_n. \quad (49)\end{aligned}$$

Here we have used the fact that the event $B = \{T_k^{(1)} > t+\frac{l}{2^n}\}$ being independent of $\mathcal{F}_{\frac{l}{2^n}}^{(2)}$, and $\mathcal{F}_{\frac{l}{2^n}}^{(2)}$ being independent of $\mathcal{F}_{\frac{l}{2^n}}^{(1)}$, $\mathbb{P}_{\phi_1\times\phi_2}(B|\mathcal{G}_{\frac{l}{2^n}}) = \mathbb{P}_{\phi_1\times\phi_2}(B|\mathcal{F}_{\frac{l}{2^n}}^{(1)})$. Now, on the event $\{\tau = T_j^{(2)}\}\cap\{T_{k-1}^{(1)} \leq \tau < T_k^{(1)}\}$, since $l/2^n > T_{k-1}^{(1)}$ and $X_{T_{k-1}^{(1)}}^{(1)} = \varnothing$,

$$\mathbb{P}_{\phi_1}(T_k^{(1)} > t+\frac{l}{2^n}|\mathcal{F}_{\frac{l}{2^n}}^{(1)}) = \mathbb{P}_{\phi_1}(T_k^{(1)} - T_{k-1}^{(1)} > t+\frac{l}{2^n} - T_{k-1}^{(1)}|\mathcal{F}_{\frac{l}{2^n}}^{(1)}) = \mathbb{P}(T_1^{(1)} > t|X_0^{(1)} = \varnothing),$$

where the last equality comes from the fact that given $X_{T_{k-1}^{(1)}}^{(1)} = \varnothing$, the distribution of $T_k^{(1)} - T_{k-1}^{(1)}$ is an exponential distribution with rate $\alpha(\varnothing)$ having the memoryless property. We get then

$$\lim_{n\to+\infty} P_{\phi_1}(T_k^{(1)} > t+\frac{l}{2^n}|\mathcal{F}_{\frac{l}{2^n}}^{(1)})\mathbf{1}_{\tau=T_j^{(2)}}\mathbf{1}_{T_{k-1}^{(1)}\leq\tau<T_k^{(1)}} = \mathbb{P}(T_1^{(1)} > t|X_0^{(1)} = \varnothing)\mathbf{1}_{\tau=T_j^{(2)}}\mathbf{1}_{T_{k-1}^{(1)}\leq\tau<T_k^{(1)}}.$$

On the other hand, $\mathbb{E}[\varepsilon_n] = \mathbb{P}_{\phi_1\times\phi_2}(t+T_j^{(2)} < T_k^{(1)} \leq t+T_{n,j}^{(2)})$ tends to 0 by the dominated convergence theorem, so there exists a subsequent $\varphi(n)$ such that $\varepsilon_{\varphi(n)} \to 0$ almost surely. From (48), taking the limit in (49) for this subsequence, we deduce that

$$\mathbb{P}_{\phi_1\times\phi_2}(T_k^{(1)} > t + T_j^{(2)} | \mathcal{G}_{T_j^{(2)}}) \mathbf{1}_{\tau = T_j^{(2)}} \mathbf{1}_{T_{k-1}^{(1)} \leq \tau < T_k^{(1)}} = \mathbb{P}(T_{(1)}^1 > t | X_0^{(1)} = \varnothing) \mathbf{1}_{\tau = T_j^{(2)}} \mathbf{1}_{T_{k-1}^{(1)} \leq \tau < T_k^{(1)}}$$

which in view of (47) gives

$$\mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t + \tau | \mathcal{G}_\tau) \mathbf{1}_{\tau = T_j^{(2)}} = \mathbb{P}(T_1^{(1)} > t | X_0^{(1)} = \varnothing) \mathbf{1}_{\tau = T_j^{(2)}}.$$

Combining (45), (46) and this last result concludes the proof. □

Lemma 13. *For every bounded measurable function g and every $t \geq 0$,*

$$\mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) | \mathcal{G}_\tau] = \mathbb{E}[g(X_t^{(1)}) | X_0^{(1)} = \varnothing]$$
$$\mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(2)}) | \mathcal{G}_\tau] = \mathbb{E}[g(X_t^{(2)}) | X_0^{(2)} = \varnothing],$$

thereby

$$\mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) | \mathcal{G}_\tau] = \mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(2)}) | \mathcal{G}_\tau].$$

In particular, for every $A \in \mathcal{E}$ and every $t \geq 0$,

$$\mathbb{P}_{\phi_1\times\phi_2}(X_t^{(1)} \in A, \tau \leq t) = \mathbb{P}_{\phi_1\times\phi_2}(X_t^{(2)} \in A, \tau \leq t).$$

Proof. Let g be a bounded measurable function and $t \geq 0$. Denote by P_t the transition function of $(X_t)_{t\geq 0}$, i.e. for any $s, t \geq 0$, $P_t g(X_s) = \mathbb{E}(g(X_{s+t}) | \mathcal{F}_s)$. Note that both processes $X^{(1)}$ and $X^{(2)}$ have the same transition function P_t. We only prove the first equality of the lemma since the rest can be verified similarly or is a straightforward consequence.

$$\begin{aligned}
\mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) | \mathcal{G}_\tau] &= \mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) \mathbf{1}_{t+\tau < T_+^{(1)}} | \mathcal{G}_\tau] + \mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) \mathbf{1}_{t+\tau \geq T_+^{(1)}} | \mathcal{G}_\tau] \\
&= g(\varnothing) \mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t + \tau | \mathcal{G}_\tau) + \mathbb{E}_{\phi_1\times\phi_2}[P_{t+\tau-T_+^{(1)}} g(X_{T_+^{(1)}}^{(1)}) \mathbf{1}_{t+\tau \geq T_+^{(1)}} | \mathcal{G}_\tau] \\
&= g(\varnothing) \mathbb{P}_{\phi_1\times\phi_2}(T_+^{(1)} > t + \tau | \mathcal{G}_\tau) + \int_{z\in E_1} \mathbb{E}_{\phi_1\times\phi_2}[P_{t+\tau-T_+^{(1)}} g(z) \mathbf{1}_{t+\tau \geq T_+^{(1)}} | \mathcal{G}_\tau] K(\varnothing, dz).
\end{aligned}$$

We get from Lemma 12 that

$$\mathbb{E}_{\phi_1\times\phi_2}[g(X_{t+\tau}^{(1)}) | \mathcal{G}_\tau] = g(\varnothing) \mathbb{P}(T_1^{(1)} > t | X_0^{(1)} = \varnothing) + \int_{z\in E_1} \mathbb{E}[P_{t-T_1^{(1)}} g(z) \mathbf{1}_{t \geq T_1^{(1)}} | X_0^{(1)} = \varnothing] K(\varnothing, dz).$$

Notice to conclude that

$$\begin{aligned}
\mathbb{E}[g(X_t^{(1)}) | X_0^{(1)} = \varnothing] &= \mathbb{E}[g(X_t^{(1)}) \mathbf{1}_{t < T_1^{(1)}} | X_0^{(1)} = \varnothing] + \mathbb{E}[g(X_t^{(1)}) \mathbf{1}_{t \geq T_1^{(1)}} | X_0^{(1)} = \varnothing] \\
&= g(\varnothing) \mathbb{P}(T_1^{(1)} > t | X_0^{(1)} = \varnothing) + \mathbb{E}[P_{t-T_1^{(1)}} g(X_{T_1}^{(1)}) \mathbf{1}_{t \geq T_1^{(1)}} | X_0^{(1)} = \varnothing] \\
&= g(\varnothing) \mathbb{P}(T_1^{(1)} > t | X_0^{(1)} = \varnothing) + \int_{z\in E_1} \mathbb{E}[P_{t-T_1^{(1)}} g(z) \mathbf{1}_{t \geq T_1^{(1)}} | X_0^{(1)} = \varnothing] K(\varnothing, dz).
\end{aligned}$$

□

Acknowledgements

We thank Boulanger et al. (2014) for making the Langerin and Rab11 sequences data analysed in Section 4.2 available. We are grateful to Cesar Augusto Valades-Cruz for assistance in tracking the proteins from these sequences. We also thank Thierry Pécot and Charles Kervrann for helpful discussion and insight on the biological context.

References

Bass, R. F. (2011). *Stochastic processes*, Volume 33. Cambridge University Press.

Bolte, S. and F. Cordelieres (2006). A guided tour into subcellular colocalization analysis in light microscopy. *J Microscopy 224*, 213–232.

Boulanger, J., C. Gueudry, D. Munch, B. Cinquin, P. Paul-Gilloteaux, S. Bardin, C. Guérin, F. Senger, L. Blanchoin, and J. Salamero (2014). Fast high-resolution 3D total internal reflection fluorescence microscopy by incidence angle scanning and azimuthal averaging. *Proc Natl Acad Sci USA 111*(48), 17164–17169.

Brémaud, P. (1981). *Point processes and queues: martingale dynamics*. Springer.

Briane, V., M. Vimond, and C. Kervrann (2019). An overview of diffusion models for intracellular dynamics analysis. *Briefings in Bioinformatics*. bbz052.

Comas, C. (2009). Modelling forest regeneration strategies through the development of a spatio-temporal growth interaction model. *Stochastic Environmental Research and Risk Assessment 23*(8), 1089–1102.

Comas, C. and J. Mateu (2008). Space–time dependence dynamics for birth–death point processes. *Statistics & probability letters 78*(16), 2715–2719.

Costes, S., D. Daelemans, E. Cho, Z. Dobbin, G. Pavlakis, and S. Lockett (2004). Automatic and quantitative measurement of protein-protein colocalization in live cells. *Biophysical J 86*(6), 3993–4003.

Feller, W. (1939). Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheoretica 5*(1), 11–40.

Gidon, A., S. Bardin, B. Cinquin, J. Boulanger, F. Waharte, L. Heliot, H. de la Salle, D. Hanau, C. Kervrann, B. Goud, and J. Salamero (2012). A Rab11A/Myosin Vb/Rab11-FIP2 complex frames two late recycling steps of Langerin from the ERC to the Plasma membrane. *Traffic 13*(6), 815–833.

Häbel, H., M. Myllymäki, and A. Pommerening (2019). New insights on the behaviour of alternative types of individual-based tree models for natural forests. *Ecological modelling 406*, 23–32.

Jaqaman, K., D. Loerke, M. Mettlen, H. Kuwata, S. Grinstein, S. L. Schmid, and G. Danuser (2008). Robust single-particle tracking in live-cell time-lapse sequences. *Nature methods 5*(8), 695.

Kendall, D. G. (1949). Stochastic processes and population growth. *Journal of the Royal Statistical Society. Series B (Methodological) 11*(2), 230–282.

Lagache, T., N. Sauvonnet, L. Danglot, and J.-C. Olivo-Marin (2015). Statistical analysis of molecule colocalization in bioimaging. *Cytometry Part A 87*(6), 568–579.

Lavancier, F., T. Pécot, L. Zengzhen, and C. Kervrann (2019). Testing independence between two random sets for the analysis of colocalization in bioimaging. *to appear in Biometrics*.

Le Guével, R. (2019). Exponential inequalities for the supremum of some counting processes and their square martingales. *submitted (preprint hal-02275583 available online)*.

Loader, C. (2006). *Local regression and likelihood*. Springer Science & Business Media.

Lotwick, H. and B. Silverman (1981). Convergence of spatial birth-and-death processes. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 90, pp. 155–165. Cambridge University Press.

Mateu, J., F. P. Schoenberg, D. M. Diez, J. A. González, and W. Lu (2015). On measures of dissimilarity between point patterns: Classification based on prototypes and multidimensional scaling. *Biometrical Journal 57*(2), 340–358.

Møller, J. (1989). On the rate of convergence of spatial birth-and-death processes. *Annals of the Institute of Statistical Mathematics 41*(3), 565–581.

Møller, J. and M. Sørensen (1994). Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. *Scandinavian journal of statistics 21*(1), 1–19.

Møller, J. and R. P. Waagepetersen (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC, Boca Raton.

Pécot, T., P. Bouthemy, J. Boulanger, A. Chessel, S. Bardin, J. Salamero, and C. Kervrann (2014). Background fluorescence estimation and vesicle segmentation in live cell imaging with conditional random fields. *IEEE Transactions on Image Processing 24*(2), 667–680.

Pécot, T., C. Kervrann, S. Bardin, B. Goud, and J. Salamero (2008). Patch-based markov models for event detection in fluorescence bioimaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 95–103. Springer.

Pécot, T., L. Zengzhen, J. Boulanger, J. Salamero, and C. Kervrann (2018). A quantitative approach for analyzing the spatio-temporal distribution of 3d intracellular events in fluorescence microscopy. *eLife 7*, e32311.

Pommerening, A. and P. Grabarnik (2019). *Individual-Based Methods in Forest Ecology and Management*. Springer.

Preston, C. (1975). Spatial birth and death processes. *Advances in applied probability 7*(3), 371–391.

Renshaw, E., C. Comas, and J. Mateu (2009). Analysis of forest thinning strategies through the development of space–time growth–interaction simulation models. *Stochastic Environmental Research and Risk Assessment 23*(3), 275–288.

Renshaw, E. and A. Särkkä (2001). Gibbs point processes for studying the development of spatial-temporal stochastic processes. *Computational statistics & data analysis 36*(1), 85–105.

Reynolds, J. F. (1973). On estimating the parameters of a birth-death process. *Australian Journal of Statistics 15*(1), 35–43.

Sadahiro, Y. (2019). Analysis of the appearance and disappearance of point objects over time. *International Journal of Geographical Information Science 33*(2), 215–239.

Särkkä, A. and E. Renshaw (2006). The analysis of marked point patterns evolving through space and time. *Computational Statistics & Data Analysis 51*(3), 1698–1718.

Schuhmacher, D. and A. Xia (2008). A new metric between distributions of point processes. *Advances in applied probability 40*(3), 651–672.

Thorisson, H. (2000). *Coupling, Stationarity, and Regeneration*. Springer, New York.

Wang, Y. and S.-C. Zhu (2002). A generative method for textured motion: Analysis and synthesis. In *European Conference on Computer Vision*, pp. 583–598. Springer.

Supporting information

The codes and data used in Section 4 are available in the GitHub repository at `https://github.com/lavancier-f/Birth-Death-Move-process`.