ESAIM: PS 18 (2014) 726–749
DOI: 10.1051/ps/2013054

ESAIM: Probability and Statistics
www.esaim-ps.org

A RECURSIVE NONPARAMETRIC ESTIMATOR FOR THE TRANSITION KERNEL OF A PIECEWISE-DETERMINISTIC MARKOV PROCESS *

Romain Azaïs

Abstract. In this paper, we investigate a nonparametric approach to provide a recursive estimator of the transition density of a piecewise-deterministic Markov process, from only one observation of the path within a long time. In this framework, we do not observe a Markov chain with transition kernel of interest. Fortunately, one may write the transition density of interest as the ratio of the invariant distributions of two embedded chains of the process. Our method consists in estimating these invariant measures. We state a result of consistency and a central limit theorem under some general assumptions about the main features of the process. A simulation study illustrates the well asymptotic behavior of our estimator.

Mathematics Subject Classification. 62G05, 62M05.

Received January 8, 2013. Revised August 2, 2013.

1. Introduction

The purpose of this paper is to investigate a nonparametric recursive method for estimating the transition kernel of a piecewise-deterministic Markov process, from only one observation of the process within a long time interval.

Piecewise-deterministic Markov processes (PDMP's) have been introduced in the literature by Davis in [7]. They are a general class of non-diffusion stochastic models involving deterministic motion broken up by random jumps, which occur either when the flow reaches the boundary of the state space or in a Poisson-like fashion. The path depends on three local features namely the flow Φ, which controls the deterministic trajectories, the jump rate λ, which governs the inter-jumping times, and the transition kernel Q, which determines the post-jump locations. An appropriate choice of the state space and the main characteristics of the process covers a large variety of stochastic models covering problems in reliability (see [7] and [5]) or in biology (see [14] and [11]) for instance. In this context, it appears natural to propose some nonparametric methods for estimating both the characteristics λ and Q, which control the randomness of the motion. Indeed, the deterministic flow is given by physical equations or deterministic biological models. In [2], Azaïs *et al.* proposed a kernel method for estimating the conditional probability density function associated with the jump rate λ, for a non-stationary PDMP defined on a general metric space. This work was based on a generalization of Aalen's multiplicative

Keywords and phrases. Piecewise-deterministic Markov processes, nonparametric estimation, recursive estimator, transition kernel, asymptotic consistency.

* *This work was supported by ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004.*

[1] INRIA Bordeaux Sud-Ouest, team CQFD, France and Université Bordeaux, IMB, CNRS UMR 5251, 200 Avenue de la Vieille Tour, 33405 Talence cedex, France. romain.azais@inria.fr

Article published by EDP Sciences
© EDP Sciences, SMAI 2014

intensity model and a discretization of the state space. In the present paper, we assume that Q admits a density q with respect to the Lebesgue measure, and we focus on the nonparametric estimation of this function, from the observation of a PDMP within a long time, without assumption of stationarity. In addition, since measured data are often processed sequentially, it is convenient to propose a recursive estimator. To the best of our knowledge, no nonparametric estimation procedures are available in the literature for general PDMP's.

Nonparametric estimation methods for stationary Markov chains have been extensively investigated, beginning with Roussas in [21]. He studied kernel methods for estimating the stationary density and the transition kernel of a Markov chain satisfying the strong Doeblin's condition. Later, Rosenblatt proposed in [20] some results on the bias and the variance of this estimator in a weaker framework. Next, Yakowitz improved in [23] the previous asymptotic normality result assuming a Harris-condition. Masry and Györfi in [18], and Basu and Sahoo in [3], have completed this survey. There exists also an extensive literature on nonparametric estimates for non-stationary Markov processes. We do not attempt to present an exhaustive survey on this topic, but refer the interested reader to [6,8,10,12,15–17] and the references therein. In this new framework, Doukhan and Ghindès have investigated in [8] a bound of the integrated risks for their estimate. Hernández-Lerma *et al.* in [12] and Duflo in [10] made inquiries about recursive methods for estimating the transition kernel or the invariant distribution of a Markov chain. Liebscher gave in [17] some results under a weaker condition than Doeblin's assumption. More recently, Clémençon in [6] proposed a quotient estimator using wavelets and provided the lower bound of the minimax $\mathbf{L}^p$-risk. Lacour suggested in [16] an estimator by projection with model selection, next she introduced in [15] an original estimate by inquiring into a new contrast derived from regression framework.

Our investigation and the studies of the literature mentioned before are different and complementary. In this paper, we propose to estimate the transition density q of a PDMP by kernel methods. Nevertheless, we do not observe a Markov chain whose transition distribution is given by q. Fortunately, one may write the function of interest as the ratio of two invariant measures: the one of the two components pre-jump location and post-jump location, over the one of the pre-jump location. Indeed, $Q(x, A)$ is defined as the conditional probability that the post-jump location is in A, given the path is in x just before the jump. Therefore, we suggest to estimate both these invariant measures in order to provide an estimator of the transition kernel Q. A major stumbling block for estimating the invariant law of the pre-jump location is related with the transition kernel of this Markov chain, which may charge the boundary of the state space. As a consequence, the transition kernel, as well as the corresponding invariant distribution, admits a density only on the interior of the state space. The investigated approach for estimating the invariant measure is based on this property of the transition kernel. But the main difficulty appears for analyzing the two-components process pre-jump location, post-jump location. This Markov chain has a special structure, because its invariant distribution admits a density function on the interior of the state space, unlike its transition kernel. Indeed, the pre-jump location is distributed on the curve governed by the deterministic flow initialized by the previous post-jump location. As a consequence, the author have to explore a new method for estimating the two-dimensional invariant measure of interest. The proposed one is more universal, but implies a more restrictive assumption on the shape of the bandwidth.

An intrinsic complication throughout the paper comes from the presence of deterministic jumps, when the path tries to cross the boundary of the state space. Indeed, this induces that the invariant distributions mentioned above may charge a subset with null Lebesgue measure. This important feature has been introduced by Davis in [7] and is very attractive for the modeling of a large number of applications. For instance, one may find in [1] an example of shock models with failures of threshold type. One may also refer the reader to [14], where the authors develop a PDMP to capture the mechanism behind antibiotic released by a bacteria. Forced jumps are used to model a deterministic switching when the concentration of nutrients rises over a certain threshold.

The paper is organized as follows. Section 2 is devoted to the precise formulation of a piecewise-deterministic Markov process. The main results of convergence are presented in Section 3. The consistency of our nonparametric recursive estimator is stated in Theorem 3.1, while a central limit theorem lies in Theorem 3.2. Section 4 deals with numerical considerations for illustrating the asymptotic behavior of our estimate. Finally, some concluding remarks are given in Section 5. The strategy and the proofs of the main results are deferred into Appendices A, B and C.

2. Definition of a PDMP

We present the definition of a piecewise-deterministic Markov process on $\mathbf{R}^d$, where d is an integer greater or equal to 1. The process evolves in an open subset E of $\mathbf{R}^d$ equipped with the Euclidean norm $|\cdot|$. The motion is defined by the three local characteristics (λ, Q, Φ).

- $\Phi : \mathbf{R}^d \times \mathbf{R} \to \mathbf{R}^d$ is the deterministic flow. It satisfies,

$$\forall \xi \in \mathbf{R}^d,\ \forall s,t \in \mathbf{R},\ \Phi_\xi(t+s) = \Phi_{\Phi_\xi(t)}(s).$$

  For each $\xi \in E$, $t^+(\xi)$ denotes the deterministic exit time from E:

$$t^+(\xi) = \inf\{t > 0\ :\ \Phi_\xi(t) \in \partial E\},$$

  with the usual convention $\inf \emptyset = +\infty$.
- $\lambda : \mathbf{R}^d \to \mathbf{R}_+$ is the jump rate. It is a measurable function which satisfies,

$$\forall \xi \in \mathbf{R}^d,\ \exists \varepsilon > 0,\ \int_0^\varepsilon \lambda\big(\Phi_\xi(s)\big)\mathrm{d}s < +\infty.$$

- Q is a Markov kernel on $(\mathbf{R}^d, \mathcal{B}(\mathbf{R}^d))$ which satisfies, for any $\xi \in \mathbf{R}^d$,

$$Q(\xi, E \setminus \{\xi\}) = 1 \quad \text{and,} \quad \forall B \in \mathcal{B}(\mathbf{R}^d),\ Q(\xi, B) = \int_B q(\xi, z)\mathrm{d}z,$$

  where the transition density q is piecewise-continuous.

There exists a filtered probability space $(\Omega, \mathcal{A}, (\mathcal{F}_t), \mathbf{P})$, on which a process (X_t) is defined (see [7]). Its motion, starting from $x \in E$, can be described as follows. T_1 is a positive random variable whose survival function is,

$$\forall t \geq 0,\ \mathbf{P}(T_1 > t | X_0 = x) = \exp\left(-\int_0^t \lambda(\Phi_x(s))\mathrm{d}s\right)\mathbf{1}_{\{0 \leq t < t^+(x)\}}.$$

One chooses an E-valued random variable Z_1 according to the distribution $Q(\Phi_x(T_1), \cdot)$. Let us remark that the post-jump location depends only on the pre-jump location $\Phi_x(T_1)$. The trajectory between the times 0 and T_1 is given by

$$X_t = \begin{cases} \Phi_x(t) & \text{for } 0 \leq t < T_1, \\ Z_1 & \text{for } t = T_1. \end{cases}$$

Now, starting from X_{T_1}, one selects the time $S_2 = T_2 - T_1$ and the post-jump location Z_2 in a similar way as before, and so on. This gives a strong Markov process with the T_k's as the jump times (with $T_0 = 0$). One often considers the embedded Markov chain (Z_n, S_n) associated to the process (X_t) with $Z_n = X_{T_n}$, $S_n = T_n - T_{n-1}$ and $S_0 = 0$. The Z_n's denote the post-jump locations of the process, and the S_n's denote the interarrival times.

In the sequel, we shall consider the discrete-time process (Z_n^-) defined by,

$$\forall n \geq 1,\ Z_n^- = \Phi_{Z_{n-1}}(S_n).$$

This sequence is naturally of interest. Indeed, the transition kernel Q describes the transition from Z_n^- to Z_n. Z_n^- stands for the location of (X_t) just before the nth jump. We shall prove that (Z_n^-) is a Markov chain in Lemma A.1.

Throughout the paper, f and G denote the conditional probability density function and the conditional survival function associated with $\lambda(\Phi_\cdot(\cdot))$. Precisely, for all $z \in \mathbf{R}^d$ and $t \geq 0$,

$$G(z,t) = \exp\left(-\int_0^t \lambda(\Phi_z(s))\mathrm{d}s\right),$$
$$f(z,t) = \lambda(\Phi_z(t))G(z,t).$$

Moreover, $\mathcal{S}$ denotes the conditional distribution of S_{n+1} given Z_n, for all integer n. For all $z \in E$ and $\Gamma \in \mathcal{B}(\mathbf{R}_+)$, we have

$$\begin{aligned}\mathcal{S}(z,\Gamma) &= \mathbf{P}\left(S_{n+1} \in \Gamma | Z_n = z, \sigma(Z_i, S_i \,:\, 0 \le i \le n)\right) \\ &= \int_{\Gamma \cap [0,t^+(z)[} f(z,s)\mathrm{d}s \;+\; \mathbf{1}_\Gamma(t^+(z))G(z,t^+(z)). \end{aligned} \tag{2.1}$$

The first term corresponds to random jumps which occur in a Poisson-like fashion, while the second one is associated with deterministic jumps. The relation between $\mathcal{S}$ and the conditional survival function G is given, for all $z \in \mathbf{R}^d$ and $t \ge 0$, by $G(z,t) = \mathcal{S}(z,]t, +\infty[)$.

3. Main results

Our main objective in this paper is to provide a recursive estimator of the transition density $q(x,y)$ for any $(x,y) \in E^2$. The recursive estimator of $q(x,y)$ that we consider may be written as follows,

$$\widehat{q}_n(x,y) = \frac{\displaystyle\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right)}{\displaystyle\sum_{j=1}^{n+1} \frac{1}{v_j^{d}} K\left(\frac{Z_j^- - x}{v_j}\right)},$$

where $w_j = w_1 j^{-\beta}$, $v_j = v_1 j^{-\alpha}$, with $v_1, w_1, \alpha, \beta > 0$. In addition, K is a kernel function from $\mathbf{R}^d$ to $\mathbf{R}_+$ satisfying the following conditions,

(i) $\operatorname{supp} K \subset B(0_{\mathbf{R}^d}, \delta)$, where $\delta > 0$ and $B(\xi, r)$ stands for the open ball centered at ξ with radius r,
(ii) K is a bounded function.

In particular, $\int_{\mathbf{R}^d} K^2(z)\mathrm{d}z$ is finite. τ^2 denotes this integral in the sequel.

Under the technical conditions given in Assumptions A.2 and A.8, we shall state that the Markov chain (Z_n^-) admits a unique invariant measure π, which has a density p on the interior of the state space (see Cor. A.7). Under these hypotheses, we have a result of consistency.

Theorem 3.1. *Let us choose v_1 and w_1 such that $\max(v_1, w_1)\delta < \operatorname{dist}(x, \partial E)$. If $p(x) > 0$, $\alpha d < 1$ and $8\beta d < 1$, then,*

$$\widehat{q}_n(x,y) \xrightarrow{a.s.} q(x,y),$$

when n goes to infinity.

Proof. The proof is stated at the end of Appendix B. □

Under an additional condition presented in Assumption C.2, we have the following central limit theorem.

Theorem 3.2. *Let us choose v_1 and w_1 such that $\max(v_1, w_1)\delta < \operatorname{dist}(x, \partial E)$. If $p(x) > 0$,*

$$\frac{1}{2+d} < \alpha < \frac{1}{d} \qquad \textit{and} \qquad 2(1-\alpha d) < 4\beta < \min\left(\frac{1}{2d}, \alpha - \frac{1}{2d}\right),$$

then, when n goes to infinity,

$$n^{(1-\alpha d)/2}\big(\widehat{q}_n(x,y) - q(x,y)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{q(x,y)^2\tau^2}{p(x)(1+\alpha d)}\right).$$

Proof. The proof is stated at the end of Appendix C. □

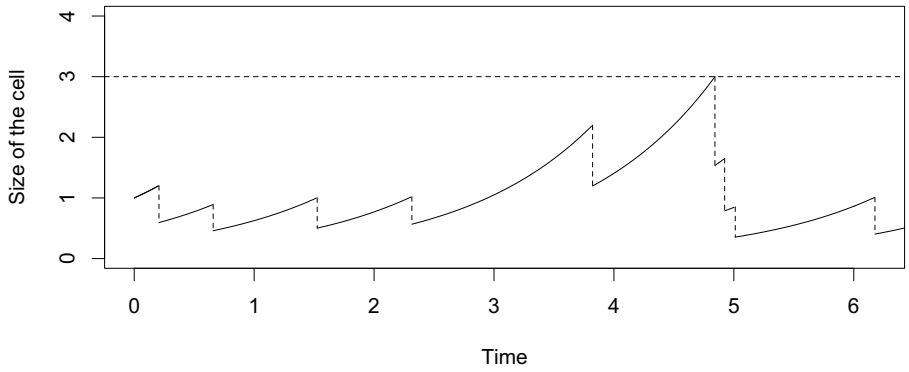

FIGURE 1. Evolution of the size of the cell when 10 jumps occur.

Remark 3.3. For any dimension d, the set of bandwidth parameters α and β satisfying the asymptotic normality presented in Theorem 3.2 is non-empty. Indeed,

$$2(1-\alpha d) < \min\left(\frac{1}{2d}\,,\,\alpha-\frac{1}{2d}\right) \quad \Leftrightarrow \quad \alpha > \alpha_{\min} = \max\left(\frac{1}{d}-\frac{1}{4d^2}\,,\,\frac{1+4d}{d\,(2+4d)}\right),$$

where $\alpha_{\min} d < 1$.

4. Simulation study

The goal of this section is to illustrate the asymptotic behavior of our recursive estimator *via* numerical experiments in the one-dimensional case. More precisely, we investigate numerical simulations for an application of PDMP's in a biological context. Our example deals with the behavior of the size of a cell over time, and is a particular growth-fragmentation model (see [9]).

We consider a continuous-time process (X_t) which models the size of a cell. This one grows exponentially in time, next the cell splits into two offsprings at a division rate λ that depends on its size. We impose that the size can not exceed a certain threshold. In our application, the state space of (X_t) is assumed to be $E =]0,3[$. For any $x \in E$, the deterministic flow $\Phi(x,t)$ is given, for any $t \geq 0$, by

$$\Phi(x,t) = x\exp(\tau t),$$

where $\tau = 0.9$ in the simulations. We assume that the inter-jumping times are distributed according to the Weibull distribution where the shape parameter is the inverse of the size of the cell. More precisely, for $x \in E$ and $t \geq 0$, the conditional density $f(x,t)$ associated with the jump rate λ is given by

$$f(x,t) = \frac{t^{(1-x)/x}\exp\left(-t^{1/x}\right)}{x}.$$

The transition kernel $Q(x,\cdot)$ is chosen to be Gaussian with mean $x/2$, a small variance σ^2 and truncated to $]x/2 \pm \sigma[\cap E$, with $\sigma = 10^{-1}$ in our simulations. A trajectory of such a PDMP is given in Figure 1.

In these numerical experiments, we begin with some investigations of the accuracy of the estimator $\widehat{q}_n(x,y)$, for $(x,y) = (1,0.5)$ and $(x,y) = (2,1)$, from different numbers n of observed jumps. The chosen bandwidth parameters are $v_1 = w_1 = 0.1$, $\alpha = 0.125$ and $\beta = 0.1$, associated with the Epanechnikov kernel. We present in Figure 2 the boxplots of the estimates over 100 replicates. The empirical distributions of the associated relative errors are given in Figure 3. On small-sampled sizes, our procedure is quite unfulfilling. However, for n large enough, our method succeeds in the pointwise estimation of the quantity of interest, especially when n is greater

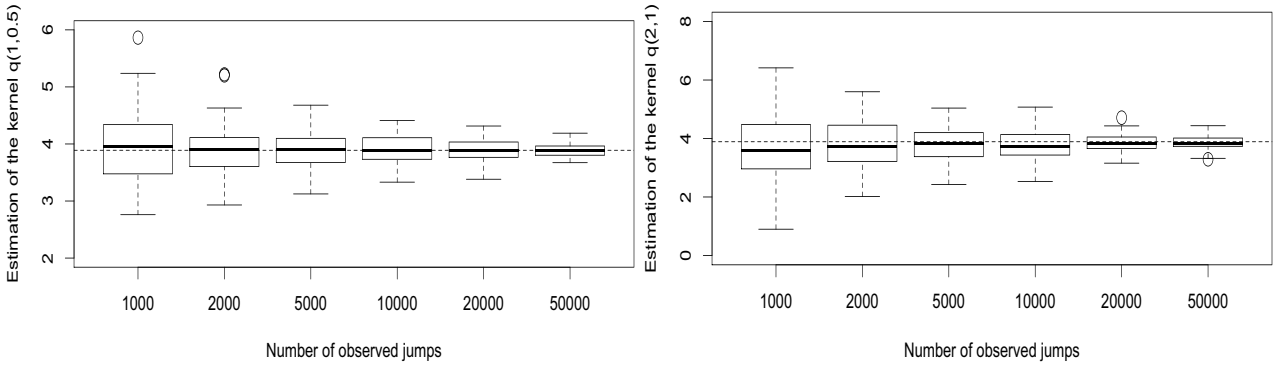

FIGURE 2. Boxplots over 100 replicates of the estimator $\widehat{q}_n(x,y)$ of $q(x,y)$ for $x=1$ and $y=0.5$ (left) and $x=2$ and $y=1$ (right) from different numbers of observed jumps.

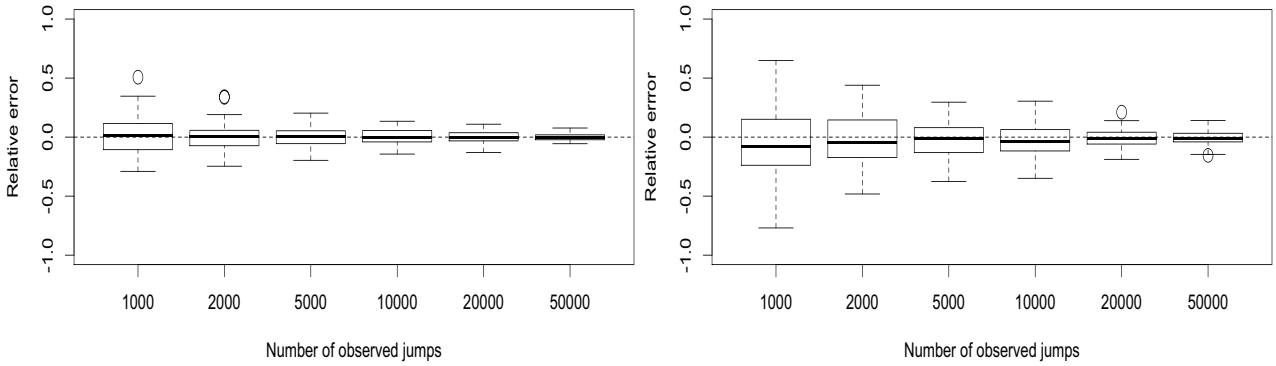

FIGURE 3. Boxplots over 100 replicates of the relative error between $\widehat{q}_n(x,y)$ and $q(x,y)$ for $x=1$ and $y=0.5$ (left) and $x=2$ and $y=1$ (right) from different numbers of observed jumps.

than 10 000. In addition, the complete curves $q(x,\cdot)$, with $x=1$ and $x=2$, and their estimates from different numbers of observed jumps are presented in Figure 4. One may observe the convergence of $\widehat{q}_n$ to the transition kernel of interest q. From $n=50\,000$ observed jumps, the estimation procedure performs very well.

The quality of the estimation is better for $(x,y)=(1,0.5)$ than for $(x,y)=(2,1)$ (see Figs. 3 and 4). Thanks to the estimation of the invariant distribution π (see Fig. 5), one may notice that the Markov chain (Z_n^-) is more often around $x=1$ than $x=2$. As a consequence, the number of data for the estimation of $q(x,y)$ is larger for $x=1$ than for $x=2$. The behavior of the limit distribution π of (Z_n^-) is a good indicator of the accuracy of the estimator $\widehat{q}_n(x,y)$.

We investigate the choice of the bandwidth parameters in Figure 6. We present the boxplots over 100 replicates of the estimates of $q(1,0.5)$ from 10 000 observed jumps, with different values of the parameters α and β. In a theoretical point of view, the almost sure convergence holds when $\alpha<1$. However, we observe that $\alpha=1/8$ and $\alpha=1/4$ are better choices than $\alpha=1/2$. According to Figure 6, a good compromise for small numerical bias and variance is $\alpha=1/8$ and $\beta=0.1$.

It is tedious to determine numerically the bandwidth parameters α and β for which we observe a central limit theorem for $\widehat{q}_n(x,y)$. An illustration of Theorem 3.2 is given in Figure 7 from 50 000 observed jumps, with $\alpha=0.5$ and $\beta=0.1$.

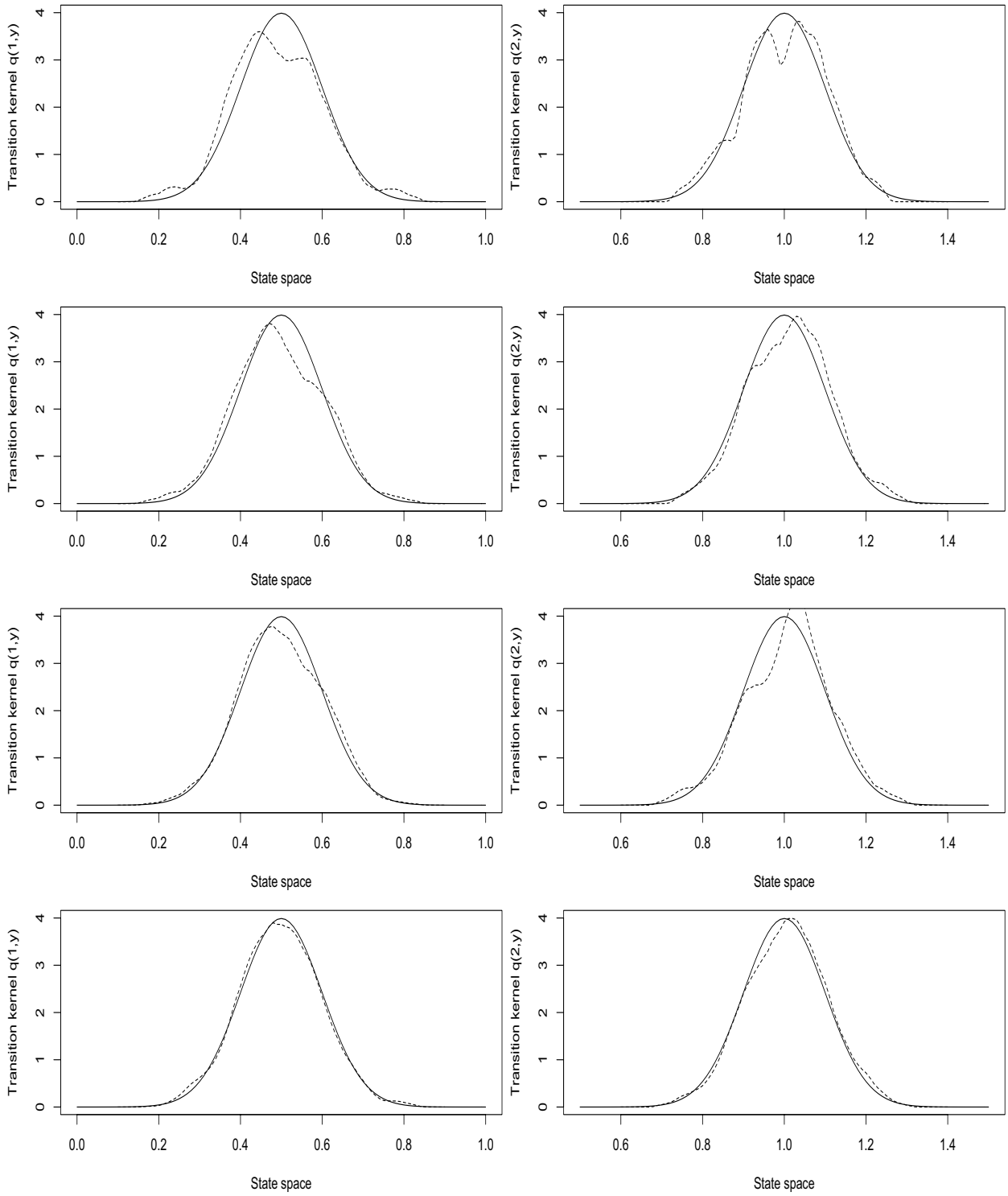

FIGURE 4. Estimation of the transition density $q(x, \cdot)$ for $x = 1$ (left) and $x = 2$ (right) from different numbers of observed jumps: 5 000, 10 000, 20 000 and 50 000 (from top to bottom).

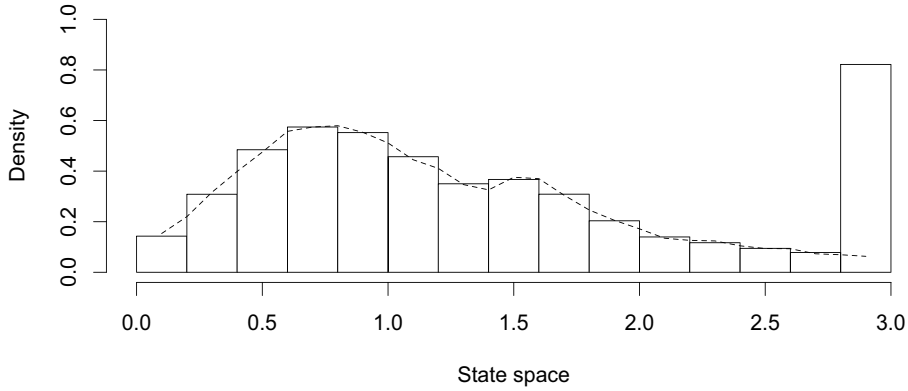

FIGURE 5. Estimation of the invariant distribution π of the Markov chain (Z_n^-). The estimator $\widehat{p}_n(x)$ is drawn in dashed line for $x \in [0.1, 2.9]$. The histogram shows the empirical distribution of (Z_n^-).

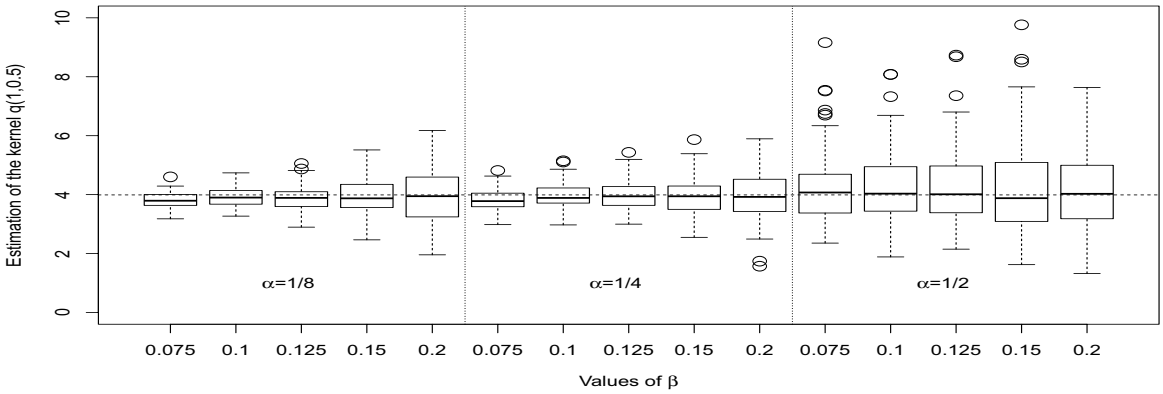

FIGURE 6. Boxplots over 100 replicates of the estimator $\widehat{q}_n(1, 0.5)$ from $n = 10\,000$ observed jumps for different values of the parameters α and β.

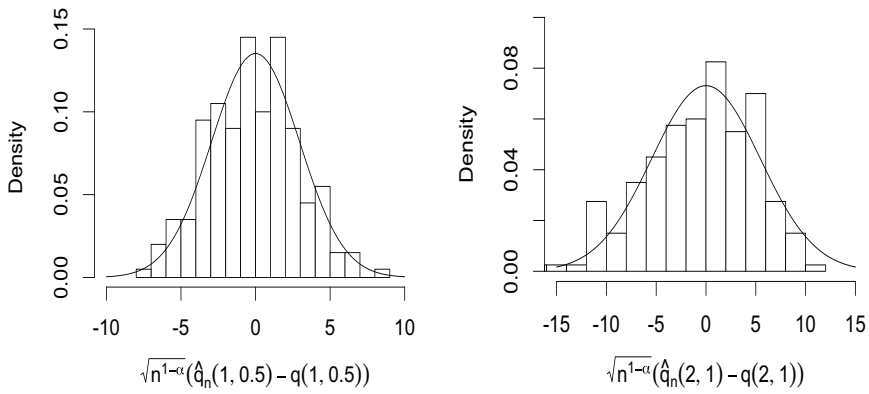

FIGURE 7. Illustration of the central limit theorem for $\widehat{q}_n(x, y)$ for $x = 1$ and $y = 0.5$ (left) and $x = 2$ and $y = 1$ (right) from $50\,000$ observed jumps. The variance of the Gaussian curve has been estimated.

5. Concluding remarks

In this paper, we have presented a recursive nonparametric estimator for the transition kernel of a d-dimensional piecewise-deterministic Markov process observed within a long time interval. A recursive procedure is suitable for lots of applications because it allows to extract information of a dynamical system in real time. The consistency and the asymptotic normality of our estimate have been stated under some general conditions on the main features of the process. In addition, the good asymptotic behavior of the estimator has been illustrated *via* numerical experiments in a biological context.

As seen in Section 3, the estimation needs the observation of the Markov chains (Z_n) and (Z_n^-). If the pre-jump location Z_n^- is not available, one may replace it by $\Phi_{Z_{n-1}}(S_n)$ when the deterministic flow is assumed to be known. If both the pre-jump locations and the deterministic motion are not available, then the estimator can not be computed. In this case, one may focus on the estimation of the conditional distribution of the post-jump location Z_n given the previous one Z_{n-1} and the inter-jumping time S_n.

The optimal choice of the bandwidth parameters could be investigated. Indeed, we have shown *via* numerical simulations for the growth-fragmentation model that the bandwidth parameters have a strong effect on the variance of the estimator for finite sample size (see Fig. 6). In addition, for real data applications, it is attractive to determine the best choice for these parameters. In a theoretical point of view, one may concentrate on minimizing the mean integrated squared error. The influence of the kernel function could also be explored. These points are currently under investigation.

Appendix A. Estimation of the invariant distribution of (Z_n^-)

The main objective of this section is the estimation of the invariant distribution of the Markov chain (Z_n^-). This section is divided into two parts. In the first one, we are interested in the existence and the uniqueness of the invariant distribution of (Z_n^-), and in the properties of its transition kernel $\mathcal{R}$. In the second part, we propose a recursive estimator of the invariant distribution of (Z_n^-) and we investigate its asymptotic behavior.

A.1. Some properties of (Z_n^-)

In this part, we focus on the process (Z_n^-), which is a Markov chain on $\overline{E}$. We especially investigate its transition kernel $\mathcal{R}$ and the existence of an invariant measure.

Lemma A.1. *(Z_n^-) is a Markov chain whose transition kernel $\mathcal{R}$ is given, for all $y \in E$ and $B \in \mathcal{B}(\overline{E})$, by*

$$\mathcal{R}(y, B) = \int_E Q(y, \mathrm{d}z)\mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbf{R}_+), \tag{A.1}$$

where the conditional distribution $\mathcal{S}$ has already been defined by (2.1).

Proof. For all integer n, by (2.1), we have

$$\begin{aligned}
&\mathbf{P}(Z_{n+1}^- \in B | Z_n = z, Z_n^-, \ldots, Z_1^-) \\
&\quad = \mathbf{P}(S_{n+1} \in \Phi_z^{-1}(B) \cap \mathbf{R}_+ | Z_n = z, Z_n^-, \ldots, Z_1^-) \\
&\quad = \mathbf{E}\left[\mathbf{E}\left[\mathbf{1}_{\{S_{n+1} \in \Phi_z^{-1}(B) \cap \mathbf{R}_+\}} \middle| Z_n = z, \sigma(Z_i, S_i : 0 \le i \le n)\right] \middle| Z_n = z, Z_n^-, \ldots, Z_1^-\right] \\
&\quad = \mathbf{E}\left[\mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbf{R}_+) \middle| Z_n = z, Z_n^-, \ldots, Z_1^-\right] \\
&\quad = \mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbf{R}_+).
\end{aligned}$$

As a consequence,

$$\begin{aligned}
\mathbf{P}(Z_{n+1}^- \in B | Z_n^-, \ldots, Z_1^-) &= \int_E \mathbf{P}(Z_{n+1}^- \in B | Z_n = z, Z_n^-, \ldots, Z_1^-) Q(Z_n^-, \mathrm{d}z) \\
&= \int_E \mathcal{S}(z, \Phi_z^{-1}(B) \cap \mathbf{R}_+) Q(Z_n^-, \mathrm{d}z),
\end{aligned}$$

which provides the result. □

We focus on the ergodicity of (Z_n^-) by using Doeblin's assumption.

Assumption A.2. We assume that the transition kernel $\mathcal{R}$ satisfies Doeblin's condition (see [19], p. 396 for instance), that is, there exist a probability measure μ on $(\overline{E}, \mathcal{B}(\overline{E}))$, a real number ε and an integer k such that,

$$\forall y \in \overline{E},\ \forall B \in \mathcal{B}(\overline{E}),\ \mathcal{R}^k(y, B) \geq \varepsilon\, \mu(B). \tag{A.2}$$

Under this assumption, one may state that the Markov chain (Z_n^-) is ergodic.

Proposition A.3. *We have the following results.*

- *The Markov chain (Z_n^-) is μ-irreducible, aperiodic and admits a unique invariant measure, which we denote by π.*
- *There exist $\rho > 1$ and $\kappa > 0$ such that,*

$$\forall n \geq 1,\ \sup_{\xi \in \overline{E}} \|\mathcal{R}^n(\xi, \cdot) - \pi\|_{TV} \leq \kappa \rho^{-n}, \tag{A.3}$$

  where $\|\cdot\|_{TV}$ stands for the total variation norm.
- *In addition, (Z_n^-) is positive Harris-recurrent.*

Proof. By definition and the inequality (A.2), (Z_n^-) is μ-irreducible and aperiodic (see [19], p. 114). Moreover, on the strength of Theorem 16.0.2 of [19], (Z_n^-) admits a unique invariant measure π since it is aperiodic and (A.3) holds. In addition, from Theorem 4.3.3 of [13], (Z_n^-) is positive Harris-recurrent. □

Remark A.4. The ergodicity of the Markov chain (Z_n^-) is often equivalent to the one of the post-jump locations (Z_n). Doeblin's assumption may be related to the existence of a Foster–Lyapunov's function for instance (see [19] for this kind of connection). Furthermore, Assumption A.2 is satisfied for a PDMP defined on a bounded state space, with Gaussian transitions and a strictly positive jump rate.

Now, we shall impose some assumptions on the characteristics relative to the flow of the process. Under these new constraints, one may provide a more useful expression of $\mathcal{R}$. In the sequel, for $\xi \in E$, $t^-(\xi)$ denotes the deterministic exit time from E for the reverse flow,

$$t^-(\xi) = \sup\{t < 0 \,:\, \Phi_x(t) \in \partial E\},$$

with the usual convention $\sup \emptyset = -\infty$. Remark that $t^-(\xi)$ is a negative number.

Assumptions A.5.

(i) The flow Φ is assumed to be $\mathcal{C}^1$-smooth. For any $(z, t) \in \mathbf{R}^d \times \mathbf{R}$, $D\Phi_z(t)$ is defined by

$$D\Phi_z(t) = \left| \det \left(\frac{\partial \Phi_x^{(i)}(t)}{\partial x_j} \right)_{1 \leq i, j \leq d} \right|. \tag{A.4}$$

(ii) For any $t \in \mathbf{R}$, $\varphi_t : \mathbf{R}^d \to \mathbf{R}^d$, defined by $\varphi_t(x) = \Phi_x(t)$, is an injective application.

A useful expression of the transition kernel $\mathcal{R}$ is stated in the following proposition.

Proposition A.6. *Let $y \in E$ and $B \in \mathcal{B}(\overline{E})$. We have*

$$\mathcal{R}(y,B) = \int_{B\cap E} r(y,z)\mathrm{d}z + \mathcal{R}(y, B\cap \partial E), \tag{A.5}$$

where the conditional density function r is given by,

$$\forall z \in E,\ r(y,z) = \int_0^{-t^-(z)} q(y, \Phi_z(-s)) f(\Phi_z(-s), s) D\Phi_z(-s)\mathrm{d}s. \tag{A.6}$$

Proof. Let $B \subset E$. First, we fix $t \geq 0$. We define the set A_t by

$$A_t = \{\Phi_\xi(-t) \,:\, \xi \in B\},$$

and we examine the function $\varphi_t : A_t \to B$ defined by $\varphi_t(x) = \Phi_x(t)$, $x \in A_t$. φ_t is a $\mathcal{C}^1$-smooth injective application (see Assumption A.5). Furthermore, for any $z \in B$,

$$\varphi_t(\Phi_z(-t)) = \Phi_{\Phi_z(-t)}(t) = z,$$

with $\Phi_z(-t) \in A_t$ by definition of A_t. Consequently, φ_t is a $\mathcal{C}^1$-one-to-one correspondence. The inverse function φ_t^{-1} is given by $\varphi_t^{-1}(x) = \Phi_x(-t)$, so it is $\mathcal{C}^1$-smooth too. Thus, φ_t is a $\mathcal{C}^1$-diffeomorphism from A_t into B, which allows us to consider it as a change of variable. In particular, this shows the relation

$$(z \in E,\, t \in \mathbf{R}_+, \Phi_z(t) \in B) \ \Leftrightarrow\ (t \in \mathbf{R}_+,\, z \in E,\, z \in A_t)\,. \tag{A.7}$$

Moreover, the Jacobian matrix $\mathbf{J}_{\varphi_t^{-1}}$ of the inverse function φ_t^{-1} satisfies,

$$\forall x \in \mathbf{R}^d,\ \mathbf{J}_{\varphi_t^{-1}}(x) = \left(\frac{\partial \Phi_x^{(i)}(-t)}{\partial x_j}\right)_{1\leq i,j\leq d}.$$

By (2.1) and (A.1), we have

$$\mathcal{R}(y,B) = \int_E \int_{\Phi_z^{-1}(B)\cap \mathbf{R}_+} q(y,z) f(z,t)\mathrm{d}t.$$

Together with (A.7), we obtain

$$\mathcal{R}(y,B) = \int_{\mathbf{R}_+} \left(\int_{A_t} \mathbf{1}_E(z) f(z,t) q(y,z)\mathrm{d}z\right)\mathrm{d}t.$$

By the change of variable φ_t, we have

$$\begin{aligned}\mathcal{R}(y,B) &= \int_{\mathbf{R}_+} \left(\int_B \mathbf{1}_E\left(\varphi_t^{-1}(\xi)\right) f\left(\varphi_t^{-1}(\xi), t\right) q\left(y, \varphi_t^{-1}(\xi)\right) \left|\det \mathbf{J}_{\varphi_t^{-1}}(\xi)\right| \mathrm{d}\xi\right)\mathrm{d}t \\ &= \int_{\mathbf{R}_+} \left(\int_B \mathbf{1}_E\left(\Phi_\xi(-t)\right) f\left(\Phi_\xi(-t), t\right) q\left(y, \Phi_\xi(-t)\right) D\Phi_\xi(-t)\mathrm{d}\xi\right)\mathrm{d}t,\end{aligned}$$

where $D\Phi_\xi$ is defined by (A.4). We remark that

$$(\Phi_\xi(-t) \in E) \Leftrightarrow (0 \leq t < -t^\star(\xi))\,,$$

so $\mathbf{1}_E(\Phi_\xi(-t)) = \mathbf{1}_{\{0\leq t < -t^\star(\xi)\}}$. By Fubini's theorem, this yields to the expected result. □

In the light of this result, one may obtain the following one about the invariant distribution π of the Markov chain (Z_n^-): π admits a density with respect to the Lebesgue measure in the interior of the state space. In addition, one may exhibit a link between this density and r.

Corollary A.7. *There exists a non-negative function p such that*

$$\forall B \in \mathcal{B}(\overline{E}),\ \pi(B) = \int_{B\cap E} p(x)\mathrm{d}x \ +\ \pi(B \cap \partial E). \tag{A.8}$$

In addition, p is given by the expression,

$$\forall x \in E,\ p(x) = \int_{\overline{E}} \pi(\mathrm{d}y) r(y,x). \tag{A.9}$$

Proof. μ is an irreducibility measure for $\mathcal{R}$. As a consequence and according to Proposition 4.2.2 of [19], the maximal irreducibility measure $\widetilde{\mu}$ is equivalent to the measure $\widetilde{\mu}'$ given for any $B \in \mathcal{B}(\overline{E})$, by

$$\widetilde{\mu}'(B) = \int_{\overline{E}} \mu(\mathrm{d}y) \sum_{n\geq 0} \mathcal{R}^n(y,B) \frac{1}{2^{n+1}}\cdot$$

$\mathcal{R}$ admits a density on the interior E of the state space $\overline{E}$ of the Markov chain (Z_n^-) (see Prop. A.6). As a consequence, this is the case for $\mathcal{R}^n$, too. Indeed, for any set B such that $\lambda_d(B\cap E) = 0$, we have

$$\mathcal{R}^n(y, B\cap E) = \int_{\overline{E}} \mathcal{R}^{n-1}(y,\mathrm{d}z)\mathcal{R}(z, B\cap E) = 0.$$

Therefore, $\widetilde{\mu}'(B\cap E) = 0$. Finally,

$$\lambda_d(B\cap E) = 0 \ \Rightarrow\ \widetilde{\mu}(B\cap E) = 0.$$

Since π and the maximal irreducibility measure $\widetilde{\mu}$ are equivalent, π admits a density on E. Now, we investigate the expression of this density. By Fubini's theorem, we have for any $B \subset E$,

$$\begin{aligned}\pi(B) &= \int_{\overline{E}} \pi(\mathrm{d}y)\mathcal{R}(y,B)\\ &= \int_{\overline{E}} \pi(\mathrm{d}y) \int_B r(y,x)\mathrm{d}x\\ &= \int_B \left(\int_{\overline{E}} \pi(\mathrm{d}y) r(y,x)\right)\mathrm{d}x.\end{aligned}$$

As a consequence, one may identify p with the function $\displaystyle\int_{\overline{E}} \pi(\mathrm{d}y) r(y,\cdot)$. □

One shall see that the regularity of the conditional probability density function r is significant in all the sequel. We state that under additional assumptions, r is Lipschitz.

Assumptions A.8. We assume the following statements.

(i) t^- is a bounded and Lipschitz function, that is,

$$\exists [t^-]_{Lip} > 0,\ \forall x,y \in E,\ \left|t^-(x) - t^-(y)\right| \leq [t^-]_{Lip}|x-y|.$$

(ii) The flow Φ is Lipschitz, that is,

$$\exists [\Phi]_{Lip} > 0,\ \forall x,y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ |\Phi_x(t) - \Phi_y(t)| \leq [\Phi]_{Lip}|x-y|.$$

(iii) f is a bounded and Lipschitz function, that is,

$$\exists [f]_{Lip} > 0,\ \forall x, y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ |f(x,t) - f(y,t)| \leq [f]_{Lip}|x-y|.$$

(iv) q is a bounded and Lipschitz function, that is, there exists $[q]_{Lip} > 0$ such that, for any $x, y, z \in \mathbf{R}^d$,

$$|q(x,y) - q(x,z)| \leq [q]_{Lip}|y-z| \quad \text{and} \quad |q(x,z) - q(y,z)| \leq [q]_{Lip}|x-y|.$$

(v) $D\Phi$ is a bounded and Lipschitz function, that is,

$$\exists [D\Phi]_{Lip} > 0,\ \forall x, y \in \mathbf{R}^d,\ \forall t \in \mathbf{R},\ |D\Phi_x(t) - D\Phi_y(t)| \leq [D\Phi]_{Lip}|x-y|.$$

Proposition A.9. *r is a bounded function. Furthermore, there exists a constant $[r]_{Lip} > 0$ such that, for any $x \in \overline{E}$, $y \in E$ and $u \in \mathbf{R}^d$ such that $y + u \in E$, we have*

$$|r(x, y+u) - r(x,y)| \leq [r]_{Lip}|u|.$$

Proof. First, we have from (A.6),

$$\|r\|_\infty \leq \|t^-\|_\infty \|q\|_\infty \|f\|_\infty \|D\Phi\|_\infty.$$

For the second point, we consider the function γ defined by,

$$\forall (y,t) \in \mathbf{R}^d \times \mathbf{R},\ \gamma(y,t) = q(x, \Phi_y(-t)) f(\Phi_y(-t), t) D\Phi_y(-t).$$

This function is Lipschitz as a compound and product of Lipschitz functions (see Assumption A.8). $[\gamma]_{Lip}$ stands for its Lipschitz constant, and we have

$$|\gamma(y,t) - \gamma(y+u,t)| \leq [\gamma]_{Lip}|u|.$$

In addition, by (A.6), the function $r(x,y)$ is given by

$$r(x,y) = \int_0^{-t^-(y)} \gamma(y,s)\mathrm{d}s.$$

We suppose that $-t^-(y) \leq -t^-(y+u)$ (recall that t^- is a negative function). We have

$$r(x,y+u) - r(x,y) = \int_0^{-t^-(y)} (\gamma(y+u,s) - \gamma(y,s))\,\mathrm{d}s + \int_{-t^-(y)}^{-t^-(y+u)} \gamma(y+u,s)\mathrm{d}s.$$

As a consequence,

$$\begin{aligned} |r(x,y+u) - r(x,y)| &\leq \int_0^{\|t^-\|_\infty} |\gamma(y+u,s) - \gamma(y,s)|\,\mathrm{d}s \\ &\quad + \|q\|_\infty \|f\|_\infty \|D\Phi\|_\infty \left|t^-(y) - t^-(y+u)\right| \\ &\leq \|t^-\|_\infty [\gamma]_{Lip}|u| + [t^-]_{Lip}\|q\|_\infty \|f\|_\infty \|D\Phi\|_\infty |u|. \end{aligned}$$

The obtained inequality for $-t^-(y) > -t^-(y+u)$ is exactly the same one. This achieves the proof. □

A.2. Estimation of p

We propose a recursive nonparametric estimator of the function p given in the Corollary A.7. For all integer n, the recursive estimator $\widehat{p}_n$ of p that we propose is given for all $x \in E$ by

$$\widehat{p}_n(x) = \frac{1}{n}\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right), \tag{A.10}$$

where the bandwith v_j satisfies

$$v_j = v_1 j^{-\alpha}, \text{ with } \alpha > 0.$$

Remark A.10. Let $x \in E$ and $j \geq 1$. Since the sequence (v_n) is decreasing, we have

$$\operatorname{supp} K\left(\frac{\cdot - x}{v_j}\right) \subset \operatorname{supp} K\left(\frac{\cdot - x}{v_1}\right) \subset B(x, v_1\delta).$$

Thus, if $v_1\delta < \operatorname{dist}(x, \partial E)$, we have

$$\operatorname{supp} K\left(\frac{\cdot - x}{v_j}\right) \subset E.$$

In the following proposition, we establish the pointwise asymptotic consistency of $\widehat{p}_n$.

Proposition A.11. *Let $x \in E$. One chooses v_1 such that $v_1\delta < \operatorname{dist}(x, \partial E)$ and α such that $\alpha d < 1$. Then,*

$$\widehat{p}_n(x) \xrightarrow{a.s.} p(x),$$

when n goes to infinity.

Proof. By the expression of $p(x)$ given by (A.9), the difference $\widehat{p}_n(x) - p(x)$ may be written in the following way,

$$\begin{aligned}\widehat{p}_n(x) - p(x) &= \frac{1}{n}\sum_{j=1}^{n+1} \frac{1}{v_j^d} K\left(\frac{Z_j^- - x}{v_j}\right) - \int_E r(u,x)\pi(\mathrm{d}u) \\ &= \frac{1}{nv_1^d} K\left(\frac{Z_1^- - x}{v_1}\right) + \frac{1}{n} M_n + R_n^{(1)} + R_n^{(2)},\end{aligned} \tag{A.11}$$

where M_n, $R_n^{(1)}$ and $R_n^{(2)}$ are given by

$$M_n = \sum_{j=1}^{n}\left[\frac{1}{v_{j+1}^d} K\left(\frac{Z_{j+1}^- - x}{v_{j+1}}\right) - \int_{\mathbf{R}^d} r(Z_j^-, x + yv_{j+1}) K(y)\mathrm{d}y\right], \tag{A.12}$$

$$R_n^{(1)} = \frac{1}{n}\sum_{j=1}^{n} \int_{\mathbf{R}^d} \left[r(Z_j^-, x + yv_{j+1}) - r(Z_j^-, x)\right] K(y)\mathrm{d}y, \tag{A.13}$$

$$R_n^{(2)} = \frac{1}{n}\sum_{j=1}^{n} r(Z_j^-, x) - \int_E r(u,x)\pi(\mathrm{d}u). \tag{A.14}$$

The dependency on x is implicit. In (A.11), the first term clearly tends to 0 as n goes to infinity. The sequel of the proof is divided into three parts: in the first one, we show that $R_n^{(2)}$ tends to 0 by the ergodic theorem. In the second one, we focus on $R_n^{(1)}$ and we prove that this term goes to 0. Finally, we state that M_n/n tends to 0

by using the second law of large numbers for martingales. Recall that the Markov chain (Z_n^-) is positive Harris-recurrent with invariant measure π, according to Proposition A.3. Thus, one may apply the ergodic theorem (see for instance Thm. 17.1.7 of [19]) and we obtain that $R_n^{(2)}$ almost surely tends to 0. For $R_n^{(1)}$, we have

$$\begin{aligned}\left|R_n^{(1)}\right| &\leq \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbf{R}^d}\left|r(Z_j^-,x+yv_{j+1})-r(Z_j^-,x)\right|K(y)\mathrm{d}y\\ &\leq \frac{1}{n}\sum_{j=1}^{n}\int_{\mathbf{R}^d}[r]_{Lip}|y|v_{j+1}K(y)\mathrm{d}y\\ &\leq \frac{1}{n}\sum_{j=1}^{n}v_{j+1}\left(\int_{\mathbf{R}^d}|y|K(y)\mathrm{d}y\right)[r]_{Lip}. \end{aligned}\tag{A.15}$$

This upper bound tends to 0 by Cesaro's lemma because the limit of the sequence (v_n) is 0. Therefore, $R_n^{(1)}$ goes to 0 as n tends to infinity. Finally, we investigate the term M_n/n. First, we show that the process (M_n) is a discrete-time martingale with respect to the filtration $(\mathcal{F}_n)$ defined by,

$$\forall n\geq 1,\ \mathcal{F}_n=\sigma(Z_1^-,\ldots,Z_{n+1}^-).$$

We have

$$\mathbf{E}\left[M_n|\mathcal{F}_{n-1}\right]=M_{n-1}+\mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^--x}{v_{n+1}}\right)\middle|Z_n^-\right]-\int_{\mathbf{R}^d}r(Z_n^-,x+yv_{n+1})K(y)\mathrm{d}y.$$

Thus, we only have to prove that

$$\mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^--x}{v_{n+1}}\right)\middle|Z_n^-\right]=\int_{\mathbf{R}^d}r(Z_n^-,x+yv_{n+1})K(y)\mathrm{d}y.\tag{A.16}$$

We have

$$\mathbf{E}\left[\frac{1}{v_{n+1}^d}K\left(\frac{Z_{n+1}^--x}{v_{n+1}}\right)\middle|Z_n^-\right]=\frac{1}{v_{n+1}^d}\int_{\overline{E}}K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-,\mathrm{d}u).$$

By the assumption on v_1 and Remark A.10,

$$\begin{aligned}\int_{\overline{E}}K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-,\mathrm{d}u)&=\int_{E}K\left(\frac{u-x}{v_{n+1}}\right)\mathcal{R}(Z_n^-,\mathrm{d}u)\\ &=\int_{E}K\left(\frac{u-x}{v_{n+1}}\right)r(Z_n^-,u)\mathrm{d}u,\end{aligned}$$

by (A.5). Finally, the change of variable $u=yv_{n+1}+x$ states (A.16). Thus, (M_n) is a martingale. We shall study the asymptotic behavior of its predictable quadratic variation $\langle M\rangle$. A straightforward calculus leads to

$$\begin{aligned}(M_n-M_{n-1})^2=&\frac{1}{v_{n+1}^{2d}}K^2\left(\frac{Z_{n+1}^--x}{v_{n+1}}\right)+\left[\int_E r(Z_n^-,x+yv_{n+1})K(y)\mathrm{d}y\right]^2\\ &-\frac{2}{v_{n+1}^d}K\left(\frac{Z_{n+1}^--x}{v_{n+1}}\right)\int_E r(Z_n^-,x+yv_{n+1})K(y)\mathrm{d}y.\end{aligned}$$

Using the method used to show (A.16), we deduce that

$$\begin{aligned}\mathbf{E}\left[(M_n-M_{n-1})^2|\mathcal{F}_{n-1}\right]=&\frac{1}{v_{n+1}^d}\int_E K^2(y)r(Z_n^-,x+yv_{n+1})\mathrm{d}y\\ &-\left[\int_E r(Z_n^-,x+yv_{n+1})K(y)\mathrm{d}y\right]^2.\end{aligned}\tag{A.17}$$

As a consequence, there exists a constant $C > 0$ such that

$$\langle M\rangle_n \leq \sum_{j=1}^{n}\left(\frac{1}{v_{j+1}^d}\|r\|_\infty\tau^2 + \|r\|_\infty\right) \sim Cn^{\alpha d+1} \text{ a.s.}$$

when n tends to infinity. By the second law of large numbers for martingales (see Thm. 1.3.15 of [10]), we have

$$M_n^2 = \mathcal{O}\left(\langle M\rangle_n \ln(\langle M\rangle_n)^{1+\gamma}\right) \text{ a.s.}$$

with $\gamma > 0$. As a consequence,

$$\frac{M_n}{n} = \mathcal{O}\left(\sqrt{n^{\alpha d-1}\ln(n^{\alpha d+1})^{1+\gamma}}\right) \text{ a.s.}$$

Thus, M_n/n almost surely tends to 0 as n goes to infinity if $\alpha d < 1$. This achieves the proof. □

Appendix B. Estimation of the invariant distribution of (Z_n^-, Z_n)

In this section, we state that the Markov chain (Z_n^-, Z_n) admits a unique invariant measure. In addition, we are interested in the recursive estimation of this measure. We prove the almost sure convergence of $\widehat{q}_n(x,y)$ given in Theorem 3.1 at the end of this section.

B.1. Some properties of (Z_n^-, Z_n)

In this part, we focus on the asymptotic behavior of the chain (Z_n^-, Z_n). In all the sequel, η_n (respectively π_n) denotes the distribution of (Z_n^-, Z_n) (resp. Z_n^-) for all integer n. We have these straightforward relations between η_n, Q or q and π_n,

$$\begin{aligned}\eta_n(A\times B) &= \int_A Q(z,B)\pi_n(\mathrm{d}z)\\ &= \int_{A\times B} q(z,y)\pi_n(\mathrm{d}z)\mathrm{d}y.\end{aligned} \tag{B.1}$$

Lemma B.1. *We have*

$$\lim_{n\to+\infty}\|\eta_n - \eta\|_{TV} = 0,$$

where the limit distribution η is defined for all $A\times B \in \mathcal{B}(\overline{E}\times E)$ by

$$\eta(A\times B) = \int_{A\times B} q(z,y)\pi(\mathrm{d}z)\mathrm{d}y. \tag{B.2}$$

Proof. Let g be a measurable function bounded by 1. By virtue of Fubini's theorem, we have

$$\left|\int_{\overline{E}\times E} g(x,y)\left(\eta_n(\mathrm{d}x\times\mathrm{d}y) - \eta(\mathrm{d}x\times\mathrm{d}y)\right)\right| \leq \left|\int_{\overline{E}}(\pi_n(\mathrm{d}x) - \pi(\mathrm{d}x))\int_E g(x,y)q(x,y)\mathrm{d}y\right|,$$

from (B.1) and (B.2). Thus,

$$\left|\int_{\overline{E}\times E} g(x,y)\left(\eta_n(\mathrm{d}x\times\mathrm{d}y) - \eta(\mathrm{d}x\times\mathrm{d}y)\right)\right| \leq \left|\int_{\overline{E}}\widetilde{g}(x)(\pi_n(\mathrm{d}x) - \pi(\mathrm{d}x))\right|,$$

where the function $\widetilde{g} : x \mapsto \int_E g(x,y)q(x,y)\mathrm{d}y$ is bounded by 1 since g is bounded by 1 and q is the conditional density associated with the Markov kernel Q. As a consequence,

$$\|\eta_n - \eta\|_{TV} \leq \|\pi_n - \pi\|_{TV}. \tag{B.3}$$

One obtains the expected limit from (A.3). □

In addition, one may prove that η admits a density on $E \times E$.

Lemma B.2. *There exists a positive function h such that*

$$\eta(A \times B) = \int_{A \times B} h(x,y)\mathrm{d}x\,\mathrm{d}y,$$

for any $A \times B \in \mathcal{B}(\overline{E} \times E)$ with $A \subset E$. In addition, h is given for all $x, y \in E$ by

$$h(x,y) = p(x)q(x,y). \tag{B.4}$$

Proof. From (B.2), we have

$$\begin{aligned}\eta(A \times B) &= \int_{A \times B} q(z,y)\pi(\mathrm{d}z)\mathrm{d}y \\ &= \int_{A \times B} q(z,y)p(z)\mathrm{d}z\,\mathrm{d}y,\end{aligned}$$

by (A.8) and because $A \subset E$. This achieves the proof. □

B.2. Estimation of h

We propose to estimate the function h by the recursive nonparametric estimator $\widehat{h}_n$ given, for any $(x,y) \in E^2$, by

$$\widehat{h}_n(x,y) = \frac{1}{n}\sum_{j=1}^{n+1} \frac{1}{w_j^{2d}} K\left(\frac{Z_j^- - x}{w_j}\right) K\left(\frac{Z_j - y}{w_j}\right), \tag{B.5}$$

where the bandwith w_j is given by

$$w_j = w_1 j^{-\beta}, \text{ with } \beta > 0.$$

In the sequel, we are interested in the pointwise convergence of the estimator at a point $(x,y) \in E^2$. We assume that w_1 is such that $w_1\delta < \mathrm{dist}(x, \partial E)$, where δ is the radius of the open ball which contains the support of the kernel function K. In this case, Remark A.10 is still valid, and we have the following inclusions, for any integer j,

$$\mathrm{supp}\, K\left(\frac{\cdot - x}{w_j}\right) \subset B(x, w_1\delta) \subset E. \tag{B.6}$$

Our main objective is to state in Proposition B.7 that $\widehat{h}_n(x,y)$ almost surely converges to $h(x,y)$. First, we show that this estimator is asymptotically unbiased (see Prop. B.4).

We state some new properties of the distribution measures π_n and π. Let us recall that π_n is the law of Z_n^-, while π is the invariant measure of the Markov chain (Z_n^-).

Lemma B.3. *We have the following statements.*

- *For any integer n, π_n admits a density function p_n on E.*
- *p_n is bounded by $\|r\|_\infty$ and is an $[r]_{Lip}$-Lipschitz function.*
- *p is Lipschitz.*
- *For any integer n, we have*

$$\sup_{x \in E} |p_n(x) - p(x)| \leq \|r\|_\infty \kappa \rho^{-(n-1)}. \tag{B.7}$$

Proof. For the first point, let $B \in \mathcal{B}(\overline{E})$ with $B \subset E$. We have

$$\begin{aligned}\pi_n(B) &= \int_{\overline{E}} \int_B \mathcal{R}(\xi, \mathrm{d}y)\pi_{n-1}(\mathrm{d}\xi) \\ &= \int_B \int_{\overline{E}} r(\xi, y)\pi_{n-1}(\mathrm{d}\xi)\mathrm{d}y,\end{aligned}$$

where r is the conditional density associated with the kernel $\mathcal{R}$ (see (A.5)). Thus, one may identify

$$p_n(y) = \int_{\overline{E}} r(\xi, y)\pi_{n-1}(\mathrm{d}\xi). \tag{B.8}$$

For the second assertion, we have stated in Proposition A.9 that r is a bounded function. As a consequence,

$$|p_n(y)| \leq \|r\|_\infty \pi_{n-1}(\overline{E}) = \|r\|_\infty.$$

In addition, since r is Lipschitz,

$$\begin{aligned}|p_n(y) - p_n(z)| &\leq \int_{\overline{E}} |r(\xi, y) - r(\xi, z)|\, \pi_{n-1}(\mathrm{d}\xi) \\ &\leq [r]_{Lip}|y - z|\pi_{n-1}(\overline{E}) = [r]_{Lip}|y - z|.\end{aligned}$$

For the third point, p is Lipschitz for the same reason than p_n since p satisfies (A.9). Finally, for the last point, we have, by (A.9) and (B.8),

$$\begin{aligned}|p_n(x) - p(x)| &\leq \int_{\overline{E}} r(y, x)\, |\pi_{n-1}(\mathrm{d}y) - \pi(\mathrm{d}y)| \\ &\leq \|r\|_\infty \|\pi_{n-1} - \pi\|_{TV} \\ &\leq \|r\|_\infty \kappa \rho^{-(n-1)},\end{aligned}$$

by (A.3). This achieves the proof. □

Now, one may state that $\widehat{h}_n(x, y)$ is an asymptotically unbiased estimator of $h(x, y)$.

Proposition B.4. *When n goes to infinity,*

$$\mathbf{E}\left[\widehat{h}_n(x, y)\right] \to h(x, y).$$

Proof. We only state that

$$\mathbf{E}\left[\widehat{h}_n(x, y)\right] - \frac{n+1}{n} h(x, y)$$

tends to 0. We have

$$\begin{aligned}\mathbf{E}\left[\widehat{h}_n(x, y)\right] - \frac{n+1}{n} h(x, y) = \frac{1}{n}\sum_{j=1}^{n+1} \Bigg[&\int_{\overline{E}\times E} \frac{1}{w_j^{2d}} K\left(\frac{u - x}{w_j}\right) K\left(\frac{v - y}{w_j}\right) \pi_j(\mathrm{d}u) q(u, v)\mathrm{d}v \\ &- \int_{E\times E} h(x, y)K(u)K(v)\mathrm{d}u\mathrm{d}v\Bigg].\end{aligned}$$

Thanks to (B.6), one may replace $\overline{E}$ by E in the first integral. As a consequence, one may replace $\pi_j(\mathrm{d}u)$ by $p_j(u)\mathrm{d}u$ (see Lem. B.3). Together with (B.4) and a change of variables, we obtain

$$\begin{aligned}\mathbf{E}\left[\widehat{h}_n(x, y)\right] - \frac{n+1}{n} h(x, y) = \frac{1}{n}\sum_{j=1}^{n+1} &\int_{E\times E} K(u)K(v) \\ &\times \Big(p_j(x + uw_j)q(x + uw_j, y + vw_j) - p(x)q(x, y)\Big)\mathrm{d}u\, \mathrm{d}v.\end{aligned}$$

Furthermore, since p_j is $[r]_{Lip}$-Lipschitz and bounded by $\|r\|_\infty$ in the light of Lemma B.3, an elementary calculus leads to

$$\begin{aligned}\big|p_j(x+uw_j)q(x+uw_j,y+vw_j)-p(x)q(x,y)\big| \\ &\le \|r\|_\infty[q]_{Lip}w_j\left(|u|+|v|\right)+\|q\|_\infty[r]_{Lip}|u|w_j+\|q\|_\infty\left|p_j(x)-p(x)\right| \\ &\le w_j\Big(|u|\big(\|r\|_\infty[q]_{Lip}+\|q\|_\infty[r]_{Lip}\big)+|v|\|r\|_\infty[q]_{Lip}\Big)+\|q\|_\infty\|r\|_\infty\kappa\rho^{-(j-1)},\end{aligned}$$

together with (B.7). Finally, we obtain

$$\begin{aligned}\left|\mathbf{E}\left[\widehat{h}_n(x,y)\right]-\frac{n+1}{n}h(x,y)\right| &\le \frac{(2\|r\|_\infty[q]_{Lip}+\|q\|_\infty[r]_{Lip})\int_E K(u)|u|\mathrm{d}u}{n}\sum_{j=1}^{n+1}w_j+\frac{\|r\|_\infty\|q\|_\infty\kappa}{n}\sum_{j=1}^{n+1}\rho^{-(j-1)} \\ &\le \frac{(2\|r\|_\infty[q]_{Lip}+\|q\|_\infty[r]_{Lip})\int_E K(u)|u|\mathrm{d}u}{n}\sum_{j=1}^{n+1}w_j+\frac{\|r\|_\infty\|q\|_\infty\kappa}{n(1-\rho^{-1})},\end{aligned} \tag{B.9}$$

which tends to 0 by Cesaro's lemma. □

In the following, we are interested in some properties of the discrete-time process

$$(A_n)=\left(\frac{1}{w_n^{2d}}K\left(\frac{Z_n^- - x}{w_n}\right)K\left(\frac{Z_n-y}{w_n}\right)\right), \tag{B.10}$$

which naturally appears in the study of the estimator $\widehat{h}_n(x,y)$. In particular, we propose to investigate its autocovariance function. On the strength of this result, we will establish the asymptotic behavior of the variance of $\widehat{h}_n(x,y)$.

Proposition B.5. *There exist two constants B and $b>1$ such that, for any integers $n\ge k$,*

$$\left|\mathbf{C}\mathrm{ov}(A_k,A_n)\right|\le\frac{\|K\|_\infty^4 B}{w_n^{4d}}b^{k-n}\left(1+b^{-k}\right). \tag{B.11}$$

In particular, one obtains by taking $k=n$ and using that $b^{-n}\le 1$,

$$\mathbf{V}\mathrm{ar}(A_n)\le\frac{2\|K\|_\infty^4 B}{w_n^{4d}}. \tag{B.12}$$

Proof. We have

$$\mathbf{C}\mathrm{ov}(A_k,A_n)=\frac{\|K\|_\infty^4}{w_n^{2d}w_k^{2d}}\mathbf{C}\mathrm{ov}\left(\frac{1}{\|K\|_\infty^2}K\left(\frac{Z_k^- - x}{w_k}\right)K\left(\frac{Z_k-y}{w_k}\right),\frac{1}{\|K\|_\infty^2}K\left(\frac{Z_n^- - x}{w_n}\right)K\left(\frac{Z_n-y}{w_n}\right)\right),$$

where both the components in the covariance are bounded by 1. We apply Theorem 16.1.5 of [19] with $V=1$, $\Phi=(Z^-,Z)$,

$$g=\frac{1}{\|K\|_\infty^2}K\left(\frac{\cdot-x}{w_n}\right)K\left(\frac{\cdot-y}{w_n}\right)\quad\text{and}\quad h=\frac{1}{\|K\|_\infty^2}K\left(\frac{\cdot-x}{w_k}\right)K\left(\frac{\cdot-y}{w_k}\right).$$

The conditions of the theorem are satisfied by (A.3) and (B.3). We obtain

$$|\mathbf{C}\mathrm{ov}(A_k,A_n)|\le\frac{\|K\|_\infty^4}{w_n^{2d}w_k^{2d}}Bb^{k-n}\left(1+b^{-k}\right).$$

Together with $1/w_k^{2d}\le 1/w_n^{2d}$, this shows (B.11). □

In the following result, we give a bound for the variance of $\widehat{h}_n(x,y)$. It is a corollary of Proposition B.5.

Corollary B.6. *Let n be an integer. We have*

$$\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) \leq \frac{8}{nw_{n+1}^{4d}} \frac{\|K\|_\infty^4 B}{1-b^{-1}}. \tag{B.13}$$

As a consequence, this variance goes to 0 when $4\beta d < 1$ (recall that $w_n = w_1 n^{-\beta}$).

Proof. This inequality is a consequence of (B.11) stated in Proposition B.5. Indeed, in light of the expressions of A_n (B.10) and $\widehat{h}_n(x,y)$ (B.5), we have

$$\begin{aligned}\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) &= \frac{2}{n^2}\sum_{k=1}^{n+1}\sum_{l=k}^{n+1}\mathbf{C}\mathrm{ov}(A_l, A_k)\\ &\leq \frac{2}{n^2}\|K\|_\infty^4 B \sum_{k=1}^{n+1}\sum_{l=k}^{n+1} b^{k-l}\left(1+b^{-k}\right) w_l^{-4d}.\end{aligned}$$

Using that $b^{-k} \leq 1$ and $w_l^{-4d} \leq w_{n+1}^{-4d}$,

$$\begin{aligned}\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right) &\leq \frac{4}{n^2 w_{n+1}^{4d}}\|K\|_\infty^4 B \sum_{k=1}^{n+1} b^k \sum_{l=k}^{n+1} b^{-l}\\ &\leq \frac{4}{n^2 w_{n+1}^{4d}}\|K\|_\infty^4\ B \sum_{k=1}^{n+1} b^k \frac{b^{-k}}{1-b^{-1}}\\ &\leq \frac{8}{nw_{n+1}^{4d}}\frac{\|K\|_\infty^4 B}{1-b^{-1}},\end{aligned}$$

with $(n+1)/n \leq 2$. □

Now, one may state the consistency of our estimator of $h(x,y)$.

Proposition B.7. *Let $(x,y) \in E^2$. One chooses $w_1\delta < \mathrm{dist}(x, \partial E)$ and $8\beta d < 1$. Then,*

$$\widehat{h}_n(x,y) \xrightarrow{\text{a.s.}} h(x,y),$$

when n goes to infinity.

Proof. According to Proposition B.4, we only have to prove that

$$Y_n = \left(\widehat{h}_n(x,y) - \mathbf{E}\left[\widehat{h}_n(x,y)\right]\right)^2 \tag{B.14}$$

almost surely converges to 0. In the sequel of the proof, we establish that there exists a random variable Y such that $Y_n \xrightarrow{a.s.} Y$. Since the sequence (Y_n) tends to 0 in $\mathbf{L}^1$ (remark that $\mathbf{E}[Y_n] = \mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right)$, together with Corollary B.6), $Y = 0$ a.s. and it induces the expected result. In order to show the almost sure convergence of the sequence (Y_n), we use Van Ryzin's lemma (see [22]). In light of this result, if the sequence (Y_n) satisfies the following conditions,

(i) $Y_n \geq 0$ a.s.,
(ii) $\mathbf{E}[Y_1] < +\infty$,
(iii) $\mathbf{E}\left[Y_{n+1}|\mathcal{S}_n\right] \leq Y_n + Y_n'$ a.s., where $\mathcal{S}_n = \sigma(Y_1, \ldots, Y_n)$ and Y_n' is $\mathcal{S}_n$-measurable,
(iv) $\sum_{n\geq 1}\mathbf{E}\left[|Y_n'|\right] < +\infty$,

then $Y_n \xrightarrow{a.s.} Y$. In our context, points (i) and (ii) are unquestionably satisfied. Let us define

$$V_n = \frac{1}{n}\sum_{k=1}^{n+1}\big(A_k - \mathbf{E}[A_k]\big). \tag{B.15}$$

By (B.5), (B.10) and (B.14), $V_n^2 = Y_n$ and we have the recurrence relation,

$$V_n = \frac{A_{n+1} - \mathbf{E}[A_{n+1}] + (n-1)V_{n-1}}{n}.$$

By squaring, we obtain

$$\begin{aligned} Y_n &= \left(\frac{n-1}{n}\right)^2 Y_{n-1} + \frac{\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)V_{n-1}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)}{n^2} \\ &\le Y_{n-1} + \frac{\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)V_{n-1}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)}{n^2}. \end{aligned}$$

Finally, $\mathbf{E}[Y_n|\mathcal{S}_{n-1}] \le Y_{n-1} + Y'_{n-1}$, where

$$Y'_{n-1} = \frac{1}{n^2}\mathbf{E}\left[\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)^2 + 2(n-1)V_{n-1}\big(A_{n+1} - \mathbf{E}[A_{n+1}]\big)\Big|\mathcal{S}_{n-1}\right]. \tag{B.16}$$

Thus, (iii) is checked. Ultimately, we have to verify (iv). By (B.16) and Cauchy–Schwarz inequality,

$$\begin{aligned} \mathbf{E}\left[|Y'_{n-1}|\right] &\le \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2(n-1)}{n^2}\mathbf{E}\left[\big|V_{n-1}\big|\,\big|A_{n+1} - \mathbf{E}[A_{n+1}]\big|\right] \\ &\le \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2}{n}\sqrt{\mathbf{E}[Y_{n-1}]\mathbf{V}\mathrm{ar}(A_{n+1})} \\ &\le \frac{1}{n^2}\mathbf{V}\mathrm{ar}(A_{n+1}) + \frac{2}{n}\sqrt{\mathbf{V}\mathrm{ar}\left(\widehat{h}_{n-1}(x,y)\right)\mathbf{V}\mathrm{ar}(A_{n+1})}. \end{aligned}$$

Thus, by (B.12) and (B.13), there exist two numbers c_1 and c_2 such that

$$\begin{aligned} \mathbf{E}\left[|Y'_{n-1}|\right] &\le \frac{c_1}{n^2 w_n^{4d}} + \frac{c_2}{n^{3/2} w_n^{4d}} \\ &\le \frac{c_1}{w_1 n^{2-4\beta d}} + \frac{c_2}{w_1 n^{3/2-4\beta d}}. \end{aligned}$$

As a consequence, $\sum \mathbf{E}[|Y'_n|]$ is a convergent series for $8\beta d < 1$. □

Now, we give the proof of the almost sure convergence of the estimator $\widehat{q}_n(x,y)$.

Proof of Theorem 3.1. In light of (B.4), if $p(x) > 0$, one may write $q(x,y) = h(x,y)/p(x)$. In addition, $\widehat{q}_n(x,y)$ is defined by the ratio $\widehat{h}_n(x,y)/\widehat{p}_n(x)$ (see (B.5) for the expression of $\widehat{h}_n(x,y)$ and (A.10) for the one of $\widehat{p}_n(x)$), where $\widehat{h}_n(x,y)$ (respectively $\widehat{p}_n(x)$) estimates $h(x,y)$ (resp. $p(x)$). In such a case, the result is a corollary of Propositions A.11 and B.7.

Appendix C. Central limit theorem

In this section, we establish a central limit theorem for $\widehat{q}_n(x,y)$ for any $(x,y) \in E^2$ (see Thm. 3.2). First, we derive the rate of convergence of the recursive estimator $\widehat{h}_n(x,y)$ of $h(x,y)$ under some conditions on the parameter β.

Proposition C.1. *Let us choose v_1 and w_1 such that* $\max(v_1, w_1)\delta < \mathrm{dist}(x, \partial E)$. *If*

$$2(1-\alpha d) < 4\beta < \min\left(\frac{1}{2d}, \alpha - \frac{1}{2d}\right),$$

then, when n goes to infinity,

$$n^{(1-\alpha d)/2}\left(\widehat{h}_n(x,y) - h(x,y)\right) \xrightarrow{a.s.} 0.$$

Proof. We take back the definitions (B.14) of Y_n and (B.15) of V_n used in the proof of Proposition B.7. Recall that we have $V_n^2 = Y_n$. Together with (B.5) and (B.10), we have

$$n^{(1-\alpha d)/2}\left(\widehat{h}_n(x,y) - h(x,y)\right) = n^{(1-\alpha d)/2}V_n + n^{(1-\alpha d)/2}\left(\mathbf{E}\left[\widehat{h}_n(x,y)\right] - h(x,y)\right).$$

By (B.9) and $w_j = w_1 j^{-\beta}$,

$$n^{(1-\alpha d)/2}\left(\mathbf{E}\left[\widehat{h}_n(x,y)\right] - h(x,y)\right) = \mathcal{O}\left(n^{1/2-\beta-\alpha d/2}\right).$$

As a consequence, the bias term

$$n^{(1-\alpha d)/2}\left(\mathbf{E}\left[\widehat{h}_n(x,y)\right] - h(x,y)\right)$$

tends to 0 when $2\beta > 1 - \alpha d$. Now, we only have to prove that $W_n = n^{1-\alpha d}Y_n$ almost surely tends to 0. By virtue of Van Ryzin's lemma and a calculus similar to the one implemented in the proof of Proposition B.7, the sequence (W_n) almost surely tends to a random variable W for β such that $4\beta d < \alpha d - 1/2$. As aforementioned in the proof of Proposition B.7,

$$\mathbf{E}[W_n] = n^{1-\alpha d}\mathbf{E}[Y_n] = n^{1-\alpha d}\mathbf{V}\mathrm{ar}\left(\widehat{h}_n(x,y)\right).$$

Thus, in light of Corollary B.6, (W_n) converges to 0 in $\mathbf{L}^1$ when $4\beta < \alpha$. Consequently, $W = 0$ a.s. This shows the expected result. □

Under an additional assumption presented in the sequel, we give a central limit theorem for $\widehat{p}_n(x)$.

Assumption C.2. For any α such that $1/(2+d) < \alpha < 1/d$, we assume that, when n goes to infinity,

$$\frac{1}{n^{(1+\alpha d)/2}}\left|\sum_{j=1}^{n}\left(r(Z_j^-, x) - p(x)\right)\right| \xrightarrow{a.s.} 0.$$

This kind of condition is satisfied for iterative models with some Lipschitz mixing properties (see [10], p. 234 for instance). We have the following central limit theorem for $\widehat{p}_n(x)$.

Proposition C.3. *Let us choose v_1 such that $v_1\delta < \mathrm{dist}(x, \partial E)$. If $1/(2+d) < \alpha < 1/d$ and $p(x) > 0$, then,*

$$n^{(1-\alpha d)/2}\left(\widehat{p}_n(x) - p(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\tau^2 p(x)}{1+\alpha d}\right),$$

when n goes to infinity, where $\tau^2 = \int_{\mathbf{R}^d} K^2(y)\mathrm{d}y$.

Proof. By (A.11), we have

$$n^{(1-\alpha d)/2}\big(\widehat{p}_n(x)-p(x)\big) = \frac{1}{n^{(1+\alpha d)/2}v_1^d}K\left(\frac{Z_1^- - x}{v_1}\right) + \frac{M_n}{n^{(1+\alpha d)/2}} + n^{(1-\alpha d)/2}R_n^{(1)} + n^{(1-\alpha d)/2}R_n^{(2)},$$

where M_n, $R_n^{(1)}$ and $R_n^{(2)}$ are defined by (A.12), (A.13) and (A.14). It is obvious that the first term almost surely tends to 0 since K is a bounded function. In addition, by (A.15) together with $v_j = v_1 j^{-\alpha}$,

$$n^{(1-\alpha d)/2}\left|R_n^{(1)}\right| = \mathcal{O}\left(n^{\frac{1-\alpha d-2\alpha}{2}}\right) \text{ a.s.}$$

As a consequence, $n^{(1-\alpha d)/2}R_n^{(1)}$ almost surely tends to 0 when $1 < \alpha(d+2)$. The term $n^{(1-\alpha d)/2}R_n^{(2)}$ almost surely converges to 0 under Assumption C.2 with (A.9). By (A.17) and the almost sure ergodic theorem, we have

$$\frac{(1+\alpha d)\langle M\rangle_n}{n^{1+\alpha d}} \xrightarrow{a.s.} \tau^2\int_{\overline{E}} r(\xi,x)\pi(\mathrm{d}\xi) \;=\; \tau^2 p(x).$$

As a consequence, by virtue of the central limit theorem for martingales (see for instance Cor. 2.1.10 of [10]), we have

$$\frac{M_n}{n^{(1+\alpha d)/2}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\tau^2 p(x)}{1+\alpha d}\right),$$

when n goes to infinity, if Lindeberg's condition is satisfied. Now, we only have to verify this technical condition in order to end the proof. By (A.12) together with $v_j = v_1 j^{-\alpha}$, we have

$$|M_j - M_{j-1}| \;\le\; \frac{\|K\|_\infty}{v_{j+1}^d} + \|r\|_\infty \;=\; \mathcal{O}\left(j^{\alpha d}\right) \text{ a.s.}$$

Thus, there exists a constant C such that we have the following inclusions, for $1 \le j \le n$,

$$\left\{|M_j - M_{j-1}| \ge \varepsilon n^{(1+\alpha d)/2}\right\} \subset \left\{Cj^{\alpha d} \ge \varepsilon n^{(1+\alpha d)/2}\right\} \subset \left\{Cn^{\alpha d} \ge \varepsilon n^{(1+\alpha d)/2}\right\},$$

for any $\varepsilon > 0$. Consequently, as $n^{(\alpha d-1)/2}$ tends to 0 when $\alpha d < 1$, we have

$$\mathbf{1}_{\left\{|M_j - M_{j-1}| \ge \varepsilon n^{(1+\alpha d)/2}\right\}} = 0 \text{ a.s.}$$

for n large enough. Finally, when n goes to infinity,

$$\frac{1}{n^{1+\alpha d}}\sum_{j=1}^n \mathbf{E}\left[(M_j - M_{j-1})^2\mathbf{1}_{\left\{|M_j - M_{j-1}| \ge \varepsilon n^{(1+\alpha d)/2}\right\}}\middle|\mathcal{F}_j\right] \xrightarrow{\text{a.s.}} 0.$$

This shows Lindeberg's condition and thus the expected result. □

Now, one may give the proof of the central limit theorem for $\widehat{q}_n(x,y)$, presented in Theorem 3.2, under some assumptions on the bandwidth parameters α and β.

Proof of Theorem 3.2. We have

$$n^{(1-\alpha d)/2}\left(\widehat{q}_n(x,y) - q(x,y)\right) = n^{(1-\alpha d)/2}\left(\frac{\widehat{h}_n(x,y)}{\widehat{p}_n(x)} - \frac{h(x,y)}{p(x)}\right) = \frac{n^{(1-\alpha d)/2}\big(\widehat{h}_n(x,y) - h(x,y)\big)}{\widehat{p}_n(x)} + \frac{q(x,y)n^{(1-\alpha d)/2}\big(\widehat{p}_n(x) - p(x)\big)}{\widehat{p}_n(x)}.$$

In light of Propositions C.3 and A.11 and Slutsky's theorem, we have

$$\frac{q(x,y)n^{(1-\alpha d)/2}\left(\widehat{p}_n(x)-p(x)\right)}{\widehat{p}_n(x)} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{q(x,y)^2\tau^2}{p(x)(1+\alpha d)}\right),$$

when n goes to infinity. By virtue of Proposition C.1, we obtain the expected result.

Acknowledgements. The author would like to thank the Associate Editor, two anonymous Reviewers and his Ph.D. advisors, François Dufour and Anne Gégout-Petit, for their suggestions and helpful comments.

References

[1] T. Aven and U. Jensen, Stochastic models in reliability, vol. 41 of *Applications of Mathematics*. Springer-Verlag, New York (1999).

[2] R. Azaïs, F. Dufour and A. Gégout-Petit, Nonparametric estimation of the conditional density of the inter-jumping times for piecewise-deterministic Markov processes. Preprint `arXiv:1202.2212v2` (2012).

[3] A.K. Basu and D.K. Sahoo, On Berry-Esseen theorem for nonparametric density estimation in Markov sequences. *Bull. Inform. Cybernet.* **30** (1998) 25–39.

[4] D. Chafaï, F. Malrieu and K. Paroux, On the long time behavior of the TCP window size process. *Stoch. Process. Appl.* **120** (2010) 1518–1534.

[5] J. Chiquet and N. Limnios, A method to compute the transition function of a piecewise deterministic Markov process with application to reliability. *Statist. Probab. Lett.* **78** (2008) 1397–1403.

[6] S.J.M. Clémençon, Adaptive estimation of the transition density of a regular Markov chain. *Math. Methods Statist.* **9** (2000) 323–357.

[7] M.H.A. Davis, Markov models and optimization, vol. 49 of *Monogr. Statist. Appl. Probab.* Chapman & Hall, London (1993).

[8] P. Doukhan and M. Ghindès, Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1. *Stoch. Process. Appl.* **15** (1983) 271–293.

[9] M. Doumic, M. Hoffmann, N. Krell and L. Robert, Statistical estimation of a growth-fragmentation model observed on a genealogical tree. Preprint (2012).

[10] M. Duflo, Random iterative models. *Appl. Math.* Springer-Verlag, Berlin (1997).

[11] A. Genadot and M. Thieullen, Averaging for a fully coupled Piecewise Deterministic Markov Process in Infinite Dimensions. *Adv. Appl. Probab.* **44** 3 (2012).

[12] O. Hernández-Lerma, S.O. Esparza and B.S. Duran, Recursive nonparametric estimation of nonstationary Markov processes. *Bol. Soc. Mat. Mexicana* **33** (1988) 57–69.

[13] O. Hernández-Lerma and J.B. Lasserre, Markov chains and invariant probabilities, vol. 211 of *Progr. Math.* Birkhäuser Verlag, Basel (2003).

[14] J. Hu, W. Wu and S. Sastry, *Modeling subtilin production in* bacillus subtilis *using stochastic hybrid systems*. Hybrid Systems: Computation and Control. Edited by R. Alur and G.J. Pappas. *Lect. Notes Comput. Sci.* Springer-Verlag, Berlin (2004).

[15] C. Lacour, Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. Henri Poincaré, Probab. Statist.* **43** (2007) 571–597.

[16] C. Lacour, Nonparametric estimation of the stationary density and the transition density of a Markov chain. *Stoch. Process. Appl.* **118** (2008) 232–260.

[17] E. Liebscher, Density estimation for Markov chains. *Statistics* **23** (1992) 27–48.

[18] E. Masry and L. Györfi, Strong consistency and rates for recursive probability density estimators of stationary processes. *J. Multivariate Anal.* **22** (1987) 79–93.

[19] S. Meyn and R.L. Tweedie, Markov chains and stochastic stability, second edition. Cambridge University Press, Cambridge (2009).

[20] M. Rosenblatt, Density estimates and Markov sequences. In *Nonparametric Techniques in Statistical Inference. Proc. of Sympos., Indiana Univ., Bloomington, Ind., 1969*. Cambridge Univ. Press, London (1970), 199–213.

[21] G.G. Roussas, Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21** (1969) 73–87.

[22] J. Van Ryzin, On strong consistency of density estimates. *Ann. Math. Statist.* **40** (1969) 1765–1772.

[23] S. Yakowitz, Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.* **30** (1989) 124–136.