



# Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes

Romain Azaïs, Aurélie Muller-Gueudin

## ► To cite this version:

Romain Azaïs, Aurélie Muller-Gueudin. Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes. 2015. hal-01168651v1

**HAL Id: hal-01168651**

**<https://hal.science/hal-01168651v1>**

Preprint submitted on 26 Jun 2015 (v1), last revised 15 Nov 2016 (v2)

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

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

# Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes

Romain Azaïs\* and Aurélie Muller-Gueudin†

*Inria Nancy-Grand Est, Team BIGS, Institut Élie Cartan de Lorraine, Vandoeuvre-lès-Nancy, France*

## Abstract

A piecewise-deterministic Markov process is a stochastic process whose behavior is governed by an ordinary differential equation punctuated by random jumps occurring at random times. We focus on the nonparametric estimation problem of the jump rate for such a stochastic model observed within a long time interval under an ergodicity condition. We introduce an uncountable class (indexed by the deterministic flow) of recursive kernel estimates of the jump rate and we establish their strong pointwise consistency as well as their asymptotic normality. We propose to choose among this class the estimator with the minimal variance, which is unfortunately unknown and thus remains to be estimated. We also discuss the choice of the bandwidth parameters by cross-validation methods.

**Keywords:** Cross-validation · Jump rate · Kernel method · Nonparametric estimation · Piecewise-deterministic Markov process

**Mathematics Subject Classification (2010):** 62M05 · 62G20 · 60J25

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Problem formulation</b>	<b>3</b>
2.1	Definition and notation	3
2.2	Assumptions	5
<b>3</b>	<b>Estimation procedure</b>	<b>6</b>
3.1	Inference for the inter-jumping times	6
3.2	Optimal estimation of the jump rate	9
3.3	How to choose bandwidth parameters $\alpha$ and $\beta$ ?	10
<b>4</b>	<b>Numerical illustration</b>	<b>12</b>
4.1	Presentation of the model	13
4.2	Estimation procedure	13
4.3	Simulation results	14
<b>A</b>	<b>Ergodicity and invariant measures</b>	<b>17</b>
<b>B</b>	<b>Proof of Theorem 3.3</b>	<b>17</b>
B.1	Sketch of the proof	18
B.2	Remainder term	18
B.3	Martingale term	20
<b>C</b>	<b>Proof of Proposition 3.14</b>	<b>24</b>

---

\*Electronic address: [romain.azais@inria.fr](mailto:romain.azais@inria.fr); Corresponding author

†Electronic address: [aurelie.gueudin@univ-lorraine.fr](mailto:aurelie.gueudin@univ-lorraine.fr)

# 1 Introduction

Piecewise-deterministic Markov processes (PDMP's in abbreviated form) have been introduced in the literature by Davis in [15] as a general class of continuous-time non-diffusion stochastic models, suitable for modeling deterministic phenomena in which the randomness appears as point events. The motion of a PDMP may be defined from three local characteristics: the flow  $\Phi(x, t)$ , the jump rate  $\lambda(x)$  and the transition measure  $\mathcal{Q}(x, dy)$ . Starting from some initial value  $X_0$ , the process evolves in a deterministic way following  $\Phi(X_0, t)$  until the first jump time  $T_1$  which occurs either when the flow reaches the boundary of the state space or in a Poisson-like fashion with non homogenous rate  $\lambda(\Phi(X_0, t))$  before. In both cases, the location of the process at time  $T_1$  is governed by the transition distribution  $\mathcal{Q}(\Phi(X_0, T_1), dy)$  and the motion restarts from this new point as before. This family of stochastic models is well-adapted for tackling various problems arising for example in biology [8, 14, 28, 29], in neuroscience [20] or in reliability [12, 16]. Proposing efficient statistical methods for this class of stochastic models is therefore of a great interest. Nevertheless, the very particular framework involving both deterministic motion and punctual random jumps imposes to consider specific methods. For instance, the authors of [5] have shown that the well-known multiplicative intensity model developed by Aalen [1] for estimating the jump rate function does not directly apply to PDMP's. Alternative approaches should be thus proposed.

In the present paper, we focus on the recursive nonparametric estimation of the jump rate of a PDMP from the observation of only one trajectory within a long time interval. More precisely, the purpose of this work is to show how one may obtain by kernel methods a class of consistent estimators for the jump rate, and how one may choose among this class in an optimal way. To the best of our knowledge, the nonparametric estimation of the jump rate in a general framework has never been investigated.

As PDMP's may model a large variety of problems, some methods have been developed by many authors for their statistical inference. As presented before, the randomness of a PDMP is governed by two characteristics: the transition kernel  $\mathcal{Q}(x, dy)$  and the jump rate  $\lambda(x)$ . As a consequence, two main questions arise in the estimation problem for such a process, namely the statistical inference for both these features. On the one hand, a few papers investigate some nonparametric methods for estimating the transition function of a PDMP either for a specific model [12] or in a more general setting for a  $d$ -dimensional process [3]. On the other hand the estimation of the jump rate  $\lambda(x)$  or of the associated density function has been more extensively studied by several authors. Without attempting to give an exhaustive survey of the literature on this topic, one may refer the reader to [5, 17, 18, 23, 26] and the references therein. In the book [23], the author studies likelihood processes for observation of PDMP's which could lead to inference methods in a parametric or semi-parametric setting. The papers [17, 18] deal with the nonparametric estimation for some PDMP's used in the modeling of a size-structured population observed along a lineage tree. In both these articles, the authors rely on the specific form of the features of the process of interest in order to derive the asymptotic behavior of their estimation procedure. These techniques have been generalized in [26] to introduce a nonparametric method for estimating the jump rate in a specific class of one-dimensional PDMP's with monotonic motion and deterministic breaks, that is to say when the transition measure  $\mathcal{Q}(x, dy)$  is a Dirac mass at some location depending on  $x$ . The procedures developed in these papers [17, 18, 26] are obviously of a great interest but strongly use the particular framework involved in the investigated models and are thus not well adapted in a more general setting. In [5], the authors show that the famous multiplicative intensity model only applies for estimating the jump rate of a modified version of the underlying PDMP. This leads to a statistical method for approximating the conditional density associated with the jump rate for a process defined on a bounded metric state space.

A main difficulty throughout the present paper and the articles [3, 5] is related to the presence of deterministic jumps when the path reaches the boundary of the state space. This feature is often used for modeling a deterministic switching when the quantitative variable rises over a certain threshold [2, 22]. In a statistical point of view, the interarrival times are therefore right-censored by a deterministic clock depending on the state space, which leads to some technical difficulties. We would like to emphasize that the techniques developed in the references [17, 18, 23, 26] do not take into account the likely presence of forced jumps in the dynamic.

One may also find in the literature a few papers [6, 7, 9, 10] which focus on the estimation of various functionals for this family of stochastic models. More precisely, the authors of [9, 10] provide numerical methods for the expectations and for the exit times of PDMP's. In addition, the article [7] deals with a

PDMP introduced for modeling the temporal evolution of exposure to a food contaminant and consider a simulation-based statistical inference procedure for estimating some functionals such as first passage times. Plug-in methods have been studied in [6] for a non-ergodic growth-fragmentation model which is absorbed under a certain threshold. In many aspects our approach and these papers are different and complementary. Indeed they are devoted to the estimation of some functionals of PDMP's while we focus on the direct estimation of the primitive data of such a process.

In this article, we introduce a three-dimensional kernel estimator computed from the observation of the embedded Markov chain of a PDMP. We establish its pointwise consistency as well as its asymptotic normality in Theorem 3.3. The estimate that we consider is recursive: it may be computed in real-time from sequential data, which may be relevant in many applications. We deduce from this result two first corollaries about the nonparametric estimation of the conditional density of the interarrival times  $f(x, t)$  (see Corollary 3.8) and their survival function  $G(x, t)$  (see Corollary 3.9). We also investigate in Corollary 3.10 the asymptotic behavior of an estimator for the composed function  $\lambda \circ \Phi(x, t)$ . We derive in (18) an uncountable class (indexed by the reverse flow  $\Phi(x, -t)$ ) of consistent estimates of the jump rate  $\lambda(x)$ , and we show how one may choose among this class by minimizing their asymptotic variance. We state in (19) that this procedure is equivalent to maximize the criterion  $\nu_\infty(\xi)G(\xi, \tau_x(\xi))$  along the curve  $\Phi(x, -t)$ , where  $\nu_\infty(\xi)$  denotes the invariant measure of the post-jump locations and  $\tau_x(\xi)$  is the deterministic time to reach  $x$  following  $\Phi(\xi, t)$ . The choice of this criterion is far to be obvious without precisely computing the limit variance in the central limit theorem presented in Corollary 3.10. Indeed, a naive criterion to maximize is the invariant distribution  $\nu_\infty(\xi)$  along  $\Phi(x, -t)$ : the larger  $\nu_\infty(\xi)$  is, the larger the number of data around  $\xi$  is and the higher the quality of the estimation should be. Nonetheless, this simple criterion does not take into account that the estimate also depends on the time of interest  $\tau_x(\xi)$  (see Remark 3.11). In Section 4, we show on synthetic data that our method performs pretty well, and that the choice of the criterion  $\nu_\infty(\xi)G(\xi, \tau_x(\xi))$  is better than the naive one  $\nu_\infty(\xi)$ . We also discuss the choice of the bandwidth parameters by minimizing the Integrated Square Error computed along the reverse flow  $\Phi(x, -t)$ : we introduce a cross-validation procedure in this Markov setting and we prove its convergence in Propositions 3.14 and 3.15. Finally, we would like to highlight that the regularity conditions that we impose are non restrictive. In particular, neither the deterministic exit time from the state space is assumed to be a bounded function, nor the transition kernel is supposed to be lower-bounded, as is the case in [5] (see eq. (2) and Assumptions 2.4). In addition, the forms of the transition measure and of the deterministic flow are not specified.

The sequel of the paper is organized as follows. We begin in Section 2 with the precise formulation of our framework (see Subsection 2.1) and the main assumptions that we need in this article (see Subsection 2.2). Section 3 is devoted to the presentation of the statistical procedure and the related results of convergence. More precisely, a three-dimensional kernel estimator for the inter-jumping times is introduced and investigated in Subsection 3.1. We derive a class of estimators of the jump rate and propose how to choose among it in Subsection 3.2. The crucial choice of the bandwidth parameters is studied in Subsection 3.3. Finally, the whole estimation procedure is illustrated by a numerical example in Section 4. The proofs and the technicalities are postponed in Appendix A, B and C at the end of the paper.

## 2 Problem formulation

This section is devoted to the definition of the PDMP's and to the presentation of the main assumptions that we impose in the paper.

### 2.1 Definition and notation

In this paper  $\mathcal{B}(\mathbb{R}^d)$  denotes the Borel algebra of  $\mathbb{R}^d$  endowed with the Euclidean norm  $|\cdot|$ . In addition, the Lebesgue measure on  $\mathcal{B}(\mathbb{R}^d)$  is denoted by  $\lambda_d(dx)$ , with the particular notation in the one-dimensional case  $\lambda_1(dx) = dx$ . The ball of  $\mathbb{R}^d$  with radius  $r$  and center  $x$  is denoted by  $B_d(x, r)$ . The closure of a set  $E$  is denoted by  $\overline{E}$  while  $\partial E$  stands for its boundary.

The motion of a PDMP on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  may always be defined from three local characteristics  $(\lambda, \mathcal{Q}, \Phi)$ :

- $\Phi : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$  is the deterministic flow. It satisfies the semigroup property,

$$\forall x \in \mathbb{R}^d, \forall t, s \in \mathbb{R}, \Phi(x, t+s) = \Phi(\Phi(x, t), s).$$

- $\lambda : \mathbb{R}^d \rightarrow \mathbb{R}_+$  is the jump rate.
- $\mathcal{Q} : (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)) \rightarrow [0, 1]$  is the transition kernel.

We define the deterministic exit times of  $E$  for the flow and for the reverse flow as, for any  $x \in E$ ,

$$t^+(x) = \inf \{t > 0 : \Phi(x, t) \in \partial E\} \quad \text{and} \quad t^-(x) = \inf \{t > 0 : \Phi(x, -t) \in \partial E\}. \quad (1)$$

In all the sequel, we consider a PDMP  $(X_t)_{t \geq 0}$  evolving on an open subset  $E$  of  $\mathbb{R}^d$ . In this context, we impose as usual [15, (24.8) Standard conditions] that,

$$\forall x \in E, \exists \varepsilon > 0, \int_0^\varepsilon \lambda(\Phi(x, t)) dt < +\infty,$$

and

$$\forall x \in \overline{E}, \mathcal{Q}(x, E \setminus \{x\}) = 1. \quad (2)$$

In addition, we restrict ourselves to the case where the transition kernel  $\mathcal{Q}$  admits a density with respect to the Lebesgue measure,

$$\forall x \in \mathbb{R}^d, \forall A \in \mathcal{B}(\mathbb{R}^d), \mathcal{Q}(x, A) = \int_A \mathcal{Q}(x, y) \lambda_d(dy). \quad (3)$$

This assumption is natural when one considers multivariate real-valued PDMP's, and is satisfied in various problems arising in biology [8], population dynamics [6] or in insurance [15, (21.11) An insurance model].

Starting from any initial condition  $X_0 = x$ , the motion of  $(X_t)_{t \geq 0}$  may be described as follows. The distribution of the first jump time  $T_1$  is given by,

$$\forall t \geq 0, \mathbb{P}(T_1 > t | X_0 = x) = \begin{cases} \exp\left(-\int_0^t \lambda(\Phi(x, s)) ds\right) & \text{if } t < t^+(x), \\ 0 & \text{else.} \end{cases} \quad (4)$$

In other words, the process jumps either when the flow hits the boundary of the state space at time  $t^+(x)$  or in a Poisson-like fashion with rate  $\lambda \circ \Phi$  before. Next the post-jump location  $Z_1$  at time  $T_1$  is defined through the transition kernel  $\mathcal{Q}$ : for any test function  $\varphi$ , we have

$$\mathbb{E}[\varphi(Z_1) | T_1, X_0 = x] = \int \varphi(u) \mathcal{Q}(\Phi(x, T_1), du).$$

The path between 0 and the first jump time  $T_1$  is given by,

$$\forall 0 \leq t \leq T_1, X_t = \begin{cases} \Phi(x, t) & \text{if } t < T_1, \\ Z_1 & \text{else.} \end{cases}$$

Now starting from the post-jump location  $X_{T_1}$ , one chooses the next inter-jumping time  $S_2 = T_2 - T_1$  and the future post-jump location  $Z_2$  in a similar way as before, and so on. One obtains a strong Markov process with  $(T_n)_{n \geq 0}$  as the sequence of the jump times (where  $T_0 = 0$  by convention). The inter-jumping times are defined by  $S_0 = 0$  and, for any integer  $n \geq 1$ ,  $S_n = T_n - T_{n-1}$ . Finally  $(Z_n)_{n \geq 0}$  denotes the stochastic sequence of the post-jump locations of  $(X_t)_{t \geq 0}$ , with for any  $n$ ,  $Z_n = X_{T_n}$ .

All the randomness of the PDMP  $(X_t)_{t \geq 0}$  is contained in the stochastic sequence  $(Z_n, S_{n+1})_{n \geq 0}$  which is a Markov chain. In addition, the post-jump locations  $(Z_n)_{n \geq 0}$  also form a discrete-time Markov process on the state space  $E$  because of the condition (2). In this paper,  $\nu_n$  denotes the distribution of the  $n^{\text{th}}$  post-jump location  $Z_n$  for any integer  $n \geq 0$ , while  $\mathcal{P}$  denotes its Markov kernel,

$$\begin{aligned} \forall x \in E, \forall A \in \mathcal{B}(\mathbb{R}^d), \mathcal{P}(x, A) &= \mathbb{P}(Z_{n+1} \in A | Z_n = x) \\ &= \int_{\mathbb{R}_+} \mathcal{S}(x, dt) \mathcal{Q}(\Phi(x, t), A), \end{aligned} \quad (5)$$

where  $\mathcal{S}$  stands for the conditional distribution of  $S_{n+1}$  given  $Z_n$  for any  $n$ ,

$$\begin{aligned} \forall x \in E, \forall t \geq 0, \mathcal{S}(x, (t, +\infty)) &= \mathbb{P}(S_{n+1} > t \mid Z_n = x) \\ &= \begin{cases} \exp\left(-\int_0^t \lambda(\Phi(x, s))ds\right) & \text{if } t < t^+(x), \\ 0 & \text{else,} \end{cases} \end{aligned} \quad (6)$$

in light of (4). We would like to highlight that the conditional distribution  $\mathcal{S}(x, \cdot)$  is absolutely continuous with respect to the unidimensional Lebesgue measure on  $(0, t^+(x))$  with sometimes a singular component at  $t^+(x)$ ,

$$\forall A \in \mathcal{B}(\mathbb{R}_+), \mathcal{S}(x, A) = \int_{A \cap (0, t^+(x))} f(x, t) dt + \mathcal{S}(x, A \cap \{t^+(x)\}), \quad (7)$$

where the conditional density  $f$  may be obtained by deriving (6),

$$\forall x \in E, \forall 0 < t < t^+(x), f(x, t) = \lambda(\Phi(x, t)) \exp\left(-\int_0^t \lambda(\Phi(x, s))ds\right). \quad (8)$$

In all the sequel  $G$  stands for the conditional survival function associated with  $f$ , that is,

$$\forall x \in E, \forall 0 < t < t^+(x), G(x, t) = \mathcal{S}(x, (t, +\infty)), \quad (9)$$

where  $\mathcal{S}$  is defined in (6). As highlighted before, the process  $(Z_n, S_{n+1})_{n \geq 0}$  forms a Markov chain on the set  $F$  defined by

$$F = \bigcup_{x \in E} \{x\} \times [0, t^+(x)]. \quad (10)$$

$\mathcal{R}$  denotes the transition kernel of this process,

$$\begin{aligned} \forall (x, t) \in F, \forall A \times B \in \mathcal{B}(\mathbb{R}^d \times \mathbb{R}_+), \mathcal{R}((x, t), A \times B) &= \mathbb{P}(Z_{n+1} \in A, S_{n+2} \in B \mid Z_n = x, S_{n+1} = t) \\ &= \int_A \mathcal{Q}(\Phi(x, t), d\xi) \mathcal{S}(\xi, B), \end{aligned} \quad (11)$$

and, for any  $n$ ,  $\mu_n$  denotes the distribution of the couple  $(Z_n, S_{n+1})$ .

## 2.2 Assumptions

The main assumption that we impose in the present paper is a condition of ergodicity on the Markov chain  $(Z_n)_{n \geq 0}$ . This property is often a keystone in statistical inference for Markov processes and may be directly imposed [3, 5] or established [26] from the primitive features of the data.

**Assumption 2.1** *There exists a distribution  $\nu_\infty$  on  $E$  such that, for any initial distribution  $\nu_0 = \delta_{\{x\}}$ ,  $x \in E$ ,*

$$\lim_{n \rightarrow +\infty} \|\nu_n - \nu_\infty\|_{TV} = 0,$$

where  $\|\cdot\|_{TV}$  stands for the total variation norm.

This assumption may be checked directly on the Markov kernel  $\mathcal{P}$  of  $(Z_n)_{n \geq 0}$  from the existence of a Foster-Lyapunov's function or Doeblin's condition for instance [27, Theorem 16.0.2]. In the following remark we establish a first property of the sequence  $(\nu_n)_{n \geq 0}$  and of its limit  $\nu_\infty$ .

**Remark 2.2** *Since the transition kernel  $\mathcal{Q}$  is assumed to be absolutely continuous with respect to the Lebesgue measure (3), the kernel  $\mathcal{P}$  given by (5) of the post-jump locations  $(Z_n)_{n \geq 0}$  also admits a density. As a consequence, for any integer  $n$ , the distribution  $\nu_n$  of  $Z_n$  and thus the invariant measure  $\nu_\infty$  introduced in Assumption 2.1 admit a density on the state space  $E$ . For the sake of clarity, we write  $\nu_\infty(dx) = \nu_\infty(x)\lambda_d(dx)$  with a slight abuse of notation.*

We add some regularity conditions on the main features of the process to show the convergence of the estimates in Theorem 3.3.

### Assumptions 2.3

- *The sup-norms  $\|\mathcal{Q}\|_\infty$  and  $\|f\|_\infty$  are finite. These conditions are used in the proof of Theorem 3.3 to find an upper bound of the non diagonal terms of some square variation process of interest.*

- The functions  $\mathcal{Q}$  and  $f$  are Lipschitz,

$$\begin{aligned} \forall x, y \in E, \forall 0 < s < t < t^+(x) \wedge t^+(y), |f(x, t) - f(y, s)| &\leq [f]_{Lip}(|t - s| + |y - x|), \\ \forall x \in \bar{E}, \forall y, z \in E, |\mathcal{Q}(x, y) - \mathcal{Q}(x, z)| &\leq [\mathcal{Q}]_{Lip}|y - z|. \end{aligned}$$

These conditions are used in the proof of Theorem 3.3 to control the diagonal terms of the same variation process and to study the convergence of some remainder terms.

- The survival function  $G$  is Lipschitz,

$$\forall x, y \in E, \forall 0 < t < t^+(x) \wedge t^+(y), |G(x, t) - G(y, t)| \leq [G]_{Lip}|x - y|.$$

This condition is used in the proof of Theorem 3.3 to investigate the convergence of the remainder terms.

- The deterministic exit time  $t^+$  is continuous. This condition is used to find some admissible initial bandwidths  $v_0$  and  $w_0$ .

Finally, we consider an additional condition on both the transition kernels  $\mathcal{P}$  and  $\mathcal{Q}$  and the flow  $\Phi$  in order to ensure the Lipschitz mixing property of the Markov chain  $(Z_n)_{n \geq 0}$ . This will be sufficient to establish the almost sure convergence to 0 of the remainder term with the adequate rate in the proof of Theorem 3.3.

**Assumptions 2.4** The transition kernel  $\mathcal{P}$  of the Markov chain  $(Z_n)_{n \geq 0}$  satisfies, for some  $a_1 \geq 1$  and  $a_2 < 1$ ,

$$\forall (x, y) \in E^2, \int_{\mathbb{R}^d \times \mathbb{R}^d} |u - v|^{a_1} \mathcal{P}(x, du) \mathcal{P}(y, dv) \leq a_2 |x - y|^{a_1}.$$

In addition, the composed function  $\mathcal{Q}(\Phi(\cdot, \cdot), \cdot)$  belongs to the regularity class  $\text{Li}(r_1, r_2)$  defined by

$$|\mathcal{Q}(\Phi(x_1, t_1), y_1) - \mathcal{Q}(\Phi(x_2, t_2), y_2)| = O(|(x_1, t_1, y_1) - (x_2, t_2, y_2)|^{r_2} (|(x_1, t_1, y_1)|^{r_1} + |(x_2, t_2, y_2)|^{r_1} + 1)),$$

for some positive numbers  $r_1$  and  $r_2$  satisfying  $2(r_1 + r_2) \leq a_1$ .

## 3 Estimation procedure

### 3.1 Inference for the inter-jumping times

For any integer  $n$ , we introduce the  $\sigma(Z_0, S_1, \dots, Z_{n-1}, S_n)$ -measurable functions  $\hat{\nu}_\infty^n$  defined on  $E$ , and  $\hat{\mathcal{F}}^n$  and  $\hat{\mathcal{G}}^n$  defined on the interior  $\mathring{F}$  (let us recall that the set  $F$  is given in (10)), by,

$$\begin{aligned} \forall (x, t) \in \mathring{F}, \hat{\mathcal{F}}^n(x, t) &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d w_i} \mathbb{K}_d \left( \frac{Z_i - x}{v_i} \right) \mathbb{K}_1 \left( \frac{S_{i+1} - t}{w_i} \right), \\ \forall (x, t) \in \mathring{F}, \hat{\mathcal{G}}^n(x, t) &= \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d \left( \frac{Z_i - x}{v_i} \right) \mathbb{1}_{\{S_{i+1} > t\}}, \end{aligned} \tag{12}$$

$$\forall x \in E, \hat{\nu}_\infty^n(x) = \frac{1}{n} \sum_{i=0}^{n-1} \frac{1}{v_i^d} \mathbb{K}_d \left( \frac{Z_i - x}{v_i} \right), \tag{13}$$

where  $\mathbb{K}_p$  denotes a kernel function on  $\mathbb{R}^p$ ,  $p \in \{1, d\}$ , and the bandwidths are defined for any integer  $k$  by  $v_k = v_0(k+1)^{-\alpha}$  and  $w_k = w_0(k+1)^{-\beta}$  for some  $\alpha, \beta > 0$  and initial positive values  $v_0$  and  $w_0$ .

It should be already noted that these quantities are of a great interest in the statistical study of the inter-jumping times of the PDMP  $(X_t)_{t \geq 0}$ . Indeed, we will see that:

- The ratio  $\frac{\hat{\mathcal{F}}^n(x, t)}{\hat{\nu}_\infty^n(x)}$  estimates the conditional density  $f(x, t)$  defined in (8).
- The ratio  $\frac{\hat{\mathcal{G}}^n(x, t)}{\hat{\nu}_\infty^n(x)}$  estimates the conditional survival function  $G(x, t)$  defined in (9).
- The ratio  $\frac{\hat{\mathcal{F}}^n(x, t)}{\hat{\mathcal{G}}^n(x, t)}$  estimates the composed function  $\lambda(\Phi(x, t))$ .

In addition, as the name suggests, we will state that  $\hat{\nu}_\infty^n(x)$  is a good estimate of the density  $\nu_\infty(x)$  of the unique invariant distribution of the post-jump locations, which is relevant in the estimation problem for PDMP's but has already been investigated in [3, Proposition A.11].

In all the sequel, we impose a few assumptions on both the kernel functions  $\mathbb{K}_1$  and  $\mathbb{K}_d$ .

**Assumptions 3.1** *For any  $p \in \{1, d\}$ , the kernel function  $\mathbb{K}_p$  is assumed to be a nonnegative smooth function satisfying the following conditions:*

- *The sup-norm  $\|\mathbb{K}_p\|_\infty$  is finite.*
- $\int_{\mathbb{R}^p} \mathbb{K}_p d\lambda_p = 1$ .
- $\text{supp } \mathbb{K}_p \subset B_p(0_p, \delta)$ . *Together with well-chosen initial bandwidths, this condition avoids to compute the kernel estimator from data located at the boundary of the state space  $F$  (see also Remark 3.4).*
- *Only for  $p = d$ . The function  $\mathbb{K}_d$  is Lipschitz,*

$$\forall x, y \in \mathbb{R}^d, |\mathbb{K}_d(x) - \mathbb{K}_d(y)| \leq [\mathbb{K}]_{Lip} |x - y|.$$

*This condition is used to show eq. (55) in the proof of Proposition 3.14.*

**Remark 3.2** *In particular, Assumptions 3.1 ensure that, for any  $p \in \{1, d\}$ ,  $\tau_p^2 = \int_{\mathbb{R}^p} \mathbb{K}_p^2 d\lambda_p$  is finite. This is used to find an upper bound of the non diagonal terms of the variation process in the proof of Theorem 3.3. In addition, the integral  $\int_{\mathbb{R}^p} |u| \mathbb{K}_p(u) d\lambda_p(u)$  is also finite, which is needed to establish the almost sure convergence of the remainder terms in the proof of the same result.*

In the sequel the admissible set for the bandwidth parameters  $\alpha$  and  $\beta$  is given by

$$\mathcal{A} = \{(\alpha, \beta) \in \mathbb{R}^2 : \alpha > 0, \beta > 0, \alpha d + \beta < 1, \alpha d + \beta + 2 \min(\alpha, \beta) > 1\}.$$

In this part, our main result is obtained from the use of vector martingales and is stated in the following theorem.

**Theorem 3.3** *For any couple  $(x, t) \in \hat{F}$  such that  $\nu_\infty(x)f(x, t) > 0$ , for any  $(\alpha, \beta) \in \mathcal{A}$  and  $(v_0, w_0)$  such that*

$$t + w_0 \delta < \inf_{\xi \in B_d(x, v_0 \delta)} t^+(\xi), \quad (14)$$

*where  $\delta$  appears in the third item of Assumptions 3.1, we have the almost sure convergence,*

$$\begin{bmatrix} \hat{\mathcal{F}}^n(x, t) \\ \hat{\mathcal{G}}^n(x, t) \\ \hat{\nu}_\infty^n(x) \end{bmatrix} \xrightarrow{a.s.} \begin{bmatrix} \nu_\infty(x)f(x, t) \\ \nu_\infty(x)G(x, t) \\ \nu_\infty(x) \end{bmatrix}$$

*and the asymptotic normality,*

$$n^{\frac{1-\alpha d-\beta}{2}} \left( \begin{bmatrix} \hat{\mathcal{F}}^n(x, t) \\ \hat{\mathcal{G}}^n(x, t) \\ \hat{\nu}_\infty^n(x) \end{bmatrix} - \begin{bmatrix} \nu_\infty(x)f(x, t) \\ \nu_\infty(x)G(x, t) \\ \nu_\infty(x) \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(0_3, \Sigma(x, t, \alpha, \beta)),$$

*where the variance-covariance matrix  $\Sigma(x, t, \alpha, \beta)$  is degenerate with only one positive term at position (1, 1).  $\Sigma(x, t, \alpha, \beta)$  is defined in (50).*

**Proof.** The proof is stated in Appendix B. □

**Remark 3.4** *The existence of a couple  $(v_0, w_0)$  satisfying (14) is obvious whenever the exit time  $t^+$  is continuous (see Assumptions 2.3). This condition ensures that all the inter-jumping times used in the calculus of  $\hat{\mathcal{F}}^n$  and  $\hat{\mathcal{G}}^n$  are not obtained from forced jumps when the process reaches the boundary of the state space. In the case where  $t^+(x) = \infty$ , it is obvious that no interarrival times are right-censored. The consistency and the asymptotic normality are therefore still accurate without any condition on  $(v_0, w_0)$ .*



**Remark 3.5** In Theorem 3.3, the choice (14) of the initial bandwidths  $v_0$  and  $w_0$  is locally dependent on the point of interest. This may appear restrictive but may be avoided by considering the elements of

$$\mathcal{C} = \left\{ C \times [0, T] : C \text{ is a compact subset of } E \text{ and } T < \inf_{x \in C} t^+(x) \right\}. \quad (15)$$

Indeed, for any  $C \times [0, T] \in \mathcal{C}$ , there always exists a couple  $(v_0, w_0)$  such that

$$T + w_0 \delta < \inf_{x \in C} \inf_{\xi \in B_d(x, v_0 \delta)} t^+(\xi).$$

Thus,  $(v_0, w_0)$  satisfies (14) for any point  $(x, t) \in C \times [0, T]$ .

**Remark 3.6** The variance-covariance matrix appearing in the asymptotic normality presented in Theorem 3.3 is degenerate with only the component  $(1, 1)$  positive. It means that the rate of the estimators  $\hat{\nu}_\infty^n$  and  $\hat{\mathcal{G}}^n$  is faster than the one of  $\hat{\mathcal{F}}^n$ . This is straightforward because  $\hat{\mathcal{F}}^n$  is obtained by smoothing the empirical distribution of the data both in the spatial and temporal directions contrary to  $\hat{\nu}_\infty^n$  and  $\hat{\mathcal{G}}^n$ . The proof of the previous result may be adapted to show the two-dimensional central limit theorem, with  $\alpha$  such that  $\alpha d < 1$  and  $\alpha(d+2) > 1$ :

$$n^{\frac{1-\alpha d}{2}} \left( \begin{bmatrix} \hat{\mathcal{G}}^n(x, t) \\ \hat{\nu}_\infty^n(x) \end{bmatrix} - \begin{bmatrix} \nu_\infty(x)G(x, t) \\ \nu_\infty(x) \end{bmatrix} \right) \xrightarrow{d} \mathcal{N}(0_2, \Sigma'(x, t, \alpha)),$$

where  $\Sigma'(x, t, \alpha)$  is a diagonal  $2 \times 2$ -matrix. The keystone to state this convergence is the behavior given in (48) of the hook of a vector martingale.

If we assume the geometric ergodicity of the Markov chain  $(Z_n)_{n \geq 0}$ , one may also obtain the rate of convergence of the variances of the estimates  $\hat{\mathcal{F}}^n$ ,  $\hat{\mathcal{G}}^n$  and  $\hat{\nu}_\infty^n$  uniformly on any compact subset of  $\mathring{F}$  such that the parameters  $v_0$  and  $w_0$  may be uniformly chosen (see Remark 3.5).

**Proposition 3.7** Let us assume that there exists  $b > 1$  such that  $\|\nu_n - \nu_\infty\|_{TV} = O(b^{-n})$ . The geometric ergodicity is in particular ensured by Doeblin's condition (see [27, Theorem 16.0.2]). Then, for any set  $C \times [0, T] \in \mathcal{C}$  and for any couple  $(\alpha, \beta)$  such that  $2(\alpha d + \beta) < 1$ , we have

$$\begin{aligned} \sup_{(x, t) \in C} \text{Var}(\hat{\mathcal{F}}^n(x, t)) &= O(n^{2(\alpha d + \beta) - 1}), \\ \sup_{(x, t) \in C} \left( \text{Var}(\hat{\mathcal{G}}^n(x, t)) + \text{Var}(\hat{\nu}_\infty^n(x)) \right) &= O(n^{2\alpha d - 1}). \end{aligned}$$

Let us recall that  $\mathcal{C}$  has been defined in (15). The rate of convergence for  $\hat{\mathcal{F}}^n$  is faster than the one given in Theorem 3.3 whenever  $3(\alpha d + \beta) < 1$ .

**Proof.** The proof is similar to the demonstrations of Proposition B.5 and Corollary B.6 of [3] and relies on the control of the covariance process of functionals of a geometrically ergodic Markov chain (see [27, Theorem 16.1.5]).  $\square$

We present in the sequel some corollaries of Theorem 3.3 that are of interest in the estimation problem for the inter-jumping times. First, we define the estimator  $\hat{f}^n(x, t)$  of  $f(x, t)$  by,

$$\forall (x, t) \in \mathring{F}, \quad \hat{f}^n(x, t) = \frac{\hat{\mathcal{F}}^n(x, t)}{\hat{\nu}_\infty^n(x)},$$

with the usual convention  $0/0 = 0$ . We have the following result of convergence.

**Corollary 3.8** For any couple  $(x, t) \in \mathring{F}$  such that  $\nu_\infty(x)f(x, t) > 0$ , for any  $(\alpha, \beta) \in \mathcal{A}$  and  $(v_0, w_0)$  satisfying (14), we have

$$\hat{f}^n(x, t) \xrightarrow{a.s.} f(x, t) \quad \text{and} \quad n^{\frac{1-\alpha d - \beta}{2}} \left( \hat{f}^n(x, t) - f(x, t) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\tau_1^2 \tau_d^2 f(x, t)}{(1 + \alpha d + \beta) \nu_\infty(x)} \right).$$

**Proof.** This result is a direct application of Theorem 3.3 and Slutsky's lemma.  $\square$

Another feature of interest for the inter-jumping times is the survival function  $G$ . One may estimate this quantity by,

$$\forall (x, t) \in \mathring{F}, \quad \widehat{G}^n(x, t) = \frac{\widehat{G}^n(x, t)}{\widehat{\nu}_\infty^n(x)},$$

with the convention  $0/0 = 0$ . Some properties of convergence are stated in the following corollary.

**Corollary 3.9** *For any couple  $(x, t) \in \mathring{F}$  such that  $\nu_\infty(x)f(x, t) > 0$ , for any  $\alpha$  such that  $\alpha d < 1$  and  $\alpha(d+2) > 1$ , and any  $v_0$ , we have*

$$\widehat{G}^n(x, t) \xrightarrow{a.s.} G(x, t) \quad \text{and} \quad n^{\frac{1-\alpha d}{2}} \left( \widehat{G}^n(x, t) - G(x, t) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\tau_d^2 G(x, t)}{(1 + \alpha d)\nu_\infty(x)} \right).$$

**Proof.** The almost sure convergence is a direct application of Theorem 3.3. The central limit theorem is a consequence of the asymptotic normality established in Remark 3.6 and Slutsky's lemma.  $\square$

### 3.2 Optimal estimation of the jump rate

We propose to estimate the composed function  $\lambda \circ \Phi$  by the ratio  $\widehat{\lambda \circ \Phi}^n$  defined by,

$$\forall (x, t) \in \mathring{F}, \quad \widehat{\lambda \circ \Phi}^n(x, t) = \frac{\widehat{\mathcal{F}}^n(x, t)}{\widehat{G}^n(x, t)}, \quad (16)$$

again with the convention  $0/0 = 0$ . Pointwise convergence and asymptotic normality are again a consequence of Theorem 3.3.

**Corollary 3.10** *For any couple  $(x, t) \in \mathring{F}$  such that  $\nu_\infty(x)f(x, t) > 0$ , for any  $(\alpha, \beta) \in \mathcal{A}$  and  $(v_0, w_0)$  satisfying (14), we have*

$$\widehat{\lambda \circ \Phi}^n(x, t) \xrightarrow{a.s.} \lambda \circ \Phi(x, t),$$

and

$$n^{\frac{1-\alpha d-\beta}{2}} \left( \widehat{\lambda \circ \Phi}^n(x, t) - \lambda \circ \Phi(x, t) \right) \xrightarrow{d} \mathcal{N} \left( 0, \frac{\tau_1^2 \tau_d^2 \lambda \circ \Phi(x, t)}{(1 + \alpha d + \beta)\nu_\infty(x)G(x, t)} \right).$$

**Proof.** This result is a direct application of Theorem 3.3 and Slutsky's lemma together with (6), (8) and (9).  $\square$

In all the sequel, we focus on the estimation of the rate  $\lambda$  (and not on the composed function  $\lambda \circ \Phi$ ) from the estimate  $\widehat{\lambda \circ \Phi}^n$  defined in (16). In particular, for some fixed value  $x \in E$ , we introduce a class of estimators  $\widehat{\lambda}_\xi^n(x)$  of  $\lambda(x)$  indexed by the elements  $\xi$  of the curve  $\mathcal{C}_x$  described by the reverse flow  $\Phi(x, -t)$ ,  $t \geq 0$ , and we propose to choose an estimate from this class in an optimal way. Before proceeding further we define the notation  $\mathcal{C}_x$  as

$$\mathcal{C}_x = \{ \Phi(x, -t) : 0 \leq t < t^-(x) \}.$$

By definition (1) of  $t^-(x)$  we have  $\mathcal{C}_x \subset E$ . In addition, for any  $\xi \in \mathcal{C}_x$ , we define  $\tau_x(\xi)$  as the unique time satisfying

$$\Phi(\xi, \tau_x(\xi)) = x, \quad \xi \in \mathcal{C}_x.$$

Thus we have the following trivial result,

$$\forall \xi \in \mathcal{C}_x, \quad \lambda \circ \Phi(\xi, \tau_x(\xi)) = \lambda(x). \quad (17)$$

As a consequence, we propose to define a class of estimators of  $\lambda(x)$  by

$$\Lambda^n(x) = \left\{ \widehat{\lambda}_\xi^n(x) = \widehat{\lambda \circ \Phi}^n(\xi, \tau_x(\xi)) : \xi \in \mathcal{C}_x \right\}, \quad (18)$$

where  $\widehat{\lambda \circ \Phi}^n$  estimates  $\lambda \circ \Phi$  (see Corollary 3.10) and has already been defined in (16). By virtue of Corollary 3.10 and only using (17), one has, as  $n$  goes to infinity,

$$\forall \xi \in \mathcal{C}_x, \quad \widehat{\lambda}_\xi^n(x) \xrightarrow{a.s.} \lambda(x) \quad \text{and} \quad n^{\frac{1-\alpha d-\beta}{2}} \left( \widehat{\lambda}_\xi^n(x) - \lambda(x) \right) \xrightarrow{d} \mathcal{N} \left( 0, \sigma_{\xi, \alpha, \beta}^2(x) \right),$$

where the asymptotic variance is given by

$$\sigma_{\xi, \alpha, \beta}^2(x) = \frac{\tau_1^2 \tau_d^2 \lambda(x)}{(1 + \alpha d + \beta) \kappa_x(\xi)}, \quad \text{with } \kappa_x(\xi) = \nu_\infty(\xi) G(\xi, \tau_x(\xi)). \quad (19)$$

In this paper, we choose to approximate  $\lambda(x)$  by the element  $\hat{\lambda}_\xi^n(x) \in \Lambda^n(x)$  minimizing the asymptotic variance  $\sigma_{\xi, \alpha, \beta}^2(x)$ . In other words, our optimal estimator of the jump rate  $\lambda(x)$  is obtained as

$$\hat{\lambda}_*^n(x) = \hat{\lambda}_{\xi_*}^n(x), \quad \text{where } \xi_* = \arg \max_{\xi \in \mathcal{C}_x} \kappa_x(\xi).$$

**Remark 3.11** *A good criterion should be to maximize the invariant measure  $\nu_\infty(\xi)$ ,  $\xi \in \mathcal{C}_x$ . Indeed, if  $\nu_\infty(\xi)$  is large, a large frequency of post-jump locations around  $\xi$  may be available in the dataset. Nevertheless, roughly speaking, the quantities of interest  $f(\xi, \tau_x(\xi))$  and  $G(\xi, \tau_x(\xi))$  are well estimated if a large number of post-jump locations are around  $\xi$  together with inter-jumping times around  $\tau_x(\xi)$ . This naive criterion may be corrected by including the quality of the estimation at time  $\tau_x(\xi)$ , that is to say, by maximizing the product  $\nu_\infty(\xi) G(\xi, \tau_x(\xi))$ . We also refer the reader to the simulation study presented in Section 4 and more precisely to Figure 7.*

The criterion  $\kappa_x(\xi)$ ,  $\xi \in \mathcal{C}_x$ , is generally uncomputable from the known features of the PDMP and thus remains to be estimated. In light of Theorem 3.3 and by definition (19) of  $\kappa_x(\xi)$ , we naturally propose to approximate this quantity by,

$$\forall \xi \in \mathcal{C}_x, \quad \hat{\kappa}_x^n(\xi) = \hat{\mathcal{G}}^n(\xi, \tau_x(\xi)). \quad (20)$$

As a consequence, we propose to estimate the jump rate  $\lambda(x)$  by its statistical approximation in  $\Lambda^n(x)$  maximizing the estimated criterion  $\hat{\kappa}_x^n(\xi)$ . More precisely,

$$\hat{\lambda}_*^n(x) = \hat{\lambda}_{\xi_*}^n(x), \quad \text{where } \xi_* = \arg \max_{\xi \in \mathcal{C}_x} \hat{\kappa}_x^n(\xi). \quad (21)$$

The high oscillations or alternatively the high smoothness of a kernel estimator with a ill-chosen bandwidth suggest that the choice of the parameter  $\alpha$  appearing in  $\hat{\kappa}_x^n$  (see eq. (12) and (20)) is crucial at this maximization step.

### 3.3 How to choose bandwidth parameters $\alpha$ and $\beta$ ?

This part is devoted to the choice of the bandwidth parameters  $\alpha$  and  $\beta$ . The criteria that we introduce in this part to choose these features are defined as line integrals along the curve  $\mathcal{C}_x$ . As a consequence, we need to ensure that this kind of quantity is well defined in our setting.

**Assumption 3.12** *For any starting condition  $\xi \in E$ , the reverse flow  $t \mapsto \Phi(\xi, -t)$  defines a change of variable, that is to say, is a diffeomorphic mapping.*

It is common in the literature to minimize the Integrated Square Error (ISE) to choose the optimal bandwidths of a kernel estimator, in particular from dependent data: one may refer the reader to [21, 24, 25]. Another classical solution is to investigate the behavior of the Mean Integrated Square Error (MISE). In the framework of Gaussian dependent data, the authors of [13] have shown that the optimal bandwidths obtained by minimizing the ISE and the MISE are very close if the dependence is of short range, that is to say, if the covariance function is integrable (see [13, Theorem 2.2]). It should be noted that a geometric ergodic Markov chain satisfies this kind of condition (see [27, Theorem 16.1.5]).

Let us recall that  $\alpha$  appears at first in the computation of the estimated criterion  $\hat{\kappa}_x^n$  which we need to maximize along the curve  $\mathcal{C}_x$ . Indeed  $\hat{\kappa}_x^n$  is computed (20) from the estimate  $\hat{\mathcal{G}}^n$  which implicitly depends on  $\alpha$ . We propose to choose the bandwidth parameter  $\alpha$  by minimizing the ISE associated with  $\hat{\kappa}_x$  and defined by

$$\begin{aligned} \text{ISE}_\kappa^n(\alpha) &= \int_{\mathcal{C}_x} (\hat{\kappa}_x^n(\xi) - \kappa_x(\xi))^2 d\xi \\ &= \int_{\mathcal{C}_x} \kappa_x(\xi)^2 d\xi + \varepsilon_\kappa^n(\alpha), \end{aligned} \quad (22)$$

where the function  $\varepsilon_\kappa^n$  is given by

$$\varepsilon_\kappa^n(\alpha) = \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\xi, \tau_x(\xi))^2 d\xi - 2 \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\xi, \tau_x(\xi)) \kappa_x(\xi) d\xi. \quad (23)$$

One may remark that here the ISE is unusually computed along a curve of interest. In (22) the dependency on  $\alpha$  only holds through the function  $\varepsilon_\kappa^n$ . As a consequence the optimal parameter  $\alpha$  minimizing this stochastic function (23) also minimizes the ISE.

The function  $\varepsilon_\kappa^n$  is generally not computable since the unknown quantity  $\kappa_x$  appears in its definition. As a consequence, we propose to estimate  $\varepsilon_\kappa^n$  by cross-validation which is a popular technique for selecting the bandwidth that minimizes the ISE. The authors would like to highlight that cross-validation involves here two main difficulties. First the estimators are computed from dependent data which are not identically distributed. In addition, there is almost surely no data on the set of integration  $\mathcal{C}_x$  whenever the dimension  $d$  is larger than 2. That is why we propose a specific procedure adapted to this framework.

Before defining our cross-validation estimate of  $\varepsilon_\kappa^n$ , we need to introduce the quantities  $\mathbb{T}_{x,\rho}$  and  $\theta_x$ . First  $\mathbb{H}_x$  denotes the hyperplane orthogonal to  $\mathcal{C}_x$  at  $x$ , that is,

$$\mathbb{H}_x = \left\{ y \in \mathbb{R}^d : y - x \perp \nabla_t \Phi(x, 0) \right\}.$$

In addition, for any  $\rho > 0$ , we introduce the notation  $\mathbb{D}_{x,\rho}$  for

$$\mathbb{D}_{x,\rho} = B_d(x, \rho) \cap \mathbb{H}_x.$$

Furthermore  $\mathbb{T}_{x,\rho}$  denotes the tube around  $\mathcal{C}_x$  with radius  $\rho$ ,

$$\mathbb{T}_{x,\rho} = \bigcup_{y \in \mathbb{D}_{x,\rho}} \mathcal{C}_y. \quad (24)$$

Finally, for any  $\xi \in \mathbb{T}_{x,\rho}$ ,  $\theta_x(\xi)$  denotes the unique time such that  $\Phi(\xi, \theta_x(\xi)) \in \mathbb{D}_{x,\rho}$ . In particular,

$$\forall \xi \in \mathcal{C}_x, \theta_x(\xi) = \tau_x(\xi). \quad (25)$$

It should be noted that, for  $\rho$  small enough,  $\mathbb{T}_{x,\rho} \subset E$ .

We focus now on a cross-validation method for estimating the quantity  $\varepsilon_\kappa^n$ . The estimate  $\widehat{\varepsilon}_\kappa^{n,\tilde{n},\rho}$  of  $\varepsilon_\kappa^n$  is defined from the observation of the embedded Markov chain  $(\tilde{Z}_k, \tilde{S}_{k+1})_{k \geq 0}$  of another PDMP  $(\tilde{X}_t)_{t \geq 0}$  (independent on the first one  $(X_t)_{t \geq 0}$  and distributed according to the same parameters), by

$$\widehat{\varepsilon}_\kappa^{n,\tilde{n},\rho}(\alpha) = \int_{\mathcal{C}_x} \widehat{\kappa}_x^n(\xi)^2 d\xi - \frac{2\Gamma\left(\frac{d-1}{2} + 1\right)}{\tilde{n}\pi^{\frac{d-1}{2}}\rho^{d-1}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{G}}^n(\tilde{Z}_k, \theta_x(\tilde{Z}_k)) \mathbb{1}_{\mathbb{T}_{x,\rho}}(\tilde{Z}_k) \mathbb{1}_{(\theta_x(\tilde{Z}_k), +\infty)}(\tilde{S}_{k+1}), \quad (26)$$

where  $\Gamma$  denotes as usually the Euler function. Some regularity conditions are necessary to investigate the asymptotic behavior of the cross-validation estimate  $\widehat{\varepsilon}_\kappa^{n,\tilde{n},\rho}$  in Proposition 3.14.

### Assumptions 3.13

- The sup-norm  $\|\nu_\infty\|_\infty$  is finite and  $\nu_\infty$  is Lipschitz,

$$\forall x, y \in E, |\nu_\infty(x) - \nu_\infty(y)| \leq [\nu_\infty]_{Lip} |x - y|.$$

- The deterministic exit time for the reverse flow  $t^-$  is Lipschitz,

$$\forall x, y \in E, |t^-(x) - t^-(y)| \leq [t^-]_{Lip} |x - y|.$$

- The flow  $\Phi$  is Lipschitz,

$$\forall x, y \in E, \forall t \in \mathbb{R}, |\Phi(x, t) - \Phi(y, t)| \leq [\Phi]_{Lip} |x - y|.$$

- The sup-norm  $\|\nabla_t \Phi\|_\infty$  is finite and  $\nabla_t \Phi$  is Lipschitz,

$$\forall x, y \in E, \forall t \in \mathbb{R}, |\nabla_t \Phi(x, t) - \nabla_t \Phi(y, t)| \leq [\nabla_t \Phi]_{Lip} |x - y|.$$

**Proposition 3.14** *Conditionally to  $\sigma(Z_0, S_1, \dots, Z_{n-1}, S_n)$ , we have*

$$\lim_{\substack{\tilde{n} \rightarrow \infty \\ \rho \rightarrow 0}} \widehat{\varepsilon}_{\kappa}^{n, \tilde{n}, \rho}(\alpha) = \varepsilon_{\kappa}^n(\alpha) \quad a.s.$$

**Proof.** The proof is stated in Appendix C.  $\square$

By virtue of Proposition 3.14, one may obtain an estimate of the optimal bandwidth parameter  $\alpha$  arising in  $\widehat{\kappa}_x^n$  by minimizing the quantity  $\widehat{\varepsilon}_{\kappa}^{n, \tilde{n}, \rho}$  for some small enough  $\rho$  and large enough  $\tilde{n}$ . In addition, by (16), (18) and (20), the quantity  $\widehat{\kappa}_x^n$  also appears in the calculus of the estimator  $\widehat{\lambda}_{\xi}^n(x)$ . In particular, the same choice of  $\alpha$  may be done for computing the denominator  $\widehat{\kappa}_x^n(\xi)$  of  $\widehat{\lambda}_{\xi}^n(x)$ . As a consequence it remains to choose in an optimal way the bandwidth parameters  $\alpha$  and  $\beta$  arising in the formula (16) of the numerator  $\widehat{\mathcal{F}}^n(\xi, \tau_x(\xi))$  of  $\widehat{\lambda}_{\xi}^n(x)$  (18).

In a similar way as before, we propose to choose  $\alpha$  and  $\beta$  by minimizing the ISE associated with  $\widehat{\mathcal{F}}^n(\cdot, \tau_x(\cdot))$  and computed along the curve  $\mathcal{C}_x$ ,

$$\begin{aligned} \text{ISE}_{\mathcal{F}}^n(\alpha, \beta) &= \int_{\mathcal{C}_x} \left( \widehat{\mathcal{F}}^n(\xi, \tau_x(\xi)) - \mathcal{F}(\xi, \tau_x(\xi)) \right)^2 d\xi \\ &= \int_{\mathcal{C}_x} \mathcal{F}(\xi, \tau_x(\xi))^2 d\xi + \varepsilon_{\mathcal{F}}^n(\alpha, \beta), \end{aligned} \quad (27)$$

where  $\widehat{\mathcal{F}}^n$  implicitly depends on  $\alpha$  and  $\beta$ ,  $\mathcal{F}(\xi, \tau_x(\xi))$  stands for  $\nu_{\infty}(\xi)f(\xi, \tau_x(\xi))$  and  $\varepsilon_{\mathcal{F}}^n$  is given by

$$\varepsilon_{\mathcal{F}}^n(\alpha, \beta) = \int_{\mathcal{C}_x} \widehat{\mathcal{F}}^n(\xi, \tau_x(\xi))^2 d\xi - 2 \int_{\mathcal{C}_x} \widehat{\mathcal{F}}^n(\xi, \tau_x(\xi)) \mathcal{F}(\xi, \tau_x(\xi)) d\xi.$$

As in the previous part we propose to estimate  $\varepsilon_{\mathcal{F}}^n$  by cross-validation from the observation of the embedded chain  $(\tilde{Z}_k, \tilde{S}_{k+1})_{k \geq 0}$  of another PDMP  $(\tilde{X}_t)_{t \geq 0}$ . We define our estimate  $\widehat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho_1, \rho_2}$  of  $\varepsilon_{\mathcal{F}}^n$  by

$$\begin{aligned} \widehat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho_1, \rho_2}(\alpha, \beta) &= \int_{\mathcal{C}_x} \widehat{\mathcal{F}}^n(\xi, \tau_x(\xi))^2 d\xi \\ &\quad - \frac{2 \Gamma\left(\frac{d-1}{2} + 1\right)}{\tilde{n} \rho_2 \pi^{\frac{d-1}{2}} \rho_1^{d-1}} \sum_{k=0}^{\tilde{n}-1} \widehat{\mathcal{F}}^n\left(\tilde{Z}_k, \theta_x(\tilde{Z}_k)\right) \mathbb{1}_{\mathbb{T}_{x, \rho_1}}(\tilde{Z}_k) \mathbb{1}_{(\theta_x(\tilde{Z}_k) - \frac{\rho_2}{2}, \theta_x(\tilde{Z}_k) + \frac{\rho_2}{2})}(\tilde{S}_{k+1}). \end{aligned} \quad (28)$$

The convergence of  $\widehat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho, \delta}$  is investigated in Proposition 3.15.

**Proposition 3.15** *Conditionally to  $\sigma(Z_0, S_1, \dots, Z_{n-1}, S_n)$ , we have*

$$\lim_{\substack{\tilde{n} \rightarrow \infty \\ \rho_{1,2} \rightarrow 0}} \widehat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho_1, \rho_2}(\alpha, \beta) = \varepsilon_{\mathcal{F}}^n(\alpha, \beta) \quad a.s.$$

**Proof.** The proof is similar to the demonstration of Proposition 3.14 stated in Appendix C.  $\square$

## 4 Numerical illustration

This section is devoted to the implementation of the whole estimation procedure presented in this paper. The application on which we focus is a variant of the famous TCP window size process appearing in the modeling of the Transmission Control Protocol used for data transmission over the Internet and presented in [11]. This protocol has been designed to adapt to the traffic conditions of the network: for a connection, the maximum number of packets that can be sent is given by a random variable called the congestion window size. At each time step, if all the packets are successfully transmitted, then one tries to transmit one more packet until a congestion appears.

## 4.1 Presentation of the model

The model presented in this part is two dimensional. For the sake of clarity, we will use the following notation: for any  $x \in \mathbb{R}^2$ ,  $x_1$  and  $x_2$  denote the components of  $x$ .

We consider a PDMP  $(X_t)_{t \geq 0}$  evolving on the state space  $E = (0, 1)^2$ . The deterministic part of the model is defined from the flow  $\Phi$  given by,

$$\forall x \in \mathbb{R}^2, \forall t \in \mathbb{R}, \Phi(x, t) = (x_1 + t, x_2).$$

The jump rate  $\lambda$  is defined by,

$$\forall x \in \mathbb{R}^2, \lambda(x) = x_1 + x_2.$$

The transition kernel  $\mathcal{Q}$  is defined by,

$$\forall x \in E, \forall A \in \mathcal{B}(E), \mathcal{Q}(x, A) \propto \int_A u(1-u)^{2/x_1-1} v(1-v) du dv.$$

Starting from  $(x_1, x_2)$ , the process evolves in the unit square, always to the right, until a jump appears either when the motion hits the boundary  $\{(\xi_1, \xi_2) : \xi_1 = 1\}$  or with the non homogeneous rate  $x_1 + x_2 + t$  before, that is according to a Weibull distribution. The two components of the post-jump location are independent and both governed by a Beta-distribution, in such a way that the process tends with a high probability to jump to the left of the location just before the jump. One obtains a TCP-like process (see Figure 1) for which the second dimension models the quality of the network (upper the second component is, higher the probability of a congestion is).

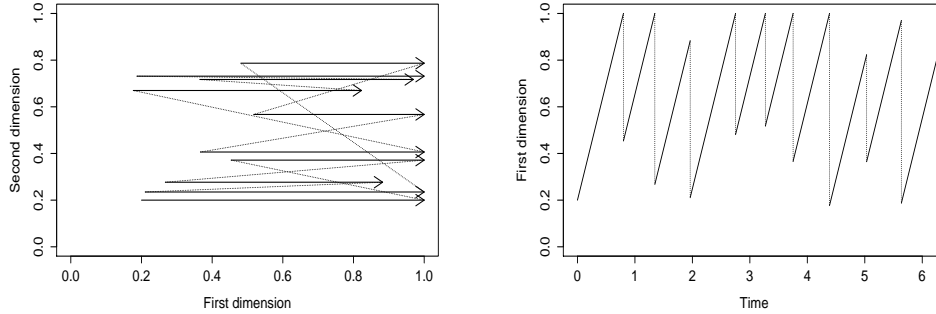


Figure 1: Two representations of the same simulated path of the TCP-like model of interest until the 10<sup>th</sup> jump. A vector field graph is given on the left, while we observe the trajectory of the first component versus time on the right.

The asymptotic behavior of the process may be represented by the invariant distribution  $\nu_\infty$  of the post-jump locations. Since this quantity is unknown, we propose to show in Figure 2 its estimate  $\hat{\nu}_\infty^n$  defined by (13) and computed from  $n = 20\,000$  observed jumps.

## 4.2 Estimation procedure

We present here all the procedure for estimating the jump rate  $\lambda$  at the location  $x = (0.75, 0.5)$  for which the quality of the network is average. It should be noted that the invariant distribution at  $x$  is quite low (see Figure 2). Nevertheless our method is expected to work pretty well even in this unfavorable framework.

In this context, the class of estimators of  $\lambda(x)$  is indexed by the elements  $\xi \in \mathcal{C}_x = (0, 0.75] \times \{0.5\}$  (see eq. (18)). Let us recall that our estimation algorithm is divided into four steps:

- Step 1** Choose the bandwidth parameter  $\alpha$  for  $\hat{\kappa}_x^n(\xi)$  defined in (20) by minimizing the error  $\text{ISE}_\kappa^n(\alpha)$  given by (22). Since this quantity is uncomputable, one maximizes its cross-validation estimate (up to an additive constant)  $\hat{\varepsilon}_\kappa^{n,\tilde{n},\rho}(\alpha)$  given in (26).
- Step 2** Compute the optimal point  $\hat{\xi}_*^n \in \mathcal{C}_x$  that maximizes  $\hat{\kappa}_x^n(\xi)$  where the implicit parameter  $\alpha$  has been chosen at Step 1 (see eq. (21)).
- Step 3** Choose the bandwidth parameters  $\alpha$  and  $\beta$  for  $\hat{\mathcal{F}}^n(\xi, \tau_x(\xi))$  by minimizing the error  $\text{ISE}_\mathcal{F}^n(\alpha, \beta)$  defined by (27). Since this quantity is uncomputable, one maximizes its cross-validation estimate (up to an additive constant)  $\hat{\varepsilon}_\mathcal{F}^{n,\tilde{n},\rho_1,\rho_2}(\alpha, \beta)$  given in (28).
- Step 4** Compute the estimate  $\hat{\lambda}_*(x) = \hat{\lambda}_{\hat{\xi}_*^n}^n(x)$ , where  $\hat{\xi}_*^n$  has been computed at Step 2, and with the optimal parameters  $\alpha$  and  $\beta$  obtained at Steps 1 and 3. More precisely, this estimate is the ratio (see eq. (16) and (18)) of  $\hat{\mathcal{F}}^n(\hat{\xi}_*^n, \tau_x(\hat{\xi}_*^n))$  (computed from the parameters  $\alpha$  and  $\beta$  obtained at Step 3) over  $\hat{\mathcal{G}}^n(\hat{\xi}_*^n, \tau_x(\hat{\xi}_*^n)) = \hat{\kappa}_x^n(\hat{\xi}_*^n)$  (computed from the parameter  $\alpha$  obtained at Step 1).

One may remark that Steps 2 and 3 are obviously permutable.

**Remark 4.1** *The time complexity of this algorithm depends on several parameters, namely the number  $n$  of observed jumps, the number  $\tilde{n}$  of observed data for the cross-validation steps, and also the numbers  $N_\xi$ ,  $N_\alpha$  and  $N_\beta$  for discretizing the state spaces of  $\xi$ ,  $\alpha$  and  $\beta$  at each maximization procedure. It is easy to see that the third step is the most complex but remains polynomial, precisely in  $O(nN_\alpha N_\beta(N_\xi + \tilde{n}))$ .*

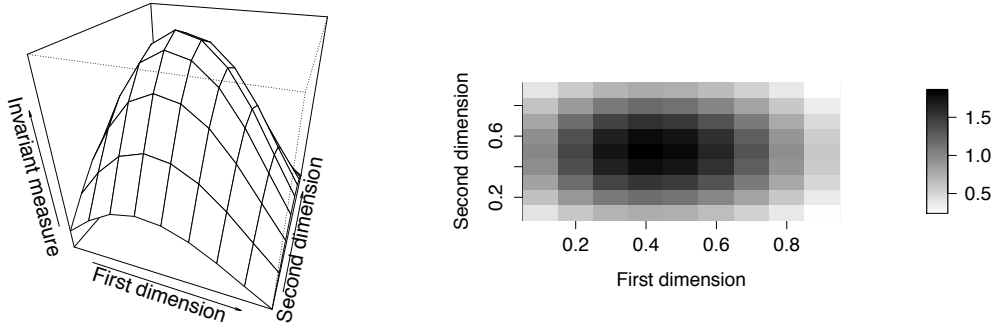


Figure 2: Estimation of the invariant distribution of the post-jump locations computed from the 20 000 first jumps of the TCP-like model.

### 4.3 Simulation results

In this simulation study, we assume that we observe the embedded Markov chain  $(Z_n, S_{n+1})_{n \geq 0}$  until the 10 000<sup>th</sup> jump. The cross-validation procedure is computed from an additional chain, independent on the first one, and observed until the 1 000<sup>th</sup> jump. When boxplots are presented, they have been computed over 100 replicates.

- Step 1** We begin with the choice of the bandwidth parameter  $\alpha$  appearing in  $\hat{\kappa}_x^n(\xi) = \hat{\mathcal{G}}^n(\xi, \tau_x(\xi))$ . The cross-validation procedure relies on the minimization of the estimate  $\hat{\varepsilon}_\kappa^{n,\tilde{n},\rho}(\alpha)$  which depends on the positive parameter  $\rho$ . We present in Figure 3 this quantity as a function of  $\alpha$  and from different values of  $\rho$ . Fortunately, this new parameter seems to have little influence over the behavior of the estimation of the ISE along  $\mathcal{C}_x$ .
- Step 2** Now we maximize the estimated criterion  $\hat{\kappa}_x^n(\xi)$  (of course computed with the optimal parameter  $\alpha$  obtained at the previous step) along the curve  $\mathcal{C}_x$  (see Figure 4): we obtain the optimal point  $\hat{\xi}_*^n$  at which we will compute our estimator of  $\lambda(x)$  at Step 4. The crucial role of  $\alpha$  at this maximization step is illustrated in Figure 5.

**Step 3** We continue with the choice of the couple  $(\alpha, \beta)$  implicitly appearing in the estimator  $\hat{\mathcal{F}}^n(\xi, \tau_x(\xi))$ . The optimal parameters are obtained by minimizing the estimate  $\hat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho_1, \rho_2}(\alpha, \beta)$  of the related ISE (see Figure 6).

**Step 4** We compute the estimators  $\hat{\lambda}_{\xi}^n(x)$  for different values of  $\xi \in \mathcal{C}_x$  and with the optimal bandwidths obtained at Steps 1 and 3. The related boxplots are presented in Figure 7. The procedure makes us able to choose the best index  $\hat{\xi}_*^n$  which most of the time corresponds with the estimate with least bias and variance. This proves the strong interest of the estimation algorithm developed in this paper.

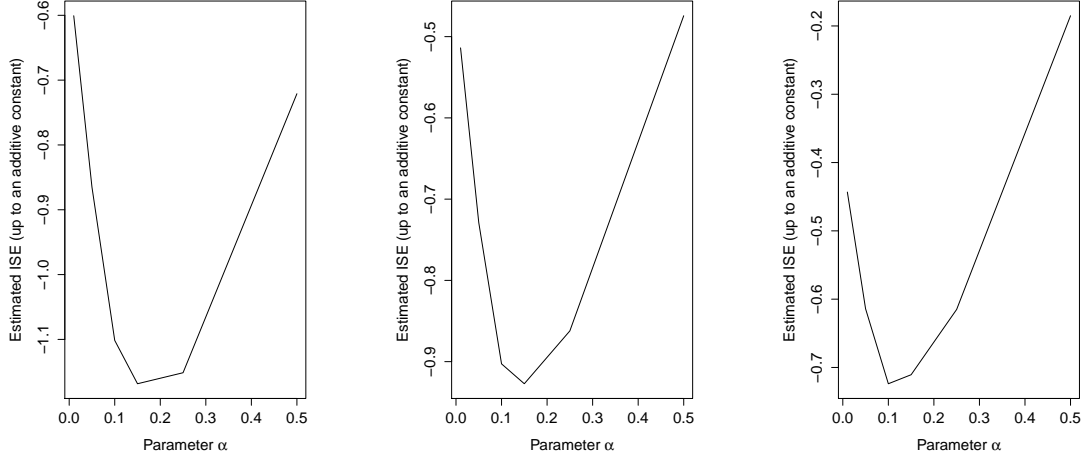


Figure 3: The choice of the bandwidth parameter  $\alpha$  appearing in the criterion  $\hat{\kappa}_x^n$  is obtained by minimizing the cross-validation estimate  $\hat{\varepsilon}_{\kappa}^{n, \tilde{n}, \rho}(\alpha)$  of the related ISE at the first step of the algorithm. The parameter  $\rho$  seems to have only a small influence over the minimization: the estimated error is computed from  $\rho = 0.005$ ,  $\rho = 0.01$  and  $\rho = 0.02$  from left to right.

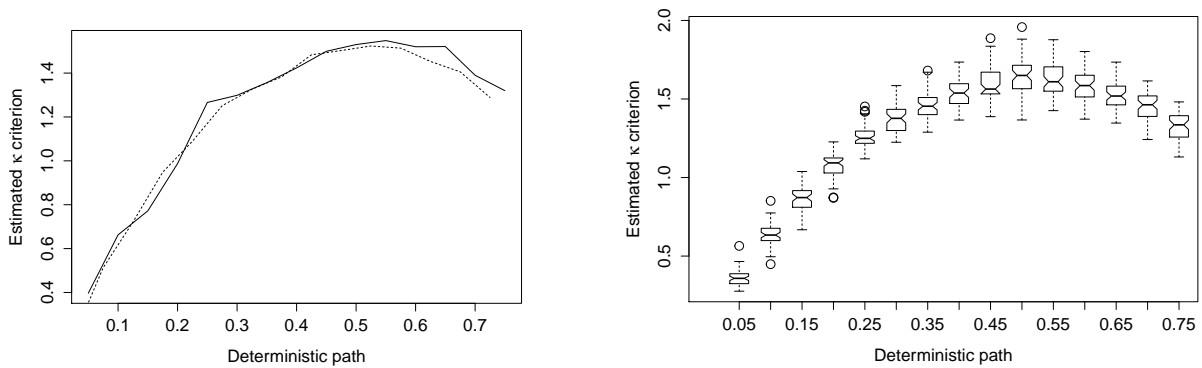


Figure 4: The optimal index  $\hat{\xi}_*^n$  is calculated by maximizing the criterion  $\hat{\kappa}_x^n(\xi)$ ,  $\xi \in \mathcal{C}_x$ , computed with the optimal parameter  $\alpha$ . On the left side of the figure, we compare the estimate  $\hat{\kappa}_x^n(\xi)$  (full line) with  $\hat{\nu}_{\infty}^n(\xi)G(\xi, \tau_x(\xi))$  (dashed line) both computed from  $\alpha = 0.1$ , which confirms that the estimation performs pretty well. This quantity seems to admit one and only one absolute maximum.



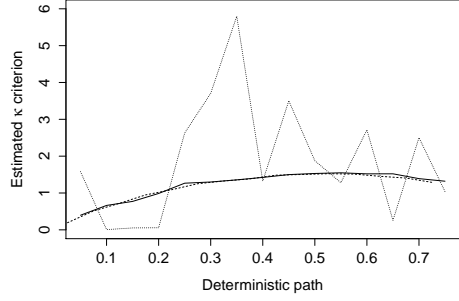


Figure 5: The parameter  $\alpha$  plays a crucial role in the estimation of  $\hat{\kappa}_x^n$ . We compare here the curves computed from  $\alpha = 0.1$  and already presented in Figure 4 (full and dashed lines) with the too oscillating estimate obtained from  $\alpha = 0.4$  (dotted line).

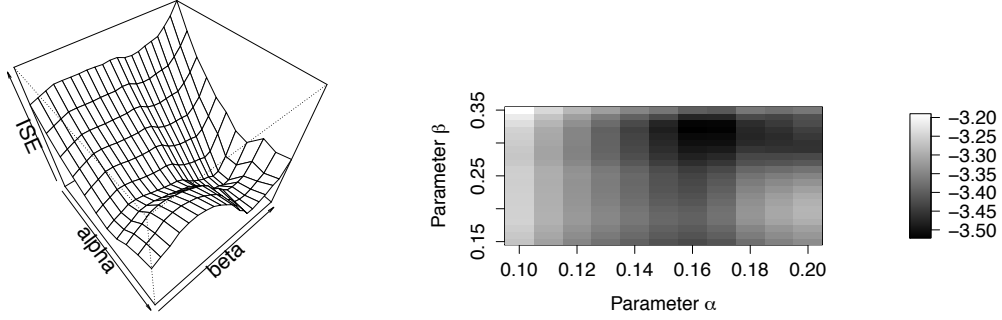


Figure 6: The optimal bandwidth parameters  $\alpha$  and  $\beta$  implicitly appearing in  $\hat{\mathcal{F}}^n(\xi, \tau_x(\xi))$ ,  $\xi \in \mathcal{C}_x$  are obtained by minimizing the cross-validation estimate  $\hat{\varepsilon}_{\mathcal{F}}^{n, \tilde{n}, \rho_1, \rho_2}(\alpha, \beta)$ , computed from  $\rho_1 = \rho_2 = 0.1$  here, of the related ISE at the third step of the procedure. The parameter  $\beta$  seems to have little influence over the estimation error in comparison to  $\alpha$ .

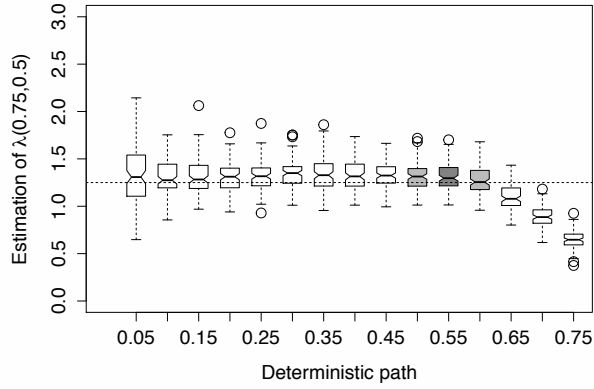


Figure 7: The estimator  $\hat{\lambda}_\xi^n(x)$  is computed from different values of the index  $\xi \in \mathcal{C}_x$ , and from the optimal bandwidth parameters obtained at Steps 1 and 3. Over the 100 replicates, the optimal points  $\hat{\xi}_*^n$  obtained at Step 2 are most of the time located at 0.55 and more than 90 times between 0.5 and 0.6 (enhanced by gray colors), which seems to correspond with the estimators with least bias and variance. In particular, we obtain a better result than by maximizing the estimated invariant measure  $\hat{\nu}_\infty^n$  (index around 0.35), see Remark 3.11.

## A Ergodicity and invariant measures

In this section we present some preliminary and technical results about the invariant distributions of the underlying Markov chains. We begin with some properties of the Markov chain  $(Z_n)_{n \geq 0}$  of the post-jump locations of the PDMP of interest. In particular, under Assumption 2.1, one may state the following result.

**Proposition A.1** *We have the following statements:*

- $(Z_n)_{n \geq 0}$  is  $\nu_\infty$ -irreducible, positive Harris-recurrent and aperiodic.
- $\nu_\infty$  is the unique invariant distribution of  $(Z_n)_{n \geq 0}$ .

**Proof.** The proof is similar to the demonstration of Proposition 4.2 of [4].  $\square$

These properties make us able to apply the law of large numbers to the Markov chain  $(Z_n)_{n \geq 0}$  (see [27, Theorem 17.1.7]). Now we propose to focus on the sequence  $(Z_n, S_{n+1})_{n \geq 0}$  which also forms a Markov process whose transition kernel is given by  $\mathcal{R}$  in (11). Let us recall that  $\mu_n$  denotes the distribution on the state space  $F$  defined by (10) of the couple  $(Z_n, S_{n+1})$ ,  $n \geq 0$ . We define the measure  $\mu_\infty$  by,

$$\forall A \in \mathcal{B}(\mathbb{R}^d), \forall t \geq 0, \mu_\infty(A \times (t, +\infty)) = \int_A \nu_\infty(dx) \mathcal{S}(x, (t, +\infty)), \quad (29)$$

where the conditional distribution  $\mathcal{S}$  is given by (6). In particular,  $\mu_\infty$  is the unique invariant distribution of  $(Z_n, S_{n+1})_{n \geq 0}$ .

**Proposition A.2** *We have the following statements:*

- For any initial distribution  $\mu_0 = \delta_{\{x, t\}}$ ,  $(x, t) \in F$ , we have

$$\lim_{n \rightarrow \infty} \|\mu_n - \mu_\infty\|_{TV} = 0.$$

- $(Z_n, S_{n+1})_{n \geq 0}$  is  $\mu_\infty$ -irreducible, positive Harris-recurrent and aperiodic.
- $\mu_\infty$  is the unique invariant distribution of  $(Z_n, S_{n+1})_{n \geq 0}$ .

**Proof.** The proof is similar to the demonstrations of Lemma 4.4 and Proposition 4.5 of [4].  $\square$

We are able again to apply the law of large numbers to the Markov chain  $(Z_n, S_{n+1})_{n \geq 0}$  by virtue of this result. It should be noted that the measures  $\mu_n$ ,  $n \geq 0$ , and  $\mu_\infty$  do not share with  $\nu_n$ ,  $n \geq 0$ , and  $\nu_\infty$  the property of absolute continuity with respect to the Lebesgue measure presented in Remark 2.2.

**Remark A.3** *For any  $n$ ,  $\mu_n$  and thus the limit  $\mu_\infty$  admit a density only on the interior  $\mathring{F}$  of the state space  $F$  because of the expression (7) of  $\mathcal{S}$  and thanks to Remark 2.2.*

Finally, we would like to highlight that the link between the measures  $\nu_\infty$  and  $\mu_\infty$  may be expressed in another way than (29). Indeed, for any  $x \in E$ , we have

$$\int_F \mathcal{Q}(\Phi(y, s), x) \mu_\infty(dy \times ds) = \nu_\infty(x), \quad (30)$$

by the expression (5) of the transition kernel  $\mathcal{P}$  of  $(Z_n)_{n \geq 0}$  and because  $\nu_\infty \mathcal{P} = \nu_\infty$  by virtue of Proposition A.1. The formula (30) of  $\nu_\infty$  will be useful in our investigations.

## B Proof of Theorem 3.3

In all this section we use the notation  $\mathbb{F}_n$  for the  $\sigma$ -algebra  $\sigma(Z_0, S_1, \dots, Z_n, S_{n+1})$ ,  $n \geq 0$ . In addition, the classical symbols  $\sim$ ,  $o$  and  $O$  must be understood to hold here in the almost sure sense.

## B.1 Sketch of the proof

The proof of Theorem 3.3 relies on the following decomposition. For any integer  $n \geq 1$ , we have

$$\begin{bmatrix} \widehat{\mathcal{F}}^n(x, t) \\ \widehat{\mathcal{G}}^n(x, t) \\ \widehat{\nu}_\infty^n(x) \end{bmatrix} - \begin{bmatrix} \nu_\infty(x)f(x, t) \\ \nu_\infty(x)G(x, t) \\ \nu_\infty(x) \end{bmatrix} = \frac{\mathcal{M}_{n-1}}{n} + \mathcal{R}_n, \quad (31)$$

where the sequence  $(\mathcal{M}_n)_{n \geq 0}$  is a  $(\mathbb{F}_n)_{n \geq 0}$ -vector martingale defined by (35) and studied in Appendix B.3, and  $\mathcal{R}_n$  is a remainder term defined by (32) and studied in Appendix B.2.

In Appendix B.2, we establish that the remainder term almost surely goes to 0 when  $n$  tends to infinity. This is a first step to show the almost sure convergence presented in Theorem 3.3. In addition we investigate the rate of convergence of the remainder term in Appendix B.2 under an additional Lipschitz mixing condition stated in Assumption 2.4. This is enough to prove the asymptotic normality given in Theorem 3.3.

The rest of the proof deals with the study of the martingale term. We prove in Appendix B.3 that the process  $(\mathcal{M}_n)_{n \geq 0}$  is a vector martingale. We investigate its asymptotic behavior by studying its square variation process  $(\langle \mathcal{M} \rangle_n)_{n \geq 0}$  in Appendix B.3. Thanks to these results we state the law of large numbers and the central limit theorem for  $(\mathcal{M}_n)_{n \geq 0}$  in Appendix B.3.

Finally the almost sure convergence presented in Theorem 3.3 is a direct application of (31) and (33) together with (49), while the asymptotic normality is obtained from (31), (34) and (52).

## B.2 Remainder term

This part of the paper is devoted to the asymptotic study of the remainder term sequence  $(\mathcal{R}_n)_{n \geq 1}$  appearing in (31).

### B.2.1 Definition of the remainder term

For any  $n$ , one may write  $\mathcal{R}_n = [R_n^{(k)}]_{1 \leq k \leq 3}$ , where each of the components is defined by

$$R_n^{(1)} = \sum_{j=0}^4 R_n^{(1,j)}, \quad R_n^{(2)} = \sum_{j=0}^3 R_n^{(2,j)} \quad \text{and} \quad R_n^{(3)} = \sum_{j=0}^2 R_n^{(3,j)}. \quad (32)$$

The five terms that define  $R_n^{(1)}$  are given by

$$\begin{aligned} R_n^{(1,0)} &= \frac{1}{nv_0^d w_0} \mathbb{K}_d \left( \frac{Z_0 - x}{v_0} \right) \mathbb{K}_1 \left( \frac{S_1 - t}{w_0} \right), \\ R_n^{(1,1)} &= \frac{1}{n} \sum_{j=1}^{n-1} \int_{\mathbb{R}^d \times \mathbb{R}} f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) (\mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) - \mathcal{Q}(\Phi(Z_{j-1}, S_j), x)) \, du \, dv, \\ R_n^{(1,2)} &= \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) \int_{\mathbb{R}^d \times \mathbb{R}} [f(x + uv_j, t + vw_j) - f(x, t + vw_j)] \mathbb{K}_d(u) \mathbb{K}_1(v) \, du \, dv, \\ R_n^{(1,3)} &= \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) \int_{\mathbb{R}^d \times \mathbb{R}} [f(x, t + vw_j) - f(x, t)] \mathbb{K}_d(u) \mathbb{K}_1(v) \, du \, dv, \\ R_n^{(1,4)} &= \left[ \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) - \nu_\infty(x) \right] f(x, t). \end{aligned}$$

The four terms that define  $R_n^{(2)}$  are given by

$$\begin{aligned}
R_n^{(2,0)} &= \frac{1}{n} \frac{1}{v_0^d} \mathbb{K}_d \left( \frac{Z_0 - x}{v_0} \right) \mathbb{1}_{\{S_1 > t\}}, \\
R_n^{(2,1)} &= \frac{1}{n} \sum_{j=1}^{n-1} \int_{\mathbb{R}^d} [\mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) - \mathcal{Q}(\Phi(Z_{j-1}, S_j), x)] G(x + uv_j, t) \mathbb{K}_d(u) du, \\
R_n^{(2,2)} &= \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) \int_{\mathbb{R}^d} [G(x + uv_j, t) - G(x, t)] \mathbb{K}_d(u) du, \\
R_n^{(2,3)} &= \left[ \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) - \nu_\infty(x) \right] G(x, t).
\end{aligned}$$

Finally, the three terms that define  $R_n^{(3)}$  are given by

$$\begin{aligned}
R_n^{(3,0)} &= \frac{1}{n} \frac{1}{v_0^d} \mathbb{K}_d \left( \frac{Z_0 - x}{v_0} \right), \\
R_n^{(3,1)} &= \frac{1}{n} \sum_{j=1}^{n-1} \int_{\mathbb{R}^d} [\mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) - \mathcal{Q}(\Phi(Z_{j-1}, S_j), x)] \mathbb{K}_d(u) du, \\
R_n^{(3,2)} &= \frac{1}{n} \sum_{j=1}^{n-1} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) - \nu_\infty(x).
\end{aligned}$$

### B.2.2 Almost sure convergence

We only investigate the first component  $(R_n^{(1)})_{n \geq 1}$  since the other terms may be treated with similar arguments. First, it is obvious that the sequence  $(R_n^{(1,0)})_{n \geq 1}$  almost surely tends to 0. In addition, the term  $(R_n^{(1,4)})_{n \geq 1}$  converges to 0 by virtue of the ergodic theorem applied to the Markov chain  $(Z_n, S_{n+1})_{n \geq 0}$  (thanks to Proposition A.2) together with (30). Because both the functions  $\mathcal{Q}$  and  $f$  are Lipschitz and bounded (see Assumptions 2.3), we have

$$\begin{aligned}
|R_n^{(1,1)}| &\leq \frac{[\mathcal{Q}]_{Lip} \|f\|_\infty}{n} \sum_{j=1}^{n-1} v_j \int_{\mathbb{R}^d} |u| \mathbb{K}_d(u) du, \\
|R_n^{(1,2)}| &\leq \frac{\|\mathcal{Q}\|_\infty [f]_{Lip}}{n} \sum_{j=1}^{n-1} v_j \int_{\mathbb{R}^d} |u| \mathbb{K}_d(u) du, \\
|R_n^{(1,3)}| &\leq \frac{\|\mathcal{Q}\|_\infty [f]_{Lip}}{n} \sum_{j=1}^{n-1} w_j \int_{\mathbb{R}} |v| \mathbb{K}_1(v) dv,
\end{aligned}$$

which are some Cesàro means because both the sequences  $(v_n)_{n \geq 0}$  and  $(w_n)_{n \geq 0}$  tend to 0. Thus,

$$\mathcal{R}_n \xrightarrow{a.s.} 0_3. \quad (33)$$

### B.2.3 Rate of convergence

Under the first item of Assumptions 2.4, the Markov chain  $(Z_n)_{n \geq 0}$  and thus the two-dimensional process  $(Z_n, S_{n+1})_{n \geq 0}$  satisfy the contraction assumption given in Theorem 6.3.17 of [19]. By applying this theorem to the function  $\varphi = \mathcal{Q}(\Phi(\cdot, \cdot), \cdot) \in \text{Li}(r_1, r_2)$  (see second item of Assumptions 2.4), we obtain that, for any  $\gamma \in (0, 1)$  and  $x \in E$ ,

$$n^{-\frac{1+\gamma}{2}} \sum_{j=1}^{n-1} [\mathcal{Q}(\Phi(Z_{j-1}, S_j), x) - \nu_\infty(x)] \xrightarrow{a.s.} 0,$$

because  $2(r_1 + r_2) \leq a_1$ . Therefore, we have for any couple  $(k, j)$  in the set  $\{(1, 4), (2, 3), (3, 2)\}$ ,

$$n^{\frac{1-\alpha d - \beta}{2}} R_n^{(k,j)} \xrightarrow{a.s.} 0.$$

The same result is obvious for the other couples  $(k, j)$  with the condition  $\alpha d + \beta + 2 \min(\alpha, \beta) > 1$ . Finally we obtain, when  $n$  goes to infinity,

$$n^{\frac{1-\alpha d-\beta}{2}} \mathcal{R}_n \xrightarrow{a.s.} 0_3. \quad (34)$$

### B.3 Martingale term

#### B.3.1 $(\mathcal{M}_n)_{n \geq 0}$ is a vector martingale

Let  $(\mathcal{M}_n)_{n \geq 0}$  be the three-dimensional process defined for any  $n$  by  $\mathcal{M}_n = [M_n^{(k)}]_{1 \leq k \leq 3}$ , where each component is defined as

$$M_n^{(k)} = \sum_{j=1}^n A_j^{(k)} - B_j^{(k)}, \quad k \in \{1, 2, 3\}, \quad (35)$$

and the terms  $A_j^{(k)}$  and  $B_j^{(k)}$  are given by

$$\begin{aligned} A_j^{(1)} &= \frac{1}{v_j^d} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right) \frac{1}{w_j} \mathbb{K}_1 \left( \frac{S_{j+1} - t}{w_j} \right), \\ B_j^{(1)} &= \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) \lambda_d(du) dv, \\ A_j^{(2)} &= \frac{1}{v_j^d} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right) \mathbb{1}_{\{S_{j+1} > t\}}, \\ B_j^{(2)} &= \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d(u) \lambda_d(du), \\ A_j^{(3)} &= \frac{1}{v_j^d} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right), \\ B_j^{(3)} &= \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d(u) \lambda_d(du). \end{aligned}$$

We claim that the process  $(\mathcal{M}_n)_{n \geq 0}$  is a  $(\mathbb{F}_n)_{n \geq 0}$ -martingale. The keystone is to show that, for any  $n$  and  $k$ , we have

$$\mathbb{E} [A_n^{(k)} \mid \mathbb{F}_{n-1}] = B_{n-1}^{(k)}.$$

The proof presents no particular difficulty except for the first component  $k = 1$  for which we provide some details. Let us recall that  $\mathcal{R}$  denotes the transition kernel of the Markov chain  $(Z_n, S_{n+1})_{n \geq 0}$  (see Subsection 2.1). We have

$$\begin{aligned} \mathbb{E} [A_n^{(1)} \mid \mathbb{F}_{n-1}] &= \frac{1}{v_n^d w_n} \int_{\mathbb{R}^d \times \mathbb{R}_+} \mathbb{K}_d \left( \frac{\xi - x}{v_n} \right) \mathbb{K}_1 \left( \frac{s - t}{w_n} \right) \mathcal{R}((Z_{n-1}, S_n), d\xi \times ds) \\ &= \frac{1}{v_n^d w_n} \int_{\mathbb{R}^d \times \mathbb{R}_+} \mathbb{K}_d \left( \frac{\xi - x}{v_n} \right) \mathbb{K}_1 \left( \frac{s - t}{w_n} \right) \mathcal{Q}(\Phi(Z_{n-1}, S_n), d\xi) \mathcal{S}(\xi, ds), \end{aligned} \quad (36)$$

with (11). Let us recall that the bandwidth sequence  $(v_n)_{n \geq 0}$  is decreasing. Together with the third item of Assumptions 3.1, we obtain

$$\text{supp } \mathbb{K}_d \left( \frac{\cdot - x}{v_n} \right) \subset \text{supp } \mathbb{K}_d \left( \frac{\cdot - x}{v_0} \right) \subset B_d(x, v_0 \delta).$$

With similar arguments, we obtain

$$\text{supp } \mathbb{K}_1 \left( \frac{\cdot - t}{w_n} \right) \subset \text{supp } \mathbb{K}_1 \left( \frac{\cdot - t}{w_0} \right) \subset B_1(t, w_0 \delta) \subset \left( -\infty, \inf_{\xi \in B_d(x, v_0 \delta)} t^+(\xi) \right), \quad (37)$$

with the condition (14) on the couple  $(v_0, w_0)$ . By (36), (37) together with the expression (7) of  $\mathcal{S}$  we obtain

$$\mathbb{E} [A_n^{(1)} \mid \mathbb{F}_{n-1}] = \frac{1}{v_n^d w_n} \int_{\mathbb{R}^d \times \mathbb{R}_+} \mathbb{K}_d \left( \frac{\xi - x}{v_n} \right) \mathbb{K}_1 \left( \frac{s - t}{w_n} \right) \mathcal{Q}(\Phi(Z_{n-1}, S_n), d\xi) f(\xi, s) ds.$$

One may conclude by a change of variable and (3).

### B.3.2 Predictable square variation process

The asymptotic behavior of the martingale  $(\mathcal{M}_n)_{n \geq 0}$  may be investigated by studying its predictable square variation process that we denote as usual  $(\langle \mathcal{M} \rangle_n)_{n \geq 0}$ . At any time  $n$ ,  $\langle \mathcal{M} \rangle_n$  is the  $3 \times 3$  symmetric matrix defined by

$$\langle \mathcal{M} \rangle_n = \sum_{k=1}^n \left[ \mathbb{E} \left[ M_k^{(i)} M_k^{(j)} \mid \mathbb{F}_{k-1} \right] - M_{k-1}^{(i)} M_{k-1}^{(j)} \right]_{1 \leq i, j \leq 3}.$$

Now, we calculate each coefficient of this matrix by beginning with the diagonal terms and the first of these  $\langle \mathcal{M} \rangle_n^{(1,1)}$ . One has

$$\langle \mathcal{M} \rangle_n^{(1,1)} = n T_n^{(1)} - T_n^{(2)},$$

where the terms  $T_n^{(1)}$  and  $T_n^{(2)}$  are defined by

$$\begin{aligned} T_n^{(1)} &= \frac{1}{n} \sum_{j=1}^n \frac{1}{v_j^d w_j} \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d^2(u) \mathbb{K}_1^2(v) du dv, \\ T_n^{(2)} &= \sum_{j=1}^n \left( \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) du dv \right)^2. \end{aligned}$$

Since the functions  $\mathcal{Q}$  and  $f$  are bounded (see Assumptions 2.3), we easily obtain that

$$T_n^{(2)} = O(n). \quad (38)$$

Let us introduce the additional notation,

$$\begin{aligned} \tilde{T}_n^{(1)} &= \frac{1}{n} \sum_{j=1}^n \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d^2(u) \mathbb{K}_1^2(v) du dv, \\ \check{T}_n^{(1)} &= \frac{1}{n} \sum_{j=1}^n \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) f(x, t) \int_{\mathbb{R}^d \times \mathbb{R}} \mathbb{K}_d^2(u) \mathbb{K}_1^2(v) du dv \\ &= \frac{\tau_d^2 \tau_1^2}{n} \sum_{j=1}^n \mathcal{Q}(\Phi(Z_{j-1}, S_j), x) f(x, t), \end{aligned}$$

by Remark 3.2 on the kernel functions  $\mathbb{K}_1$  and  $\mathbb{K}_d$ . Since both  $\mathcal{Q}$  and  $f$  are Lipschitz and bounded (see Assumptions 2.3), together with  $\text{supp } \mathbb{K}_p \subset B_p(0_p, \delta)$  (see Assumptions 3.1), the stochastic sequences  $(\tilde{T}_n^{(1)})_{n \geq 0}$  and  $(\check{T}_n^{(1)})_{n \geq 0}$  have the same limit thanks to

$$\left| \tilde{T}_n^{(1)} - \check{T}_n^{(1)} \right| \leq \frac{(\|\mathcal{Q}\|_\infty [f]_{Lip} + [\mathcal{Q}]_{Lip} \|f\|_\infty) \delta \tau_d^2 \tau_1^2}{n} \sum_{j=1}^n (v_j + w_j), \quad (39)$$

which is a Cesàro mean. In addition, by the almost sure ergodic theorem (see Proposition A.2) together with (30), we obtain

$$\check{T}_n^{(1)} \xrightarrow{a.s.} \tau_d^2 \tau_1^2 \nu_\infty(x) f(x, t). \quad (40)$$

By (39) and (40) and by virtue of Lemma 7.1.5 of [19], we obtain

$$\frac{T_n^{(1)}}{n^{\alpha d + \beta}} \xrightarrow{a.s.} \frac{\tau_d^2 \tau_1^2 \nu_\infty(x) f(x, t)}{1 + \alpha d + \beta}.$$

Together with (38), as  $n$  goes to infinity,

$$\frac{\langle \mathcal{M} \rangle_n^{(1,1)}}{n^{1 + \alpha d + \beta}} \xrightarrow{a.s.} \frac{\tau_d^2 \tau_1^2 \nu_\infty(x) f(x, t)}{1 + \alpha d + \beta} = \Delta_1. \quad (41)$$

The second diagonal term of the predictable variation process may be studied in a similar way as before. We obtain

$$\begin{aligned} \langle \mathcal{M} \rangle_n^{(2,2)} &= \sum_{j=1}^n \left[ \frac{1}{v_j^d} \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d^2(u) du \right. \\ &\quad \left. - \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d(u) du \right)^2 \right], \end{aligned}$$

and, as  $n$  goes to infinity,

$$\frac{\langle \mathcal{M} \rangle_n^{(2,2)}}{n^{1+\alpha d}} \xrightarrow{a.s.} \frac{\tau_d^2 \nu_\infty(x) G(x, t)}{1 + \alpha d} = \Delta_2. \quad (42)$$

The third and last diagonal term may also be investigated in the same way. We have

$$\begin{aligned} \langle \mathcal{M} \rangle_n^{(3,3)} &= \sum_{j=1}^n \left[ \frac{1}{v_j^d} \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d^2(u) du \right. \\ &\quad \left. - \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d(u) du \right)^2 \right], \end{aligned}$$

and, when  $n$  goes to infinity,

$$\frac{\langle \mathcal{M} \rangle_n^{(3,3)}}{n^{1+\alpha d}} \xrightarrow{a.s.} \frac{\tau_d^2 \nu_\infty(x)}{1 + \alpha d} = \Delta_3. \quad (43)$$

Now we focus on the non diagonal terms. For any integer  $n$ , we have

$$\begin{aligned} \langle \mathcal{M} \rangle_n^{(1,2)} &= \sum_{j=1}^n \left[ \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j)^2 f(x + uv_j, t + vw_j) G(x + uv_j, t) \mathbb{K}_d^2(u) \mathbb{K}_1(v) du dv \right. \\ &\quad - \left( \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) du dv \right) \\ &\quad \left. \times \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d(u) du \right) \right], \end{aligned} \quad (44)$$

$$\begin{aligned} \langle \mathcal{M} \rangle_n^{(1,3)} &= \sum_{j=1}^n \left[ \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j)^2 f(x + uv_j, t + vw_j) \mathbb{K}_d^2(u) \mathbb{K}_1(v) du dv \right. \\ &\quad - \left( \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) du dv \right) \\ &\quad \left. \times \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d(u) du \right) \right], \end{aligned} \quad (45)$$

$$\begin{aligned} \langle \mathcal{M} \rangle_n^{(2,3)} &= \sum_{j=1}^n \left[ \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j)^2 G(x + uv_j, t) \mathbb{K}_d^2(u) du \right. \\ &\quad - \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d(u) du \right) \\ &\quad \left. \times \left( \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d(u) du \right) \right]. \end{aligned} \quad (46)$$

From (44), (45) and (46), and because  $\mathcal{Q}$  and  $f$  are bounded (see Assumptions 2.3) together with the fact that the integral  $\int \mathbb{K}_p^2 d\lambda_p$  is finite (see Remark 3.2), we easily obtain that, for any  $i \neq j$ ,

$$\langle \mathcal{M} \rangle_n^{(i,j)} = O(n). \quad (47)$$

As a conclusion, by (41), (42), (43) and (47), one may sum up the asymptotic behavior of the predictable variation process  $(\langle \mathcal{M} \rangle_n)_{n \geq 0}$  by the following formula,

$$\frac{\langle \mathcal{M} \rangle_n}{n^{1+\alpha d}} \sim \begin{bmatrix} n^\beta \Delta_1 & 0 & 0 \\ 0 & \Delta_2 & 0 \\ 0 & 0 & \Delta_3 \end{bmatrix}. \quad (48)$$

It should be noted that the coefficients  $\Delta_i$ ,  $1 \leq i \leq 3$ , are positive because we assume that  $\nu_\infty(x) > 0$  and  $f(x, t) > 0$  in the statement of Theorem 3.3.

### B.3.3 Limit theorems for the vector martingale

**Law of large numbers.** We propose to apply the law of large numbers for vector martingales (see [19, Theorem 4.3.15]) to the process of interest  $(\mathcal{M}_n)_{n \geq 0}$ . In the sequel, for any  $n$ ,  $\mathcal{T}_n$  denotes the trace of the matrix  $\langle \mathcal{M} \rangle_n$ , while  $\mathcal{E}_n$  stands for its minimum eigenvalue. First, in light of (48), the trace  $(\mathcal{T}_n)_{n \geq 0}$  almost surely tends to infinity. Thus we are able to apply the third item of Theorem 4.3.15 of [19] with any function  $a(t) = t^{1+\eta}$ ,  $\eta > 0$ , and we obtain,

$$\left\| \langle \mathcal{M} \rangle_n^{-1/2} \mathcal{M}_n \right\|^2 = o\left(\frac{\log(\mathcal{T}_n)^{1+\eta}}{\mathcal{E}_n}\right).$$

By (48) again, we have, when  $n$  goes to infinity,

$$\mathcal{T}_n = O\left(n^{1+\alpha d+\beta}\right) \quad \text{and} \quad \mathcal{E}_n \sim \min(\Delta_2, \Delta_3) n^{1+\alpha d}.$$

As a consequence, using that  $n^{1+\alpha d+\beta} = o(n^2)$  whenever  $\alpha d + \beta < 1$  together with (48), we have the law of large numbers, as  $n$  tends to infinity,

$$\frac{\|\mathcal{M}_n\|}{n} \xrightarrow{a.s.} 0. \quad (49)$$

**Central limit theorem.** We now investigate the asymptotic normality of the vector martingale  $(\mathcal{M}_n)_{n \geq 0}$ . We apply Corollary 2.1.10 of [19] with the sequence  $(a_n)_{n \geq 0}$  defined by  $a_n = n^{1+\alpha d+\beta}$ . The first assumption of this result is obviously satisfied: the sequence  $(a_n^{-1} \langle \mathcal{M} \rangle_n)_{n \geq 0}$  almost surely converges to some positive semi-definite matrix. Indeed, by (48),

$$n^{-1-\alpha d-\beta} \langle \mathcal{M} \rangle_n \xrightarrow{a.s.} \begin{bmatrix} \frac{\tau_d^2 \tau_1^2 \nu_\infty(x) f(x, t)}{1+\alpha d+\beta} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \Sigma(x, t, \alpha, \beta), \quad (50)$$

where  $\Sigma(x, t, \alpha, \beta)$  is a degenerate variance-covariance matrix. As a consequence, we only have to check Lindeberg's condition in order to establish the central limit theorem for  $(\mathcal{M}_n)_{n \geq 0}$ . In other words, we have to prove that, for any  $\epsilon > 0$ ,

$$\frac{1}{n^{1+\alpha d+\beta}} \sum_{j=1}^n \mathbb{E} \left[ |M_j - M_{j-1}|^2 \mathbb{1}_{\left\{ |M_j - M_{j-1}| \geq \epsilon n^{\frac{1+\alpha d+\beta}{2}} \right\}} \middle| \mathbb{F}_{j-1} \right] \xrightarrow{\mathbb{P}} 0. \quad (51)$$

For any  $1 \leq j \leq n$  and  $1 \leq k \leq 3$ , we study the three components  $M_j^{(k)} - M_{j-1}^{(k)}$ . We have

$$\begin{aligned} M_j^{(1)} - M_{j-1}^{(1)} &= \frac{(j+1)^{\alpha d+\beta}}{v_0^d w_0} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right) \mathbb{K}_1 \left( \frac{S_{j+1} - t}{w_j} \right) \\ &\quad - \int_{\mathbb{R}^d \times \mathbb{R}} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) f(x + uv_j, t + vw_j) \mathbb{K}_d(u) \mathbb{K}_1(v) \lambda_d(du) dv, \\ M_j^{(2)} - M_{j-1}^{(2)} &= \frac{(j+1)^{\alpha d}}{v_0^d} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right) \mathbb{1}_{\{S_{j+1} > t\}} \\ &\quad - \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) G(x + uv_j, t) \mathbb{K}_d(u) \lambda_d(du), \\ M_j^{(3)} - M_{j-1}^{(3)} &= \frac{(j+1)^{\alpha d}}{v_0^d} \mathbb{K}_d \left( \frac{Z_j - x}{v_j} \right) - \int_{\mathbb{R}^d} \mathcal{Q}(\Phi(Z_{j-1}, S_j), x + uv_j) \mathbb{K}_d(u) \lambda_d(du). \end{aligned}$$

Thus, we obtain

$$\begin{aligned} |M_j^{(1)} - M_{j-1}^{(1)}| &\leq \frac{(n+1)^{\alpha d+\beta}}{v_0^d w_0} \|\mathbb{K}_d\|_\infty \|\mathbb{K}_1\|_\infty + \|Q\|_\infty \|f\|_\infty, \\ |M_j^{(2)} - M_{j-1}^{(2)}| &\leq \frac{(n+1)^{\alpha d}}{v_0^d} \|\mathbb{K}_d\|_\infty + \|Q\|_\infty, \\ |M_j^{(3)} - M_{j-1}^{(3)}| &\leq \frac{(n+1)^{\alpha d}}{v_0^d} \|\mathbb{K}_d\|_\infty + \|Q\|_\infty. \end{aligned}$$



Together with the condition  $\alpha d + \beta < 1$ ,  $|M_j - M_{j-1}| = o\left(n^{\frac{1+\alpha d+\beta}{2}}\right)$ . As a consequence, there exists an integer  $N_\epsilon$  such that for any  $n \geq N_\epsilon$ , the event  $\left\{|M_j - M_{j-1}| \geq \epsilon n^{\frac{1+\alpha d+\beta}{2}}\right\}$  is almost surely empty. This shows Lindeberg's condition (51). Finally, by (50) and (51), we obtain, as  $n$  tends to infinity,

$$n^{-\frac{1+\alpha d+\beta}{2}} \mathcal{M}_n \xrightarrow{d} \mathcal{N}(0_3, \Sigma(x, t, \alpha, \beta)). \quad (52)$$

## C Proof of Proposition 3.14

By virtue of the ergodic theorem (see Proposition A.2) applied to the Markov chain  $(\tilde{Z}_n, \tilde{S}_{n+1})_{n \geq 0}$ , we have, as  $\tilde{n}$  tends to infinity,

$$\widehat{\varepsilon}_{\kappa}^{n, \tilde{n}, \rho}(\alpha) \xrightarrow{a.s.} \int_{\mathcal{C}_x} \widehat{\kappa}_x^n(\xi)^2 d\xi - \frac{2\Gamma\left(\frac{d-1}{2} + 1\right)}{\pi^{\frac{d-1}{2}} \rho^{d-1}} \int_{\mathbb{T}_{x, \rho} \times \mathbb{R}_+} \widehat{\mathcal{G}}^n(\xi, \theta_x(\xi)) \mathbb{1}_{(\theta_x(\xi), +\infty)}(t) \mu_\infty(d\xi \times dt),$$

conditionally to  $\sigma(Z_0, S_1, \dots, Z_{n-1}, S_n)$ . Together with the definition of  $\mu_\infty$  (29), the expression (9) of  $S(x, \cdot)$  and Remark 2.2, we obtain

$$\widehat{\varepsilon}_{\kappa}^{n, \tilde{n}, \rho}(\alpha) \xrightarrow{a.s.} \int_{\mathcal{C}_x} \widehat{\kappa}_x^n(\xi)^2 d\xi - \frac{2\Gamma\left(\frac{d-1}{2} + 1\right)}{\pi^{\frac{d-1}{2}} \rho^{d-1}} \int_{\mathbb{T}_{x, \rho}} \widehat{\mathcal{G}}^n(\xi, \theta_x(\xi)) G(\xi, \theta_x(\xi)) \nu_\infty(\xi) \lambda_d(d\xi), \quad (53)$$

conditionally to  $\sigma(Z_0, S_1, \dots, Z_{n-1}, S_n)$ . By definition (24) of  $\mathbb{T}_{x, \rho}$  we have

$$\int_{\mathbb{T}_{x, \rho}} \widehat{\mathcal{G}}^n(\xi, \theta_x(\xi)) G(\xi, \theta_x(\xi)) \nu_\infty(\xi) \lambda_d(d\xi) = \int_{\mathbb{D}_{x, \rho}} \left( \int_{\mathcal{C}_y} \widehat{\mathcal{G}}^n(\xi, \theta_y(\xi)) G(\xi, \theta_y(\xi)) \nu_\infty(\xi) d\xi \right) \lambda_{d-1}(dy). \quad (54)$$

In addition, one has

$$\int_{\mathcal{C}_y} \widehat{\mathcal{G}}^n(\xi, \theta_y(\xi)) G(\xi, \theta_y(\xi)) \nu_\infty(\xi) d\xi = \int_0^{t^-(y)} \widehat{\mathcal{G}}^n(\Phi(y, -t), t) G(\Phi(y, -t), t) \nu_\infty(\Phi(y, -t)) |\nabla_t \Phi(y, -t)| dt.$$

As a consequence, under Assumptions 3.13, a conscientious calculus together with eq. (25) shows that

$$\left| \int_{\mathcal{C}_y} \widehat{\mathcal{G}}^n(\xi, \theta_y(\xi)) G(\xi, \theta_y(\xi)) \nu_\infty(\xi) d\xi - \int_{\mathcal{C}_x} \widehat{\mathcal{G}}^n(\xi, \tau_x(\xi)) G(\xi, \tau_x(\xi)) \nu_\infty(\xi) d\xi \right| = O(|x - y|). \quad (55)$$

We obtain the expected result from (53), (54) and (55) and by remarking that the normalizing constant is  $\lambda_{d-1}(\mathbb{D}_{x, \rho})$ .

**Remark C.1** *In order to prove (55), one may split the integral of interest into two terms: an integral between 0 and  $t^-(x) \wedge t^-(y)$  and a remainder term integrated between  $t^-(x) \wedge t^-(y)$  and  $t^-(x) \vee t^-(y)$ . The first one is clearly upper bounded by the integral of a Lipschitz function between 0 and  $t^-(x) + |x - y| [t^-]_{Lip}$ , thus bounded by an integral between 0 and  $t^-(x) + \rho [t^-]_{Lip}$ . The second integral is obviously bounded by  $\|\widehat{\mathcal{G}}^n\|_\infty \|\nu_\infty\|_\infty \|\nabla_t \Phi\|_\infty [t^-]_{Lip} |x - y|$ .*

## References

- [1] Odd Olai Aalen. *Statistical inference for a family of counting processes*. ProQuest LLC, Ann Arbor, MI, 1975. Thesis (Ph.D.)—University of California, Berkeley.
- [2] Terje Aven and Uwe Jensen. *Stochastic models in reliability*, volume 41 of *Applications of Mathematics (New York)*. Springer-Verlag, New York, 1999.
- [3] Romain Azaïs. A recursive nonparametric estimator for the transition kernel of a piecewise-deterministic Markov process. *ESAIM: Probability and Statistics*, 18:726–749, January 2014.
- [4] Romain Azaïs, François Dufour, and Anne Gégout-Petit. Nonparametric estimation of the jump rate for non-homogeneous marked renewal processes. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(4):1204–1231, 11 2013.

- [5] Romain Azaïs, François Dufour, and Anne Gégout-Petit. Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic markov processes. *Scandinavian Journal of Statistics*, 41(4):950–969, 2014.
- [6] Romain Azaïs and Alexandre Genadot. Semi-parametric inference for the absorption features of a growth-fragmentation model. *TEST*, 24(2):341–360, 2015.
- [7] Patrice Bertail, Stephan Clemençon, and Jessica Tressou. A storage model with random release rate for modelling exposure to food contaminants. *Mathematical Bioscience and Engineering*, 5(1):35–60, 2008.
- [8] Patrice Bertail, Stephan Clemençon, and Jessica Tressou. Statistical analysis of a dynamic model for dietary contaminant exposure. *Journal of Biological Dynamics*, 4(2):212–234, 2010.
- [9] Adrien Brandejsky, Benoîte de Saporta, and François Dufour. Numerical method for expectations of piecewise-deterministic markov processes. *CAMCoS*, 7(1):63–104, 2012.
- [10] Adrien Brandejsky, Benoîte De Saporta, and François Dufour. Numerical methods for the exit time of a piecewise-deterministic markov process. *Adv. in Appl. Probab.*, 44(1):196–225, 03 2012.
- [11] Djalil Chafai, Florent Malrieu, and Katy Paroux. On the long time behavior of the TCP window size process. *Stochastic Processes and their Applications*, 120(8):1518–1534, April 2010.
- [12] Julien Chiquet and Nikolaos Limnios. A method to compute the transition function of a piecewise deterministic Markov process with application to reliability. *Statist. Probab. Lett.*, 78(12):1397–1403, 2008.
- [13] Gerda Claeskens and Peter Hall. Effect of dependence on stochastic measures of accuracy of density estimations. *The Annals of Statistics*, 30(2):431–454, 2002.
- [14] Alina Crudu, Arnaud Debussche, Aurélie Muller, Ovidiu Radulescu, et al. Convergence of stochastic gene networks to hybrid piecewise deterministic processes. *The Annals of Applied Probability*, 22(5):1822–1859, 2012.
- [15] Mark H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.
- [16] Benoîte de Saporta, François Dufour, Huilong Zhang, and Charles Elegbede. Optimal stopping for the predictive maintenance of a structure subject to corrosion. *Journal of Risk and Reliability*, 226(2):169–181, 2012.
- [17] Marie Doumic, Marc Hoffmann, Nathalie Krell, and Lydia Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli (to appear)*, 2015.
- [18] Marie Doumic, Marc Hoffmann, Patricia Reynaud-Bouret, and Vincent Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis*, 50(2):925–950, 2012.
- [19] Marie Duflo. *Random iterative models*. Applications of Mathematics. Springer-Verlag, Berlin, 1997.
- [20] Alexandre Genadot and Michèle Thieullen. Averaging for a fully coupled piecewise-deterministic markov process in infinite dimensions. *Adv. in Appl. Probab.*, 44(3):749–773, 09 2012.
- [21] Jeffrey D. Hart and Philippe Vieu. Data-driven bandwidth choice for density estimation based on dependent data. *The Annals of Statistics*, pages 873–890, 1990.
- [22] Jianghai Hu, Wei-Chung Wu, and Shankar Sastry. *Modeling subtilin production in bacillus subtilis using stochastic hybrid systems*. In R. Alur and G.J. Pappas, editors, *Hybrid Systems: Computation and Control*, number 2993 in LNCS, Springer-Verlag, Berlin, 2004.
- [23] Martin Jacobsen. *Point Process Theory and Applications: Marked Point and Piecewise Deterministic Processes*. Probability and its Applications. Birkhäuser, Boston-Basel-Berlin, 2006.
- [24] Tae Yoon Kim. Asymptotically optimal bandwidth selection rules for the kernel density estimator with dependent observations. *Journal of statistical planning and inference*, 59(2):321–336, 1997.
- [25] Tae Yoon Kim and Dennis D. Cox. Bandwidth selection in kernel smoothing of time series. *Journal of Time Series Analysis*, 17(1):49–63, 1996.
- [26] Nathalie Krell. Statistical estimation of jump rates for a specific class of piecewise deterministic markov processes. *arXiv:1406.2845*, March 2015.

- [27] Sean Meyn and Richard L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.
- [28] Ovidiu Radulescu, Aurélie Muller, and Alina Crudu. Théorèmes limites pour les processus de markov à sauts. *Technique et Science Informatiques*, 26(3-4):443–469, 2007.
- [29] Lydia Robert, Marc Hoffmann, Nathalie Krell, Stephane Aymerich, Jerome Robert, and Marie Doumic. Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC Biology*, 12(1):17, 2014.