



**HAL**  
open science

# Nonparametric estimation of jump rates for a specific class of Piecewise Deterministic Markov Processes

Nathalie Krell, Emeline Schmisser

► **To cite this version:**

Nathalie Krell, Emeline Schmisser. Nonparametric estimation of jump rates for a specific class of Piecewise Deterministic Markov Processes. *Bernoulli*, 2021, 27 (4), pp.2362-2388. 10.3150/20-BEJ1312 . hal-01996064v2

**HAL Id: hal-01996064**

**<https://hal.science/hal-01996064v2>**

Submitted on 8 Dec 2020

**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.

# Nonparametric estimation of jump rates for a specific class of piecewise deterministic Markov processes

N. Krell\*

E. Schmisser†

December 8, 2020

## Abstract

In this paper, we consider a unidimensional piecewise deterministic Markov process (PDMP), with homogeneous jump rate  $\lambda(x)$ . This process is observed continuously, so the flow  $\phi$  is known. To estimate nonparametrically the jump rate, we first construct an adaptive estimator of the stationary density, then we derive a quotient estimator  $\hat{\lambda}_n$  of  $\lambda$ . Under some ergodicity conditions, we bound the risk of these estimators (and give a uniform bound on a small class of functions), and prove that the estimator of the jump rate is nearly minimax (up to a  $\ln^2(n)$  factor). The simulations illustrate our theoretical results.

*Keywords:* Piecewise deterministic Markov processes, model selection, nonparametric estimation

*Mathematical Subject Classification:* 62G05, 62G07, 62M05, 60J25

## 1 Introduction

Piecewise deterministic Markov processes are a large class of continuous-time stochastic models first introduced by Davis [13]. They are used to model deterministic phenomena in which randomness appears as point events. They are not diffusions, which adds complexity to their study. This family of stochastic processes is well adapted to model various problems in biology (see for instance Cloez et al. [10], Rudnicki and Tyran-Kamińska [29]), neuroscience (Höpfner et al [22], Renault et al [28]), physics (Blanchard and Jadczyk [9]), reliability

---

\*Université de Rennes 1, Institut de Recherche mathématique de Rennes, CNRS-UMR 6625, Campus de Beaulieu. Bâtiment 22, 35042 Rennes Cedex, France. **email:** nathalie.krell@univ-rennes1.fr

†Laboratoire Paul Painlevé Université des Sciences et Technologies de Lille, Bureau 314, Bâtiment M3, Cité Scientifique, 59 655 Villeneuve d'Ascq Cedex **email:** emeline.schmisser@math.univ-lille1.fr

(De Saporta et al. [14]), optimal consumption and exploration (Farid and David [18]), risk insurance, seismology, . . . See also the references in the survey Azaïs et al. [4].

In this article, we consider a filtered piecewise deterministic Markov process (PDMP)  $(X_t)_{t \geq 0}$  taking values in  $\mathbb{R}^+$ , with flow  $\phi$ , transition measure  $Q(x, dy)$  and homogeneous jump rate  $\lambda(x)$ . Starting from initial value  $x_0$ , the process follows the flow  $\phi$  until the first jump time  $T_1$  which occurs spontaneously in a Poisson-like fashion with rate  $\lambda(\phi(x, t))$ . The post-jump location of the process at time  $T_1$  is governed by the transition distribution  $Q(\phi(x_0, T_1), dy)$  and the motion restarts from this new point as before.

To fix the ideas, let us consider two major examples of unidimensional PDMP.

The TCP (transmission control protocol) (see Dumas et al. [17], Guillemin et al. [20] for instance) is one of the main data transmission protocol in Internet. The maximum number of packets that can be sent at time  $t_k$  in a round is a random variable  $X_{t_k}$ . If the transmission is successful, then the maximum number of packets is increased by one:  $X_{t_{k+1}} = X_{t_k} + 1$ . If the transmission fails, then we set  $X_{t_{k+1}} = \kappa X_{t_k}$  with  $\kappa \in (0, 1)$ . A correct scaling of this process leads to a piecewise deterministic Markov process  $(X_t)$  with flow  $\phi(x, t) = x + ct$  and deterministic transition measure  $Q(x, y) = \mathbb{1}_{\{y = \kappa x\}}$ . This process grows linearly (by construction) and the constant  $\kappa$  can be configured in the server implementation (so it is also known), but the moment when the transmission fails is of course unknown. In the literature it is usually supposed that the jump rate satisfies  $\lambda(x) = x$ , but with this work we can check whether it is a realistic assumption or not.

Another example of PDMP is the size of a marked bacteria (see Doumic et al. [16], Robert et al. [26], Laurençot and Perthame [25]). We randomly choose a bacteria, and follow its growth, until it divides in two. Then we randomly choose one of its daughters, and so on. Between the jumps, the bacteria grows exponentially:  $\phi(x, t) = xe^{ct}$ . The size of the bacteria after the division is random, as the bacteria does not divide itself in two equal parts.

The process  $(X_t)$  is observed continuously without errors (so the flow  $\phi$  is known); it is assumed to be ergodic, with fast convergence toward the stationary measure, and exponentially  $\beta$ -mixing. We denote by  $(T_1, \dots, T_n)$  the jump times and consider the Markov chain  $(Z_0 = x_0, (Y_k = X_{T_k^-}, Z_k = X_{T_k})_{k \in \mathbb{N}})$ . Our aim is to construct a non-parametric adaptive estimator of the jump rate  $\lambda$  on a compact interval.

There exist few results concerning PDMP's estimation. Azaïs et al. [5] and Azaïs and Muller-Gueudin [3] consider a more general model, for a multidimensional PDMP. They construct a quotient of kernel estimators, which estimate the compound function  $\lambda(\phi(x, t))$ . Their estimator is consistent ([5]), asymptotically normal, and its pointwise rate of convergence depends on the bandwidth of the kernel (see [3]). They explain how to construct an adaptive estimator, but do not bound its risk.

Doumic et al. [16] and Hodara et al. [21] also consider multi-dimensional PDMPs but for very specific biological models.

Fujii [19] and Krell [24] both consider unidimensional PDMP, and provide estimators of  $\lambda(x)$ . [19] constructs an estimator of  $\lambda(x)$  thanks to a Rice formula, by estimating local times. He proves the consistency of his estimators. [24] considers a deterministic transition measure (so  $Y_k$  is a function of  $Z_k$ ). Her estimator of  $\lambda$  is a quotient of a kernel estimator

of the stationary density of  $Z_k$  and an empirical estimator  $\hat{\mathbf{D}}_n$  of another function  $\mathbf{D}$  with the parametric rate of convergence  $n^{1/2}$ . This nonparametric estimator is asymptotically normal, and bounds for the pointwise risk are provided. In a very recent article, Azaïs and Genadot [2] construct a nonparametric estimator of  $\lambda(x)$  for a multidimensional PDMP and prove its consistency.

This article is an extension of the work of [24]. We consider a wider class of models (in particular, the transition measure  $Q$  does not need to be deterministic any more). We bound the  $L^2$  risk of the adaptive estimator, whereas [24] only considers the pointwise risk of the nonparametric estimator with fixed bandwidth  $h$ . We also prove that our estimator is minimax (up to a  $\ln^2(n)$  factor).

For this purpose, in analogy with [24], we use the equality

$$\lambda(x) = \frac{\nu(x)}{\mathbf{D}(x)}$$

where  $\nu$  is the stationary density of pre-jump locations  $Y_k$  (see Assumption A2 for the existence of this stationary density) and  $\mathbf{D}$  a function defined in equation (5). We get an estimator  $\hat{\mathbf{D}}_n(x)$ , which converges with rate  $n^{1/2}$ . To estimate the density function  $\nu$ , we use a projection method. We obtain a series of estimators  $(\hat{\nu}_0, \hat{\nu}_1, \dots, \hat{\nu}_m, \dots)$  of  $\nu$ . Then we choose the "best" estimator by a penalization method, in the same way as Barron et al. [6], and give an oracle inequality for the adaptive estimator  $\hat{\nu}_{\hat{m}}$ . The constant in the penalty term is intractable, but can be estimated thanks to a slope heuristic. Finally, we construct a quotient estimator of  $\lambda$ ,  $\hat{\lambda} = \hat{\nu}_{\hat{m}}/\hat{\mathbf{D}}_n$ , and bound its  $L^2$ -risk. In Section 2, we specify the model and its assumptions. The main results are stated in Section 3. Proofs are gathered in Section 4 and in Appendix A for the technical results. In Appendix B, some simulations for the TCP protocol and the bacterial growth are provided, with various functions  $\lambda$ . The outcomes are consistent with the theoretical results.

## 2 PDMP

A piecewise deterministic Markov process (PDMP) is defined by its local characteristics, namely, the jump rate  $\lambda$ , the flow  $\phi$  and the transition measure  $Q$  according to which the location of the process is chosen after the jump. In this article, we consider a unidimensional PDMP  $\{X(t)\}_{t \geq 0}$ . More precisely,

### Assumption A1.

- a. *The flow  $\phi : \mathbb{R}^+ \times \mathbb{R}^+ \mapsto \mathbb{R}^+$  is a one-parameter group of homeomorphisms:  $\phi$  is  $\mathcal{C}^1$ , for each  $t \in \mathbb{R}^+$ ,  $\phi(\cdot, t)$  is an homeomorphism satisfying the semigroup property:  $\phi(\cdot, t + s) = \phi(\phi(\cdot, s), t)$  and for each  $x \in \mathbb{R}^+$ ,  $\phi_x(\cdot) := \phi(x, \cdot)$  is an increasing  $\mathcal{C}^1$ -diffeomorphism. This implies that  $\phi(x, 0) = x$ .*

b. The jump rate  $\lambda : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a measurable function satisfying

$$\forall x \in \mathbb{R}^+, \exists \varepsilon' > 0 \text{ such that } \int_0^{\varepsilon'} \lambda(\phi(x, s)) ds < \infty$$

that is, the jump rate does not explode.

c.  $\forall x \in \mathbb{R}^+, Q(x, \mathbb{R}^+ \setminus \{x\}) = 1$ .

For instance, we can take  $\phi(x, t) = x + ct$  (linear flow) or  $\phi(x, t) = xe^{ct}$  (exponential flow). The transition measure may be continuous with respect to the Lebesgue measure or deterministic ( $Q(x, \{y\}) = \mathbb{1}_{\{y=f(x)\}}$ ).

Given these three characteristics, it can be shown (Davis [13, p62-66]), that there exists a filtered probability space  $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \{\mathbb{P}_x\})$  such that the motion of the process  $\{X(t)\}_{t \geq 0}$  starting from a point  $x_0 \in \mathbb{R}^+$  may be constructed as follows. Consider a random variable  $T_1$  with survival function

$$\mathbb{P}(T_1 > t | X_0 = x_0) = e^{-\Lambda(x_0, t)}, \text{ where } \Lambda(x, t) = \int_0^t \lambda(\phi(x, s)) ds. \quad (1)$$

If  $T_1$  is equal to infinity, then the process  $\{X(t)\}_{t \geq 0}$  follows the flow, i.e. for  $t \in \mathbb{R}^+$ ,  $X(t) = \phi(x_0, t)$ . Otherwise let  $Y_1 = \phi(x_0, T_1^-)$  the pre-jump location and  $Z_1$  the post-jump location.  $Z_1$  is defined through the transition kernel  $Q$ :  $\mathbb{P}(Z_1 \in A | Y_1 = y) = \int_A Q(y, dz)$ . The trajectory of  $\{X(t)\}$  starting at  $x_0$ , for  $t \in [0, T_1]$ , is given by

$$X(t) = \begin{cases} \phi(x_0, t) & \text{for } t < T_1, \\ Z_1 & \text{for } t = T_1. \end{cases}$$

Inductively starting from  $X(T_n) = Z_n$ , we now select the next inter-jump time  $T_{n+1} - T_n$  and post-jump location  $X(T_{n+1}) = Z_{n+1}$  in a similar way. This construction properly defines a strong Markov process  $\{X(t)\}_{t \geq 0}$  with jump times  $\{T_k\}_{k \in \mathbb{N}}$  (where  $T_0 = 0$ ). A very natural Markov chain is linked to  $\{X(t)\}_{t \geq 0}$ , namely the jump chain  $\{Y_n, Z_n\}_{n \in \mathbb{N}}$  (or, equivalently,  $\{T_n, Z_n\}_{n \in \mathbb{N}}$ ).

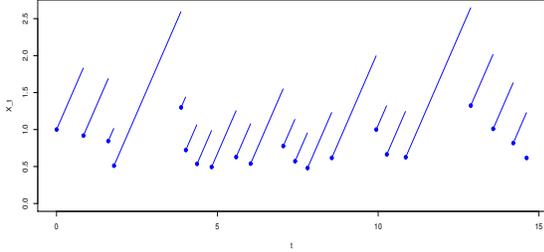
To simplify the notations, let us set  $\phi_x(t) = \phi(x, t)$  and  $z_0 = x_0$ . By (1),

$$\begin{aligned} \mathbb{P}(Y_1 > y | Z_0 = z_0) &= \mathbb{P}(T_1 > (\phi_{z_0})^{-1}(y) | Z_0 = z_0) \\ &= \exp\left(-\int_0^{(\phi_{z_0})^{-1}(y)} \lambda(\phi_{z_0}(s)) ds\right) \mathbb{1}_{\{y \geq z_0\}} \end{aligned}$$

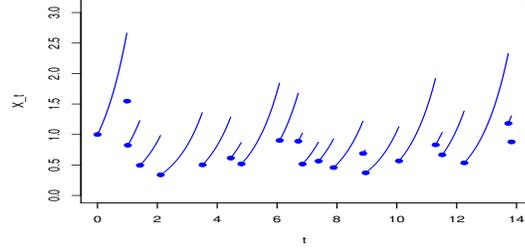
and by the change of variable  $u = \phi_{z_0}(s)$  (we recall that for any  $z \in \mathbb{R}^+$ ,  $\phi_z$  is a monotonic function), we get

$$\mathbb{P}(Y_1 > y | Z_0 = z_0) = \exp\left(-\int_{z_0}^y \lambda(u) (\phi_{z_0}^{-1})'(u) du\right) \mathbb{1}_{\{y \geq z_0\}}. \quad (2)$$

Figure 1: Examples of simulations of processes  $\{X_t\}_{t \geq 0}$  and  $\{Z_k\}_{k \in \mathbb{N}}$   
TCP protocol Bacterial growth



$$\phi(x, t) = x + t, \quad Z_k = Y_k/2, \\ \lambda(x) = \sqrt{x}$$



$$\phi(x, t) = xe^t, \quad Z_k = Y_k U, \quad U \sim \beta(20, 20), \\ \lambda(x) = x^2$$

• : process  $\{Z_k\}_{k \in \mathbb{N}}$     — : process  $\{X(t)\}_{t \geq 0}$

If the function  $\lambda(y)(\phi_{z_0}^{-1})'(y)$  is finite, we obtain the conditional density:

$$\mathcal{P}(z_0, y) := \lambda(y)(\phi_{z_0}^{-1})'(y) e^{-\int_{z_0}^y \lambda(u)(\phi_{z_0}^{-1})'(u) du} \mathbb{1}_{\{y \geq z_0\}}. \quad (3)$$

By analogy, we set  $\mathcal{P}(z_0, dy) = \mathbb{P}(Y_1 \in dy | Z_0 = z_0)$ .

Our aim is to estimate the jump rate  $\lambda$  on the compact interval  $\mathcal{I} := [i_1, i_2] \subset (0, \infty)$ .

The ergodicity is often a keystone in statistical inference for Markov processes. We also assume fast convergence toward the stationary density.

**Assumption A2.**

- a. The jump rate does not explode before  $i_2$ : for all  $x \leq i_1$ ,  $\int_0^{i_1} \lambda(y)(\phi_x^{-1})'(y) dy < \infty$  and  $\sup_{y \in [i_1, i_2]} \lambda(y) < \infty$ .
- b. The process  $(Y_k, Z_k)$  is recurrent positive and strongly ergodic. We denote by  $\nu$  the stationary measure of  $Y_k$ , by  $\mu$  that of  $Z_k$ , by  $\rho$  the stationary measure of the couple  $(Y_k, Z_k)$  and by  $\xi$  that of  $(Z_k, Y_{k+1})$ . We have that:

$$\mu(dz) = \int_{\mathbb{R}^+} \nu(dy) Q(y, dz) = \int_{\mathbb{R}^+} \rho(dy, dz), \quad \rho(dy, dz) = \nu(dy) Q(y, dz), \\ \nu(dy) = \int_{\mathbb{R}^+} \xi(dx, dy) = \int_{\mathbb{R}^+} \mathcal{P}(z, dy) \mu(dz), \quad \xi(dz, dy) = \mu(dz) \mathcal{P}(z, dy). \quad (4)$$

- c. There exist a function  $\mathbf{V}_\lambda$  greater than 1, two constants  $\gamma \in ]0, 1[$ ,  $R \in \mathbb{R}^{+*}$  such that, for any function  $\psi : (\mathbb{R}^+)^2 \mapsto \mathbb{R}^+$ ,  $|\psi| \leq \mathbf{V}_\lambda$ , for any integer  $k$ :

$$|\mathbb{E}(\psi(Y_k, Z_k) | Z_0 = z_0) - \mathbb{E}_\rho(\psi(Y_1, Z_1))| \leq R \mathbf{V}_\lambda(z_0) \gamma^k.$$

The inequality  $|\psi| \leq \mathbf{V}_\lambda$  means that, for any  $(y, z) \in (\mathbb{R}^+)^2$ ,  $|\psi(y, z)| \leq \mathbf{V}_\lambda(z)$ . This inequality is true in particular for any function  $\psi$  bounded by 1 and for  $\psi(y, z) = \mathbf{V}_\lambda(z)$ .

Under Assumption A2a, the conditional measure  $\mathcal{P}$  is continuous with respect to the Lebesgue measure on  $[0, i_2] \times [i_1, i_2]$  and  $\sup_{x, y \in [0, i_2] \times [i_1, i_2]} \mathcal{P}(x, y) < \infty$ . So is  $y \rightarrow \nu(y)$ :  $\nu(dy) = \nu(y)dy$ . Moreover,

$$\sup_{y \in [i_1, i_2]} \nu(y) = \sup_{y \in [i_1, i_2]} \int_0^{i_2} \mathcal{P}(z, y) \mu(dz) < \infty.$$

We can also remark that, for any  $x > 0$ ,  $|\mathbb{E}_\mu(\mathbf{V}_\lambda(Z_1))| \leq \mathbf{V}_\lambda(x) + R\mathbf{V}_\lambda(x) < \infty$ .

Let us set  $\mathbb{E}_{z_0}(U) = \mathbb{E}(U | Z_0 = z_0)$ . Under Assumption A2, the empirical mean is close to its expectation under the stationary density, as shown by the following lemma (proved in the Appendix).

**Lemma 1.** *Under Assumptions A1-A2, for any bounded function  $s$ :*

$$\left| \mathbb{E}_{z_0} \left( \frac{1}{n} \sum_{k=1}^n s(Y_k, Z_k) \right) - \int s(y, z) \rho(dy, dz) \right| \leq \|s\|_\infty \frac{R\mathbf{V}_\lambda(z_0)}{n(1-\gamma)}$$

and

$$\begin{aligned} \text{Var}_{z_0} \left( \frac{1}{n} \sum_{k=1}^n s(Y_k, Z_k) \right) &\leq \frac{1}{n} \int s^2(y, z) \rho(dy, dz) \\ &\quad + \frac{\|s\|_\infty}{n} \int |s(y, z)| G_\lambda(z) \rho(dy, dz) + \frac{c_\lambda \|s\|_\infty^2}{n^2} \end{aligned}$$

where  $G_\lambda(z) = \frac{R}{1-\gamma} (\mathbf{V}_\lambda(z) + \int \mathbf{V}_\lambda(u) \mu(du))$  and  $c_\lambda$  depends explicitly on  $(\gamma, R, \mathbf{V}_\lambda)$ . We can remark that  $C_\lambda := \int G_\lambda(z) \mu(dz) = \frac{2R}{1-\gamma} \int \mathbf{V}_\lambda(z) \mu(dz)$ .

In the bound of the variance, the first term is the same as for i.i.d variables. The second term is due to covariance terms (we found a similar term for stationary  $\beta$ -mixing processes), the third comes from the non-stationarity of the random vectors  $(Y_k, Z_k)$ .

To study an adaptive estimator of  $\nu$ , we need to prove that the Markov chain  $(Y_k, Z_k)$  is weakly dependent. It is the case if the process is  $\beta$ -mixing.

**Definition 2.** *Let  $(X_k)_{k \geq 0}$  be a Markov process. Let us define the  $\sigma$ -algebra*

$$\mathcal{O}_a^b = \sigma(\{X_{j_1} \in I_1, \dots, X_{j_n} \in I_n\}, a \leq j_1 \leq \dots \leq j_n \leq b, \mathbf{n} \in \mathbb{N}, I_k \in \mathcal{B}(\mathbb{R}^+)).$$

The  $\beta$ -mixing coefficient of the Markov chain  $(X_k)$  is

$$\beta_X(t) = \sup_k \sup_{E \in \mathcal{O}_0^k \times \mathcal{O}_{t+k}^\infty} |P_{\mathcal{O}_0^k, \mathcal{O}_{t+k}^\infty}(E) - P_{\mathcal{O}_0^k} \otimes P_{\mathcal{O}_{t+k}^\infty}(E)|$$

where  $P_{\mathcal{O}, \mathcal{S}}$  is the joint law of an event on  $\mathcal{O} \times \mathcal{S}$ . The  $\beta$ -mixing coefficient characterizes the dependence between what happens before  $T_k$  and what happens after  $T_{t+k}$ . The process  $(X_k)_{k \geq 0}$  is  $\beta$ -mixing if  $\lim_{k \rightarrow \infty} \beta_X(k) = 0$ . It is exponentially (or geometrically)  $\beta$ -mixing if there exists two positive constants  $c, \beta$  such that  $\beta_X(k) \leq ce^{-\beta k}$ .

The following lemma is a consequence of Assumption A2. It is proved in the Appendix.

**Lemma 3.** *Under Assumptions A1-A2, the Markov chain  $(Y_k, Z_k)$  is geometrically  $\beta$ -mixing. Moreover, its  $\beta$ -mixing coefficient satisfies:  $\forall k \in \mathbb{N}$ :*

$$\beta_{Y,Z}(k) \leq c\gamma^k \quad \text{where} \quad c = R \int \mathbf{V}_\lambda(z)\mu(dz) + R(1+R)\mathbf{V}_\lambda(x_0).$$

Estimating directly  $\lambda$  is difficult, but we can construct a quotient estimator. By (2) and (3), we get that, for any  $y \in \mathcal{I}$ ,

$$\begin{aligned} \lambda(y)(\phi_{z_0}^{-1})'(y)\mathbb{1}_{\{z_0 \leq y\}}\mathbb{P}(Y_1 > y | Z_0 = z_0) &= \mathcal{P}(z_0, y) \\ \lambda(y)\mathbb{E}\left(\mathbb{1}_{\{Z_0 \leq y < Y_1\}}(\phi_{Z_0}^{-1})'(y) \mid Z_0 = z_0\right) &= \mathcal{P}(z_0, y) \end{aligned}$$

and we integrate with respect to the stationary distribution  $\mu$  of  $Z_0$

$$\lambda(y)\mathbb{E}_\xi\left(\left(\phi_{Z_0}^{-1}\right)'(y)\mathbb{1}_{\{Z_0 \leq y < Y_1\}}\right) = \int \mathcal{P}(z, y)\mu(dz) = \nu(y)$$

recalling that  $\xi$  is the stationary measure of the couple  $(Z_0, Y_1)$ . Let us set

$$\mathbf{D}(y) := \mathbb{E}_\xi\left(\left(\phi_{Z_0}^{-1}\right)'(y)\mathbb{1}_{\{Z_0 \leq y < Y_1\}}\right). \quad (5)$$

Then, if  $\mathbf{D}(y) > 0$ , we get:

$$\lambda(y) = \frac{\nu(y)}{\mathbf{D}(y)}. \quad (6)$$

It remains to ensure that  $\mathbf{D}(y) > 0$  on  $\mathcal{I} = [i_1, i_2]$ .

**Assumption A3.** *There exists  $D_0 > 0$  such that*

$$\inf_{y \in \mathcal{I}} \mathbf{D}(y) \geq D_0 > 0.$$

*Remark.* Assumption A3 is very natural; indeed, let us set  $\Phi_0 := \inf_{x \leq i_2, y \in \mathcal{I}} (\phi_x^{-1})'(y)$ . As  $\phi_x$  is invertible, and  $\phi'(\cdot)$  is continuous,  $\Phi_0 > 0$ . Then

$$\mathbf{D}(y) \geq \Phi_0 \mathbb{P}_\xi(Z_0 \leq y < Y_1).$$

If the probability  $\mathbb{P}_\xi(Z_0 \leq y < Y_1)$  is null, then under the stationary distribution, the probability that  $(X_t)$  passes through  $y$  is null and the jump rate at that point can not be measured.

We can remark that if  $\mathbf{D} > 0$  for some point  $y$ , then so is  $\mathbf{P}(1_{Z_0 \leq y \leq Y_1}) > 0$  and its estimator

$$\hat{D}_n(y) = \frac{1}{n} \sum_{k=1}^n (\phi_{Z_k}^{-1})'(y) 1_{Z_{k-1} \leq y \leq Y_k} > 0.$$

Then if we take an interval  $[\hat{i}_1, \hat{i}_2]$  such that for some  $n$ , and some observation  $(X_t)_{t \geq 0}$ ,  $\hat{D}_n$  is positive on this interval, then Assumption A3 is satisfied on  $[\hat{i}_1, \hat{i}_2]$ . However, the true

value of  $D_0$  is unknown in that case. It should be noted that the interval  $[\hat{i}_1, \hat{i}_2]$  should not be changed for each simulation, otherwise the convergence of the estimator on the whole interval can not be guaranteed (the interval of estimation would become larger and larger, and as  $D$  is smaller on the edges on the new interval, and the convergence of the estimator is therefore slower).

Assumptions A2 and A3 are not explicit in  $(\lambda, Q, \phi)$ , so it is not easy to check that a particular model satisfies those assumptions. We give some explicit sufficient conditions on the coefficients  $(\lambda, Q, \phi)$ . For the next assumption, we use the Hölder spaces  $H^\alpha$ , as defined in Appendix A.4.

**Assumption (S).**

- a. *The transition kernel is a contraction mapping: there exists  $\kappa < 1$ , such that  $\mathbb{P}(Z_1 \leq \kappa Y_1) = 1$ .*
- b. *The flow is bounded: there exist two functions  $\mathbf{m}$  and  $\mathbf{M}$  such that,  $\forall x, y \in (\mathbb{R}^+)^2$ :*

$$0 < \mathbf{m}(y) \leq (\phi_x^{-1})'(y) \leq \mathbf{M}(y).$$

- c. *The jump rate is positive on  $[i_1, \infty[$  and there exists  $\mathbf{a} > 0$ ,  $b > -1$  such that*

$$\forall y \geq i_1, \quad \lambda(y)\mathbf{m}(y) \geq \mathbf{a} \frac{y^b}{b+1}.$$

*Then  $\forall y \geq z$ ,  $\mathbb{P}_z(Y_1 \geq y) \leq \exp(-\mathbf{a}(y^{b+1} - z^{b+1}))$  and  $\lim_{y \rightarrow \infty} \mathbb{P}_z(Y_1 \geq y) = 0$ .*

- d. *The jump rate does not explode too soon: there exist two positive constants  $\mathbf{L}, \mathbf{l}$ , such that  $\|\lambda\|_{L^\infty([i_1, i'_2])} \leq \mathbf{L}$  and  $\int_0^{i_1} \lambda(u)\mathbf{M}(u)du \leq \mathbf{l}$  where*

$$i'_2 = \max \left( i_2, (i_2 - i_1) + \left( \frac{1}{\mathbf{a}(1 - \kappa^{b+1})} \ln \left( \frac{2\kappa^{b+1}}{1 - \kappa^{b+1}} \right) \right)^{1/(b+1)} \mathbb{1}_{\{\kappa^{b+1} \geq 1/3\}} \right).$$

*These conditions ensure that Assumptions A2 and A3 are satisfied. The following two assumptions allow us to control the regularity of  $\nu$  (the rate of convergence of the estimator  $\hat{\lambda}_n$  depends on the regularity of  $\nu$ , not on the regularity of  $\lambda$ ).*

- e. *For any  $y \in \mathbb{R}^+$ ,  $\lambda(y) < \infty$ . This ensures that  $\nu$  and  $\mathcal{P}$  are continuous with respect to the Lebesgue measure on  $\mathbb{R}^+$ .*
- f. *There exists  $\alpha > 0$  such that:*

- $\forall K \subset \mathbb{R}^{+*}$  compact,  $\forall z \in \mathbb{R}^{+*}$ , the function  $(\phi_x^{-1})'(\cdot)$  belongs to  $H^\alpha([0, z] \times K)$ .
- $\forall K \subset \mathbb{R}^{+*}$  compact,  $\lambda \in H^\alpha(K)$ .

- The transition measure  $Q$  can be written

$$Q(x, dy) = Q_1(x, y)dy + p_0(x)\delta_0(dy) + \sum_{i=1}^{j_Q} p_i(x)\delta_{f_i(x)}(dy)$$

with, for any compact  $K$ ,  $Q_1$  and  $(p_i)_{0 \leq i \leq j_Q}$  in  $H^{\alpha-1}(K)$ , and  $(f_i)_{1 \leq i \leq j_Q}$  invertible functions such that  $(f_i^{-1})_{1 \leq i \leq j_Q} \in H^\alpha(K)$ .

If Assumption (S) is satisfied, for fixed flow  $\phi$  and transition measure  $Q$ , we can introduce the class of functions

$$\mathcal{E}(\mathfrak{s}, b, \alpha) = \left\{ \lambda \in H^\alpha(\mathcal{J}), \forall y \geq i_1, \lambda(y)\mathbf{m}(y) \geq \frac{\mathbf{a}y^b}{b+1}, \int_0^{i_1} \lambda(u)\mathbf{M}(u) \leq \mathbf{1}, \|\lambda\|_{H^\alpha(\mathcal{J})} \leq \mathbf{L} \right\}$$

with  $\mathfrak{s} = (\mathbf{a}, \mathbf{1}, \mathbf{L}) \in (\mathbb{R}^+)^3$  and the convex set

$$\mathcal{J} = \mathcal{J}_{[\alpha]} \cup [i_1, i_2'] := [j_1, j_2] \quad (7)$$

is defined by the recurrence:

$$\mathcal{J}_0 = \mathcal{I} \quad \text{and} \quad \mathcal{J}_{k+1} = \text{Conv} \left( \mathcal{I} \cup \bigcup_{i=1}^{j_Q} f_i^{-1}(\mathcal{J}_k) \right).$$

The following lemmas are proved in the Appendix.

**Lemma 4.** *Under Assumptions A1 and (S)*

- Assumption A2 is satisfied for  $\mathbb{V}_b(x) := \exp(\mathbf{a}x^{b+1})$ : there exists  $R, \gamma$ , for any function  $|\psi| \leq \mathbb{V}_b$ ,

$$\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} |\mathbb{E}_{z_0}(\psi(Y_k, Z_k)) - \mathbb{E}_\rho(\psi(Y_0, Z_0))| \leq R\mathbb{V}_b(z_0)\gamma^k$$

recalling that the inequality  $|\psi| \leq \mathbb{V}_b$  means that, for any  $(y, z) \in (\mathbb{R}^+)^2$ ,  $|\psi(y, z)| \leq \mathbb{V}_b(z)$ .

- Assumption A3 is satisfied. Moreover, there exists  $\eta > 0, D_0 > 0$  such that

$$\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mu([0, i_1]) \geq \eta \quad \text{and} \quad \inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \inf_{y \in \mathcal{I}} \mathbf{D}(y) \geq D_0.$$

**Lemma 5.** *If Assumptions A1 and (S) are satisfied, we can control the regularity of  $\nu$ :*

$$\|\nu\|_{H^\alpha(\mathcal{I})} \leq \psi_Q \left( \|\lambda\|_{H^\alpha(\mathcal{J})}, \|(\phi^{-1})'\|_{H^\alpha([0, j_2] \times \mathcal{J})} \right) \quad \text{with } \mathcal{J} = [j_1, j_2] \text{ defined in (7)}.$$

*Remark.* In [24], the author introduces the set of functions  $\mathcal{F}(\mathbf{c}, b)$  with very similar conditions. As she considers a transition measure  $Q$  deterministic, the sets  $\mathcal{F}(\mathbf{c}, b)$  and  $\mathcal{E}(\mathfrak{s}, b, \alpha) \cap H^\alpha$  may not be equal. In particular, if  $\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)$ , then there exists  $\mathbf{c}$  such that  $\lambda \in \mathcal{F}(\mathbf{c}, b) \cap H^\alpha$ . On the contrary, if  $\lambda$  belongs to  $\mathcal{F}(\mathbf{c}, b) \cap H^\alpha$  and the deterministic transition  $f$  is  $f(x) = \kappa x$ , then for  $i_1$  large enough, there exists  $\mathfrak{s}$  such that  $\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)$ . This is no longer the case if, for instance,  $f(x) \propto x^\beta$ . As the transition measure  $Q$  is unknown, it is not possible to exploit its characteristics.

Another difference between the two sets is that  $\lambda$  is estimated on the fixed interval  $[i_1, i_2]$  and the assumptions depends on  $(i_1, i_2)$ , whereas in [24], the interval of estimation depends on the set  $\mathcal{F}(\mathbf{c}, b)$ .

## 3 Estimation of the jump rate

### 3.1 The observation scheme

As in [3] and [24], the statistical inference is based on the observation scheme  $(X(t), t \leq T_n)$  and asymptotics are considered when the number of jumps of the process,  $n$ , goes to infinity. Actually the simpler observation scheme:  $(X(0), (X(T_{i-}), X(T_i)), 1 \leq i \leq n) = (Z_0, (Y_i, Z_i), 1 \leq i \leq n)$  is sufficient, as  $\phi$  is known and one can remark that for all  $n \geq 1$ ,  $T_n = \phi_{Z_{n-1}}^{-1}(Y_n)$ .

### 3.2 Methodology

[24] and [3] construct a pointwise kernel estimator of  $\nu$  before deriving an estimator of  $\lambda$ . Indeed, densities are often approximated by kernels methods (see Tsybakov [30] for instance). If the kernel is positive, the estimator is also a density. However, we want to control the  $L^2$  risk of our estimator (not the pointwise risk), and also to construct an adaptive estimator. Estimators by projection are well adapted for  $L^2$  estimation: if they are longer to compute at a single point than pointwise estimators, it is sufficient to know the estimated coefficients to construct the whole function. Furthermore, to find an adaptive estimator, we minimize a function of the norm of our estimator, that is the sum of the square of the coefficients, and the dimension. That is the reason why we choose an estimation by projection.

We first aim at estimating  $\nu$  on the compact set  $\mathcal{I}$ . We construct a sequence of  $L^2$  estimators by projection on an orthonormal basis. As usual in nonparametric estimation, their risks can be decomposed in a variance term and a bias term which depends of the regularity of the density function  $\nu$ . We choose to use the Besov spaces (see Section A.4) to characterize the regularity, which are well adapted to  $L^2$  estimation (particularly for the wavelet decomposition). The "best" estimator is then selected by penalization. To construct the sequence of estimators, we introduce a sequence of vectorial subspaces  $S_m$ . We construct an estimator  $\hat{\nu}_m$  of  $\nu$  on each subspace and then select the best estimator  $\hat{\nu}_{\hat{m}}$ .

**Assumption A4.**

a. The subspaces  $S_m$  are increasing and have finite dimension  $D_m$ .

b. The  $L^2$ -norm and the  $L^\infty$ -norm are connected:

$$\exists \psi_1 > 0, \forall m \in \mathbb{N}, \forall s \in S_m, \quad \|s\|_\infty^2 \leq \psi_1 D_m \|s\|_{L^2}^2.$$

This implies that, for any orthonormal basis  $(\varphi_l)$  of  $S_m$ ,

$$\left\| \sum_{l=1}^{D_m} \varphi_l^2 \right\|_\infty \leq \psi_1 D_m.$$

c. There exists a constant  $\psi_2 > 0$  such that, for any  $m \in \mathbb{N}$ , there exists an orthonormal basis  $\varphi_l$  such that:

$$\left\| \sum_{l=1}^{D_m} \|\varphi_l\|_\infty |\varphi_l(x)| \right\|_\infty \leq \psi_2 D_m.$$

d. There exists  $\mathbf{r} \in \mathbb{N}$ , called the regularity of the decomposition, such that:

$$\exists C > 0, \forall \alpha \leq \mathbf{r}, \forall s \in B_{2,\infty}^\alpha, \quad \|s - s_m\|_{L^2} \leq C D_m^{-\alpha} \|s\|_{B_{2,\infty}^\alpha}$$

where  $s_m$  is the orthogonal projection of  $s$  on  $S_m$  and  $B_{2,\infty}^\alpha$  is a Besov space (see Appendix A.4).

Conditions a, b and d are usual (see Comte et al. [12, section 2.3] for instance). They are satisfied for subspaces generated by wavelets, piecewise polynomials or trigonometric polynomials (see DeVore and Lorentz [15] for trigonometric polynomials and piecewise polynomials and Meyer [27] for wavelets). Condition c is necessary because we are not in the stationary case: it helps us to control some covariance terms. It is obviously satisfied for bounded bases (trigonometric polynomials), and localized bases (piecewise polynomials). Let us prove it for a wavelet basis. Let  $\varphi$  be a father wavelet function, then  $D_m = 2^m$  and  $\varphi_l(x) = 2^{m/2} \varphi(2^m x - l)$ . We get that  $\left\| \sum_{l=1}^{D_m} \|\varphi_l\|_\infty |\varphi_l(x)| \right\|_\infty \leq 2^m \|\varphi\|_\infty \left\| \sum_{l \in \mathbb{Z}} |\varphi(x - l)| \right\|_\infty$ . As  $\varphi$  is at least 0-regular, for  $m = 2$ , there exists a constant  $C$  such that  $|\varphi(x)| \leq C(1 + |x|^{-2})$ . Then  $\sup_x \sum_{l \in \mathbb{Z}} |\varphi(x - l)| \leq C \sup_x \sum_{l \in \mathbb{Z}} (1 + |x - l|^{-2}) < \infty$  and condition c is satisfied.

### 3.3 Estimation of the stationary density

Let us now construct an estimator  $\hat{\nu}_m$  of  $\nu$  on the vectorial subspace  $S_m$ . We consider an orthonormal basis  $(\varphi_l)$  of  $S_m$  satisfying Assumption A4. Let us set

$$a_l = \langle \varphi_l, \nu \rangle = \int_{\mathcal{I}} \varphi_l(x) \nu(x) dx \quad \text{and} \quad \nu_m(x) = \sum_{l=1}^{D_m} a_l \varphi_l(x).$$

The function  $\nu_m$  is the orthogonal projection of  $\nu$  on  $L^2(\mathcal{I})$ . We consider the estimator

$$\hat{\nu}_m(x) = \sum_{l=1}^{D_m} \hat{a}_l \varphi_l(x) \quad \text{with} \quad \hat{a}_l = \frac{1}{n} \sum_{k=1}^n \varphi_l(Y_k).$$

**Proposition 6.** *If  $D_m^2 \leq n$ , under Assumptions A1-A2 and A4,*

$$\mathbb{E}_{z_0} \left( \|\hat{\nu}_m - \nu\|_{L^2(\mathcal{I})}^2 \right) \leq \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + (\psi_1 + C_\lambda \psi_2) \frac{D_m}{n} + \frac{c}{n}$$

where  $C_\lambda = \frac{2R}{1-\gamma} \int \mathbf{V}_\lambda(z) \mu(dz)$  and  $c$  depends explicitly on  $\mathbf{V}_\lambda, \gamma, R$ .

When  $m$  increases, the bias term decreases whereas the variance term increases. It is important to find a good bias-variance compromise. If  $\nu$  belongs to the Besov space  $B_{2,\infty}^\alpha(\mathcal{I})$ , then  $\|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 \leq C \|\nu\|_{B_{2,\infty}^\alpha(\mathcal{I})} D_m^{-2\alpha}$  (see Assumption A4d). If  $\alpha \geq 1/2$ , the risk is then minimum for  $D_{m_{opt}} \propto n^{1/(2\alpha+1)}$  and we have, for some continuous function  $\psi$ :

$$\mathbb{E}_{z_0} \left( \|\hat{\nu}_{m_{opt}} - \nu\|_{L^2(\mathcal{I})}^2 \right) \leq \psi \left( \|\nu\|_{B_{2,\infty}^\alpha(\mathcal{I})}, \mathbf{V}_\lambda, R, \gamma \right) n^{-2\alpha/(2\alpha+1)}.$$

This is the usual nonparametric convergence rate (see Tsybakov [30]). If  $\alpha < 1/2$ , then the risk is minimum for  $D_m = n^{1/2}$  and the bias term is greater than the variance term. We can remark that a piecewise continuous function belongs to  $B_{2,\infty}^{1/2}$ .

Let us now construct the adaptive estimator. We compute  $(\hat{\nu}_0, \dots, \hat{\nu}_m, \dots)$  for  $m \in \mathcal{M}_n = \{m, D_m^2 \leq n\}$ . Our aim is to select automatically  $m$ , without knowing the regularity of the stationary density  $\nu$ . Let us introduce the contrast function  $\gamma_n(s) = \|s\|_{L^2}^2 - \frac{2}{n} \sum_{k=1}^n s(Y_k)$ . If  $s \in S_m$ , then we can write  $s = \sum_l b_l \varphi_l$  and

$$\gamma_n(s) = \sum_{l=1}^{D_m} b_l^2 - \sum_{l=1}^{D_m} b_l \frac{2}{n} \sum_{k=1}^n \varphi_l(Y_k).$$

The minimum is obtained for  $b_l = \hat{a}_l = \frac{1}{n} \sum_{k=1}^n \varphi_l(Y_k)$ . Therefore

$$\hat{\nu}_m = \arg \min_{s \in S_m} \gamma_n(s). \tag{8}$$

As the subspaces  $S_m$  are increasing, the function  $\gamma_n(\hat{\nu}_m)$  decreases when  $m$  increases. To find an adaptive estimator, we need to add a penalty term  $pen(m)$ . Let us set  $pen(m) = \frac{48(\psi_1 + C_\lambda \psi_2) D_m}{n} + \frac{48c_\lambda \psi_1}{n}$  (or more generally  $pen(m) = \frac{\sigma D_m}{n} + \frac{\sigma'}{n}$ , with  $\sigma \geq 48(\psi_1 + C_\lambda \psi_2)$ ,  $\sigma' \geq 48c_\lambda \psi_1$ ) and choose

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \gamma_n(\hat{\nu}_m) + pen(m). \tag{9}$$

We obtain an adaptive estimator  $\hat{\nu}_{\hat{m}}$ .

**Theorem 7** (Risk of the adaptive estimator). *Under Assumptions A1-A2 and A4,  $\forall \sigma \geq 48(\psi_1 + C_\lambda \psi_2)$ ,  $\sigma' \geq 48c_\lambda \psi_1$ ,  $\text{pen}(m) = \frac{\sigma D_m}{n} + \frac{\sigma'}{n}$ ,*

$$\mathbb{E}_{z_0} \left( \|\nu - \hat{\nu}_{\hat{m}}\|_{L^2(\mathcal{I})}^2 \right) \leq \min_{m \in \mathcal{M}_n} \left( 3 \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + 4\text{pen}(m) \right) + \frac{c'}{n}.$$

where  $c'$  is a function of  $(\mathbf{V}_\lambda, R, \gamma, \|\nu\|_{L^2(\mathcal{I})})$ . We recall that  $\mathcal{M}_n = \{m, D_m^2 \leq n\}$ .

The estimator is adaptive: it realizes the best bias-variance compromise, up to a multiplicative constant. We have an explicit rate of convergence if  $\nu$  belongs to some (unknown) Besov space  $B_{2,\infty}^\alpha$ : in that case,

$$\|\nu - \nu_m\|_{L^2(\mathcal{I})}^2 \leq 3 \|\nu_{m_{\text{opt}}} - \nu\|_{L^2(\mathcal{I})}^2 + 4\text{pen}(m_{\text{opt}}) + \frac{c}{n} \leq C \|\nu\|_{B_{2,\infty}^\alpha} D_m^{-2\alpha}$$

and if  $\alpha \geq 1/2$ ,

$$\mathbb{E}_{z_0} \left( \|\nu - \hat{\nu}_{\hat{m}}\|_{L^2(\mathcal{I})}^2 \right) \leq \psi \left( \|\nu\|_{B_{2,\infty}^\alpha(\mathcal{I})}, \mathbf{V}_\lambda, R, \gamma \right) n^{-2\alpha/(2\alpha+1)} \quad (10)$$

for some continuous function  $\psi$ .

### 3.4 Estimation of the jump rate

By (6), we have

$$\lambda(y) = \frac{\nu(y)}{\mathbf{D}(y)} \quad \text{recalling that} \quad \mathbf{D}(y) = \mathbb{E}_\xi \left( (\phi_{Z_0}^{-1})'(y) \mathbb{1}_{\{Z_0 \leq y \leq Y_1\}} \right)$$

where  $\xi$  is the stationary measure of  $(Z_k, Y_{k+1})$ .

*Remark.* We notice that this formula is different as the one used in [24]

$$\lambda(y) = \frac{f(\nu(y))}{\tilde{\mathbf{D}}(y)}$$

where

$$\tilde{\mathbf{D}}(y) := \mathbb{E}_\nu \left( ((f \circ \phi_{Z_0})^{-1})'(f(y)) \mathbb{1}_{\{f(Z_0) \leq f(y)\}} \mathbb{1}_{\{Z_1 \geq f(y)\}} \right).$$

As in [24], the author works under the assumption that  $Q(x, \{y\}) = \mathbb{1}_{\{y=f(x)\}}$ , the study was easier, here we need to consider the Markov chain  $(Y_k, Z_k)_{k \in \mathbb{N}}$ .

To estimate the jump rate, we construct a quotient estimator. Let us consider the estimator

$$\hat{\lambda}_n(y) = \frac{\hat{\nu}_{\hat{m}}(y)}{\hat{\mathbf{D}}_n(y)} \mathbb{1}_{\{\hat{\nu}_{\hat{m}}(y) \geq 0\}} \mathbb{1}_{\{\hat{\mathbf{D}}_n(y) \geq \ln(n)^{-1}\}} \quad (11)$$

where

$$\hat{\mathbf{D}}_n(y) := \frac{1}{n} \sum_{k=1}^n (\phi_{Z_{k-1}}^{-1})'(y) \mathbb{1}_{\{Z_{k-1} \leq y \leq Y_k\}}.$$

*Remark.* As the process  $\{X(t)\}$  is observed continuously without errors,  $\phi^{-1}$  (and therefore  $(\phi^{-1})'$ ) is known on  $\cup_k [Z_{k-1}, Y_k]$  so  $\hat{\mathbf{D}}_n(y)$  is computable.

The estimator  $\hat{\lambda}_n$  converges with nearly the same rate of convergence as  $\hat{\nu}$ :

**Theorem 8.** *Under A1-A4, as soon as  $\ln(n)^{-1} \leq D_0/2$ ,*

$$\begin{aligned} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) &\leq 3 \ln^2(n) \mathbb{E}_{z_0} \left( \left\| \hat{\nu}_{\hat{m}} - \nu \right\|_{L^2(\mathcal{I})}^2 \right) + c'_\lambda \frac{\ln^2(n)}{n} \\ &\leq 3 \ln^2(n) \min_{m \in \mathcal{M}_n} \left\{ 3 \left\| \hat{\nu}_m - \nu \right\|_{L^2(\mathcal{I})}^2 + 4 \text{pen}(m) \right\} + c'_\lambda \frac{\ln^2(n)}{n} \end{aligned}$$

where

$$c'_\lambda = \Phi_1^2 \frac{2 + C_\lambda}{D_0^2} \left( 3 \left\| \lambda \right\|_{L^2(\mathcal{I})}^2 + 12 \left\| \nu \right\|_{L^2(\mathcal{I})}^2 \right), \quad \Phi_1 = \sup_{x \in [0, i_2], y \in \mathcal{I}} (\phi_x^{-1})'(y).$$

The bias term depends of the regularity of the stationary density  $\nu$ , not of the regularity of  $\lambda$ . If we consider  $\lambda$  and  $\nu$  as functions of a Besov space, their regularities are not related: the Besov spaces are not stable by product (as they are subspaces of  $L^2(\mathcal{I})$ ). We would like to link the rate of convergence of  $\hat{\lambda}_n$  to the regularity of  $\lambda$  rather than  $\nu$ , at least when  $\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)$ . In that case,  $\lambda$  belong to some Hölder space, which is stable by product, composition and integration. See Appendix A.4 for the definition and properties of Besov and Hölder spaces. We obtain the following corollary:

**Corollary 9.** *Under A1, (S) and A4, as soon as  $\ln(n)^{-1} \leq D_0/2$ , for any  $\alpha \geq 1/2$ ,*

$$\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \lesssim \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

*Remark.* [24] obtain the same rate of convergence for a kernel estimator (with the regularity of  $\lambda$  known).

### 3.5 Minimax bound for the estimator of the jump rate

We have proved that, under assumptions A1, (S) and A4,

$$\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \lesssim \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

We would like to verify that our estimator converges with the minimax rate of convergence, i.e:

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \geq C \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

The  $\ln^2(n)$  factor comes from the quotient estimator, we can not expect it will stay in the minimax bound. Indeed, it is clear that one could replace  $\ln^{-1}(n)$  in (11) by any function

$w(n)$  greater than  $D_0/2$ . The best estimator will be obtained of course by taking  $w(n) = D_0/2$  and the risk of this estimator (unreachable as  $D_0$  is unknown) will be proportional to  $n^{-2\alpha/(2\alpha+1)}$ .

**Theorem 10** (Minimax bound). *If A1, (S) and A4 are satisfied, then*

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(s,b,\alpha)} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \geq C n^{-2\alpha/(2\alpha+1)}$$

where the infimum is taken among all estimators.

## 4 Proofs

Lemmas 1, 3, 4 and 5 are proved in the Appendix.

### 4.1 Proof of Proposition 6

We have the following bias-variance decomposition:

$$\begin{aligned} \mathbb{E}_{z_0} \left( \left\| \nu - \hat{\nu}_m \right\|_{L^2(\mathcal{I})}^2 \right) &= \int_{\mathcal{I}} \mathbb{E}_{z_0} \left( (\nu(x) - \hat{\nu}_m(x))^2 \right) dx \\ &= \int_{\mathcal{I}} (\nu(x) - \mathbb{E}_{z_0}(\hat{\nu}_m(x)))^2 dx + \int_{\mathcal{I}} \text{Var}_{z_0}(\hat{\nu}_m(x)) dx \\ &= \left\| \mathbb{E}_{z_0}(\hat{\nu}_m) - \nu \right\|_{L^2(\mathcal{I})}^2 + \int_{\mathcal{I}} \text{Var}_{z_0}(\hat{\nu}_m(x)) dx. \end{aligned}$$

The estimator  $\hat{\nu}_m$  (and therefore its expectation  $\mathbb{E}_{z_0}(\hat{\nu}_m)$ ) belongs to the subspace  $S_m$ . Then, by orthogonality

$$\mathbb{E}_{z_0} \left( \left\| \nu - \hat{\nu}_m \right\|_{L^2(\mathcal{I})}^2 \right) = \left\| \nu - \nu_m \right\|_{L^2(\mathcal{I})}^2 + \left\| \mathbb{E}_{z_0}(\hat{\nu}_m) - \nu_m \right\|_{L^2(\mathcal{I})}^2 + \int_{\mathcal{I}} \text{Var}_{z_0}(\hat{\nu}_m(x)) dx.$$

The first terms are two terms of bias, the third is a variance term. Let us first bound the second term of bias. As the functions  $(\varphi_l)_{1 \leq l \leq D_m}$  form an orthonormal basis of  $S_m$ , we have

$$\begin{aligned} \left\| \mathbb{E}_{z_0}(\hat{\nu}_m) - \nu_m \right\|_{L^2(\mathcal{I})}^2 &= \sum_{l=1}^{D_m} (\mathbb{E}_{z_0}(\hat{a}_l) - a_l)^2 \\ &= \sum_{l=1}^{D_m} \left( \frac{1}{n} \sum_{k=1}^n \mathbb{E}_{z_0}(\varphi_l(Y_k)) - \int_{\mathcal{I}} \varphi_l(x) \nu(x) dx \right)^2. \end{aligned}$$

By Lemma 1,

$$\left| \frac{1}{n} \sum_{k=1}^n \mathbb{E}_{z_0}(\varphi_l(Y_k)) - \int_{\mathcal{I}} \varphi_l(x) \nu(x) dx \right| \leq \|\varphi_l\|_{\infty} \frac{R \mathbf{V}_{\lambda}(z_0)}{n(1-\gamma)}.$$

As the  $L^2$  and the  $L^\infty$ -norms are connected (see Assumption A4b),  $\|\varphi_l\|_\infty^2 \leq \psi_1 D_m$  and, since  $D_m^2 \leq n$ , we get:

$$\|\mathbb{E}_{z_0}(\hat{\nu}_m - \nu_m)\|_{L^2(\mathcal{I})}^2 \leq \sum_{l=1}^{D_m} \frac{\|\varphi_l\|_\infty^2}{n^2} \frac{R^2 \mathbf{V}_\lambda^2(z_0)}{(1-\gamma)^2} \leq \psi_1 \frac{D_m^2}{n^2} \frac{R^2 \mathbf{V}_\lambda^2(z_0)}{(1-\gamma)^2} \leq \psi_1 \frac{1}{n} \frac{R^2 \mathbf{V}_\lambda^2(z_0)}{(1-\gamma)^2}.$$

Let us now consider the variance term. As the functions  $(\varphi_l)$  form an orthonormal basis of  $S_m$ , the integrated variance of  $\hat{\nu}_m$  is the sum of the variances of the coefficients  $\hat{a}_\lambda$ :

$$\begin{aligned} \int_{\mathcal{I}} \text{Var}_{z_0}(\hat{\nu}_m(x)) dx &= \int_{\mathcal{I}} \text{Var}_{z_0} \left( \sum_{l=1}^{D_m} \hat{a}_l \varphi_l(x) \right) dx = \sum_{k,l} \text{Cov}_{z_0}(\hat{a}_k, \hat{a}_l) \langle \varphi_l, \varphi_k \rangle_{L^2(\mathcal{I})} \\ &= \sum_{l=1}^{D_m} \text{Var}_{z_0}(\hat{a}_l). \end{aligned}$$

By Lemma 1, as  $\int_{\mathbb{R}^+} \rho(x, dz) = \nu(x)$ , we get:

$$\begin{aligned} \text{Var}_{z_0}(\hat{a}_l) &= \text{Var}_{z_0} \left( \frac{1}{n} \sum_{k=1}^n \varphi_l(Y_k) \right) \\ &\leq \frac{1}{n} \int_{\mathcal{I}} \varphi_l^2(x) \nu(x) dx + \frac{\|\varphi_l\|_\infty}{n} \int_{\mathcal{I} \times \mathbb{R}^+} |\varphi_l(x)| G_\lambda(z) \rho(dx, dz) + \frac{c_\lambda \|\varphi_l\|_\infty^2}{n^2}. \end{aligned}$$

By Assumptions A4b and c,  $\forall x, \sum_{l=1}^{D_m} \varphi_l^2(x) \leq \psi_1 D_m$ ,  $\sum_{l=1}^{D_m} \|\varphi_l\|_\infty |\varphi_l(x)| \leq \psi_2 D_m$  and  $\sum_{l=1}^{D_m} \|\varphi_l\|_\infty^2 \leq \psi_1 D_m^2 \leq \psi_1 n$ . Therefore:

$$\int_{\mathcal{I}} \text{Var}_{z_0}(\hat{\nu}_m(x)) dx = \sum_{l=1}^{D_m} \text{Var}_{z_0}(\hat{a}_l) \leq (\psi_1 + C_\lambda \psi_2) \frac{D_m}{n} + \frac{c_\lambda}{n} \psi_1 \quad (12)$$

where  $C_\lambda = \int G_\lambda(z) \mu(dz) = \frac{2R}{1-\gamma} \int_{\mathcal{I}} \mathbf{V}_\lambda(z) \mu(dz)$  and  $c_\lambda$  depends only on  $\mathbf{V}_\lambda$ ,  $R$  and  $\gamma$ .

## 4.2 Proof of Theorem 7

The number of coefficients in the adaptive estimator is random. If we are still able to control easily the bias term, we can not simply control the variance of our estimator by adding the variances of its coefficients. For any  $m \in \mathcal{M}_n$ , by definition of  $\hat{m}$  (see (8) and (9)), we have the following inequality:

$$\gamma_n(\hat{\nu}_{\hat{m}}) \leq \gamma_n(\hat{\nu}_m) + \text{pen}(m) - \text{pen}(\hat{m}) \leq \gamma_n(\nu_m) + \text{pen}(m) - \text{pen}(\hat{m}),$$

with  $\gamma_n(s) = \|s\|_{L^2(\mathcal{I})}^2 - 2n^{-1} \sum_{k=1}^n s(Y_k)$ . Then

$$\|\hat{\nu}_{\hat{m}}\|_{L^2(\mathcal{I})}^2 \leq \|\nu_m\|_{L^2(\mathcal{I})}^2 + \text{pen}(m) - \text{pen}(\hat{m}) + \frac{2}{n} \sum_{k=1}^n (\hat{\nu}_{\hat{m}}(Y_k) - \nu_m(Y_k)). \quad (13)$$

We have that, for any function  $s \in L^2(\mathcal{I})$ ,  $\|s\|_{L^2(\mathcal{I})}^2 = \|s - \nu\|_{L^2(\mathcal{I})}^2 - \|\nu\|_{L^2(\mathcal{I})}^2 + 2 \int_{\mathcal{I}} s(x)\nu(x)dx$ . We apply this equality to  $\hat{\nu}_{\hat{m}}$  and  $\nu_m$ . Equation (13) becomes:

$$\begin{aligned} \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 &\leq \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + pen(m) - pen(\hat{m}) \\ &\quad + \frac{2}{n} \sum_{k=1}^n \hat{\nu}_{\hat{m}}(Y_k) - \nu_m(Y_k) - 2 \int_{\mathcal{I}} (\hat{\nu}_{\hat{m}}(x) - \nu_m(x))\nu(x)dx. \end{aligned}$$

The function  $\hat{\nu}_{\hat{m}} - \nu_m$  belongs to the vectorial subspace  $S_{\hat{m}} + S_m$ . Therefore:

$$\begin{aligned} \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 &\leq \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + pen(m) - pen(\hat{m}) \\ &\quad + 2 \|\hat{\nu}_{\hat{m}} - \nu_m\|_{L^2(\mathcal{I})} \sup_{s \in \mathcal{B}_{m, \hat{m}}} \left| \sum_{k=1}^n \frac{1}{n} s(Y_k) - \int_{\mathcal{I}} s(x)\nu(x)dx \right| \end{aligned}$$

where  $\mathcal{B}_{m, m'} = \{s \in S_m + S_{m'}, \|s\|_{L^2(\mathcal{I})} = 1\}$ . As the sequence  $(S_m)$  is increasing,  $S_m + S_{m'}$  is simply the largest of the two subspaces. By the inequality of arithmetic and geometric means,

$$\begin{aligned} \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 &\leq \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + pen(m) - pen(\hat{m}) + \frac{1}{4} \|\hat{\nu}_{\hat{m}} - \nu_m\|_{L^2(\mathcal{I})}^2 \\ &\quad + \sup_{s \in \mathcal{B}_{m, \hat{m}}} 4 \left( \frac{1}{n} \sum_{k=1}^n s(Y_k) - \int_{\mathcal{I}} s(x)\nu(x)dx \right)^2. \end{aligned}$$

By the triangular inequality,  $\|\hat{\nu}_{\hat{m}} - \nu_m\|_{L^2(\mathcal{I})}^2 \leq 2 \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 + 2 \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2$ , and:

$$\begin{aligned} \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 &\leq 3 \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + 2pen(m) - 2pen(\hat{m}) \\ &\quad + 8 \sup_{s \in \mathcal{B}_{m, \hat{m}}} \left( \frac{1}{n} \sum_{k=1}^n s(Y_k) - \int_{\mathcal{I}} s(x)\nu(x)dx \right)^2. \end{aligned}$$

We can decompose the last term in a bias term and a variance term. Let us set:

$$I_n(s) := \frac{1}{n} \sum_{k=1}^n s(Y_k) - \mathbb{E}_{z_0}(s(Y_k)), \quad J_n(s) := \frac{1}{n} \sum_{k=1}^n \left( \mathbb{E}_{z_0}(s(Y_k)) - \int_{\mathcal{I}} s(x)\nu(x)dx \right) \quad (14)$$

and  $p(m, m') := (pen(m) + pen(m'))/8$ . Then:

$$\begin{aligned} \mathbb{E}_{z_0} \left( \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 \right) &\leq 3 \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + 4pen(m) \\ &\quad + 16 \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) + J_n^2(s) \right) - 16p(m, \hat{m}). \end{aligned} \quad (15)$$

By Assumption A4b,  $s \in \mathcal{B}_{m, \hat{m}}$  implies that  $\|s\|_\infty^2 \leq \psi_1(D_m + D_{\hat{m}}) \leq 2\psi_1 n^{1/2}$  (we recall that  $D_m$  and  $D_{\hat{m}}$  are smaller than  $n^{1/2}$ ). Then by Lemma 1,

$$\sup_{s \in \mathcal{B}_{m, \hat{m}}} J_n^2(s) \leq \sup_{s \in \mathcal{B}_{m, \hat{m}}} \frac{R^2 \mathbf{V}_\lambda^2(z_0) \|s\|_\infty^2}{n^2(1-\gamma)^2} \leq \frac{4\psi_1^2 R^2 \mathbf{V}_\lambda^2(z_0)}{n(1-\gamma)^2}. \quad (16)$$

It remains to bound  $\mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right)_+$ . The unit ball  $\mathcal{B}_{m, \hat{m}}$  is random. We can not bound  $I_n^2(s)$  on it, we have to control the risk on the fixed balls  $\mathcal{B}_{m, m'}$ . We can write:

$$\mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right)_+ \leq \sum_{m, m' \in \mathcal{M}_n} \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, m'}} I_n^2(s) - p(m, m') \right)_+. \quad (17)$$

The Markov chain  $(Y_1, \dots, Y_n)$  is exponentially  $\beta$ -mixing with  $\beta$ -mixing coefficient  $\beta_Y(k) \leq c\gamma^k = ce^{-\ln(1/\gamma)k}$ . The following lemma is deduced from the Berbee's coupling lemma and a Talagrand inequality. It is proved in the appendix.

**Lemma 11** (Talagrand's inequality for  $\beta$ -mixing variables). *Let  $Y_1, \dots, Y_n$  be a Markov chain exponentially  $\beta$ -mixing, with  $\beta$ -mixing coefficient  $\beta_Y(k) \leq ce^{-b_0 k}$ . We choose  $q_n := c_q \ln(n)$  with  $c_q \geq 2/b_0$ ,  $p_n = n/(2q_n)$ . We have that  $\beta_Y(q_n) \leq c\gamma^{2\ln(n)} \lesssim n^{-2}$ . Let us consider*

$$I_n(s) = \frac{1}{n} \sum_{k=1}^n s(Y_k) - \mathbb{E}_{z_0}(s(Y_k)).$$

If we can find a triplet  $(M_2, V$  and  $H)$  such that:

$$\begin{aligned} \forall i, \sup_{s \in \mathcal{B}_{m, m'}} \text{Var}_{z_0} \left( \frac{1}{q_n} \sum_{k=i}^{q_n+i} s(Y_k) \right) &\leq \frac{V}{q_n}, \\ \sup_{s \in \mathcal{B}_{m, m'}} \|s\|_\infty \leq M_2 \quad \text{and} \quad \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, m'}} |I_n(s)| \right) &\leq \frac{H}{\sqrt{n}}, \end{aligned}$$

then we have:

$$\begin{aligned} \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, m'}} |I_n^2(s) - 6H^2| \right)_+ &\leq K_1 \frac{V}{n} \exp \left( -k_1 \frac{H^2}{V} \right) + K_2 \frac{M_2^2}{p_n^2} \exp \left( -k_2 \frac{\sqrt{p_n} H}{\sqrt{q_n} M_2} \right) \\ &\quad + 2 \frac{M_2^2}{n^2} \end{aligned}$$

where  $K_1, K_2, k_1$  and  $k_2$  are universal constants.

For the sake of simplicity, let us set  $D = D_m + D_{m'}$  and  $\mathcal{B} = \mathcal{B}_{m,m'}$ . By Assumption A4b,

$$\sup_{s \in \mathcal{B}} \|s\|_\infty \leq \sup_{s \in \mathcal{B}} \psi_1^{1/2} D^{1/2} \|s\|_{L^2(\mathcal{I})} = \psi_1^{1/2} D^{1/2} := M_2.$$

By Lemma 1,

$$\text{Var}_{z_0} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} s(Y_k) \right) \leq \frac{1}{q_n} \int_{\mathcal{I}} s^2(z) \nu(z) dz + \frac{\|s\|_\infty}{q_n} \int_{\mathcal{I}} |s(z)| \nu(z) G_\lambda(z) dz + \frac{c_\lambda \|s\|_\infty^2}{q_n^2}.$$

By Cauchy-Schwarz,

$$\|s^2 \nu\|_{L^1(\mathcal{I})} \leq \|s\|_{L^2(\mathcal{I})} \|s \nu\|_{L^2(\mathcal{I})} \leq \|s\|_{L^2(\mathcal{I})} \|s\|_\infty \|\nu\|_{L^2(\mathcal{I})}$$

and

$$\|s \nu G_\lambda\|_{L^1(\mathcal{I})} \leq \|G_\lambda\|_{L^\infty(\mathcal{I})} \|s\|_{L^2(\mathcal{I})} \|\nu\|_{L^2(\mathcal{I})}.$$

Then

$$\text{Var}_{z_0} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} s(Z_k) \right) \leq \frac{\|s\|_{L^2(\mathcal{I})} \|\nu\|_{L^2(\mathcal{I})}}{q_n} \left( \|s\|_\infty + \|G_\lambda\|_{L^\infty(\mathcal{I})} \right) + c_\lambda \frac{\|s\|_\infty^2}{q_n^2}.$$

By Assumption A4b,  $\|s\|_\infty \leq \psi_1^{1/2} D^{1/2}$ , moreover,  $\sup_{s \in \mathcal{B}} \|s\|_{L^2(\mathcal{I})} = 1$  and then

$$\begin{aligned} \sup_{s \in \mathcal{B}} \text{Var}_{z_0} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} s(Z_k) \right) &\leq \frac{\psi_1^{1/2} D^{1/2} \|\nu\|_{L^2(\mathcal{I})} \left( 1 + \|G_\lambda\|_{L^\infty(\mathcal{I})} \right)}{q_n} + \frac{c_\lambda \psi_1 D}{q_n^2} \\ &\leq \frac{c_1 D^{1/2}}{q_n} + c_2 \frac{D}{q_n^2} := \frac{V}{q_n}. \end{aligned}$$

It remains to find  $H$  such that  $\mathbb{E}_{z_0} (\sup_{s \in \mathcal{B}} |I_n(s)|) \leq H/\sqrt{n}$ . Let us introduce  $(\varphi_l)_{1 \leq l \leq D}$  an orthonormal basis of  $S_m + S_{m'} = S_{\max(m,m')}$  satisfying Assumption A4. Then we can write  $s = \sum_l b_l \varphi_l$ . As the function  $s \rightarrow I_n(s)$  is linear:

$$\sup_{s \in \mathcal{B}} I_n^2(s) = \sup_{\sum b_l^2 = 1} \left( \sum_{l=1}^D b_l I_n(\varphi_l) \right)^2 \leq \sup_{\sum b_l^2 = 1} \left( \sum_{l=1}^D b_l^2 \right) \left( \sum_{l=1}^D I_n^2(\varphi_l) \right) = \sum_{l=1}^D I_n^2(\varphi_l).$$

We can remark that  $I_n(\varphi_l) = \hat{a}_l - \mathbb{E}_{z_0}(\hat{a}_l)$  (see equation (14)) and by consequence,  $\mathbb{E}_{z_0}(I_n^2(\varphi_l)) = \text{Var}_{z_0}(\hat{a}_l)$ . By (12):

$$\begin{aligned} \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}} I_n^2(s) \right) &\leq \sum_{l=1}^D \mathbb{E}_{z_0}(I_n^2(\varphi_l)) = \sum_{l=1}^D \text{Var}_{z_0}(\hat{a}_l) \\ &\leq \frac{(\psi_1 + C_\lambda \psi_2) D}{n} + \frac{c_\lambda \psi_1}{n} := \frac{H^2}{n}. \end{aligned}$$

We can now apply Lemma 11 with

$$M_2 = \psi_1^{1/2} D^{1/2}, \quad V = c_1 D^{1/2} + c_2 D/q_n \quad \text{and} \quad H^2 = (\psi_1 + C_\lambda \psi_2) D + c_\lambda \psi_1.$$

For  $p(m, m') \geq 6(\psi_1 + C_\lambda \psi_2) D/n + 6c_\lambda \psi_1/n$ , we get

$$\begin{aligned} E_1 &:= \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}} I_n^2(s) - p(m, m') \right) \\ &\leq K_1 \frac{V}{n} \exp \left( -k_1 \frac{H^2}{V} \right) + K_2 \frac{M_2^2}{p_n^2} \exp \left( -k_2 \frac{\sqrt{p_n} H}{\sqrt{q_n} M_2} \right) + 2 \frac{M_2^2}{n^2}. \end{aligned}$$

As  $2/(x+y) \geq \min(1/x, 1/y)$ ,

$$\begin{aligned} \exp \left( -k_1 \frac{H^2}{V} \right) &\leq \exp \left( -\frac{k_1}{2} \min \left( \frac{H^2}{c_1 D^{1/2}}, \frac{H^2}{c_2 D/q_n} \right) \right) \\ &\leq \exp \left( -\frac{k_1 H^2}{2c_1 D^{1/2}} \right) + \exp \left( -\frac{k_1 H^2}{2c_2 D q_n} \right) \end{aligned}$$

and therefore

$$\begin{aligned} E_1 &\leq K'_1 \left( \frac{D^{1/2}}{n} + \frac{D}{n q_n} \right) \left( \exp \left( -\frac{k'_1 D}{D^{1/2}} \right) + \exp \left( -\frac{k''_1 D q_n}{D} \right) \right) \\ &\quad + \frac{K'_2 D}{p_n^2} \exp \left( -k'_2 p_n^{1/2} \frac{D^{1/2}}{q_n^{1/2}} \frac{1}{D^{1/2}} \right) + K'_3 \frac{D}{n^2} \\ &\leq K_1^\lambda \left( \frac{D}{n} \exp(-k_1^\lambda D^{1/2}) + \frac{D}{n} \exp(-k_1^\lambda q_n) \right) + K_2^\lambda \frac{D \ln^2(n)}{n^2} \exp \left( -k_2^\lambda \frac{n^{1/2}}{\ln(n)} \right) + K_3^\lambda \frac{D}{n^2} \end{aligned}$$

where  $(K_i^\lambda)_{1 \leq i \leq 3}$  and  $(k_i^\lambda)_{1 \leq i \leq 2}$  depend on  $(\mathbf{V}_\lambda, R, \gamma, (\psi_1, \psi_2), \|\nu\|_{L^2(\mathcal{I})})$ . The second term can be made smaller than  $n^{-2}$  for  $c_q$  large enough. The third is also smaller to  $n^{-2}$  thanks to the exponential term. Then

$$\mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, m'}} I_n^2(s) - p(m, m') \right) \leq K_1^\lambda \frac{D_{m, m'}}{n} \exp(-k_1^\lambda D_{m, m'}^{1/2}) + K_4^\lambda \frac{D_{m, m'}}{n^2}.$$

All the dimensions  $D_{m, m'}$  are different, so  $\sum_{m' \in \mathcal{M}} D_{m, m'} e^{-c D_{m, m'}^{1/2}} \leq \sum_{l=1}^\infty l e^{-c l^{1/2}} < \infty$ . Moreover, as  $\sup_{m' \in \mathcal{M}_n} D_{m, m'} \leq \sqrt{n}$ ,  $\sum_{m' \in \mathcal{M}_n} D_{m, m'} \leq \max_{m' \in \mathcal{M}_n} D_{m, m'}^2 \leq n$ . Then by (17),

$$\mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right) \leq \frac{c}{n}. \quad (18)$$

Collecting (15), (16) and (18), for any  $m \in \mathcal{M}_n$ :

$$\mathbb{E}_{z_0} (\|\hat{\nu}_{\hat{m}} - \nu\|_{L^2}^2) \leq 3 \|\nu_m - \nu\|_{L^2}^2 + 4 \text{pen}(m) + \frac{c}{n}.$$

All the constants involved in the bound of  $J_n^2$  and  $I_n^2$  ( $M_2, H, V$ ) depends on  $\mathbf{V}_\lambda, \gamma, R$  and  $\|\nu\|_{L^2(\mathcal{I})} \leq \|\nu\|_{B_{2,\infty}^\alpha(\mathcal{I})}$ . Then there exists a continuous function  $\psi$  such that  $x \rightarrow \psi(x, v, c, r)$  is increasing and

$$\mathbb{E}_{z_0} \left( \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 \right) \leq 3 \|\nu_m - \nu\|_{L^2(\mathcal{I})}^2 + 4pen(m) + \frac{\psi \left( \|\nu\|_{B_{2,\infty}^\alpha}, \mathbf{V}_\lambda, \gamma, R \right)}{n}.$$

### 4.3 Proof of Theorem 8

Let us first control  $\mathbb{E}_{z_0} \left( (\hat{\mathbf{D}}_n(y) - \mathbf{D}(y))^2 \right)$ . As  $\phi$  is a diffeomorphism, the function  $(\phi_x^{-1})'(y)$  is bounded on  $[0, i_2] \times \mathcal{I}$ . The function  $s_{x,z}(y) = (\phi_x^{-1})'(y) \mathbb{1}_{\{x \leq y \leq z\}}$  is bounded by a constant on  $\mathcal{I}$ :

$$\|s_{x,z}\|_{L^\infty(\mathcal{I})} \leq \sup_{x \in [0, i_2], y \in \mathcal{I}} (\phi_x^{-1})'(y) := \Phi_1. \quad (19)$$

We have that

$$\mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) = \mathbb{E}_{z_0} \left( \left( \frac{1}{n} \sum_{k=1}^n s_{Z_{k-1}, Y_k}(y) - \mathbb{E}_\xi (s_{Z_0, Y_1}(y)) \right)^2 \right)$$

with  $\xi$  the stationary density of  $(Z_{k-1}, Y_k)$  introduced in Assumption A2. By Lemma 1, we have

$$\begin{aligned} \mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) &\leq \frac{1}{n} \left( \Phi_1^2 + \Phi_1^2 \int G_\lambda(z) \mu(dz) \right) + \frac{\Phi_1^2 c_\lambda}{n^2} + \Phi_1^2 \frac{R^2 \mathbf{V}_\lambda^2(z_0)}{n^2 (1 - \gamma)^2} \\ &\leq \frac{\Phi_1^2 (1 + C_\lambda)}{n} + \frac{\Phi_1^2 (c_\lambda + R^2 V_\lambda^2(z_0) / (1 - \gamma)^2)}{n^2} \end{aligned}$$

and therefore

$$\mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) \leq \frac{\Phi_1^2}{n} (1 + C_\lambda) + \frac{c}{n^2}. \quad (20)$$

For  $n$  large enough,  $1/\ln(n)$  is smaller than  $D_0/2$  ( $D_0$  is defined in Assumption A3) and then by Markov inequality,

$$\mathbb{P} \left( \hat{\mathbf{D}}_n(y) \leq 1/\ln(n) \right) \leq \mathbb{P} \left( |\hat{\mathbf{D}}_n(y) - \mathbf{D}(y)| \geq D_0/2 \right) \leq \frac{4}{D_0^2} \mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right). \quad (21)$$

As  $\nu$  is a positive function,  $|\hat{\lambda}_n(y) - \lambda(y)| \mathbb{1}_{\{\hat{\nu}_{\hat{m}}(y) \geq 0\}} \leq |\hat{\lambda}_n(y) - \lambda(y)|$  and therefore, according to the definition of the estimator  $\hat{\lambda}_n$  (see (11)),

$$|\hat{\lambda}_n(y) - \lambda(y)| \leq \left| \frac{\hat{\nu}_{\hat{m}}(y)}{\hat{\mathbf{D}}_n(y)} - \frac{\nu(y)}{\mathbf{D}(y)} \right| \mathbb{1}_{\{\hat{\mathbf{D}}_n(y) \geq 1/\ln(n)\}} + \lambda(y) \mathbb{1}_{\{\hat{\mathbf{D}}_n(y) \leq 1/\ln(n)\}}.$$

We can write:

$$\left| \frac{\hat{\nu}_{\hat{m}}(y)}{\hat{\mathbf{D}}_n(y)} - \frac{\nu(y)}{\mathbf{D}(y)} \right| \leq \left| \frac{\hat{\nu}_{\hat{m}}(y) - \nu(y)}{\hat{\mathbf{D}}_n(y)} + \frac{\nu(y)}{\hat{\mathbf{D}}_n(y)\mathbf{D}(y)} (\mathbf{D}(y) - \hat{\mathbf{D}}_n(y)) \right|.$$

As  $\mathbf{D} \geq D_0$  by Assumption A3:

$$|\hat{\lambda}_n(y) - \lambda(y)| \leq \ln(n) (|\hat{\nu}_{\hat{m}}(y) - \nu(y)|) + \ln(n) \frac{|\hat{\mathbf{D}}_n(y) - \mathbf{D}(y)|}{D_0} \nu(y) + \lambda(y) \mathbb{1}_{\{\hat{\mathbf{D}}_n(y) \leq 1/\ln(n)\}}.$$

By (20) and (21),

$$\begin{aligned} \mathbb{E}_{z_0} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) &\leq 3 \ln^2(n) \mathbb{E}_{z_0} \left( \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 \right) \\ &\quad + 3D_0^{-2} \ln^2(n) \int_{\mathcal{I}} \mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) \nu^2(y) dy \\ &\quad + 12D_0^{-2} \int_{\mathcal{I}} \mathbb{E}_{z_0} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) \lambda^2(y) dy \\ &\leq 3 \ln^2(n) \mathbb{E}_{z_0} \left( \|\hat{\nu}_{\hat{m}} - \nu\|_{L^2(\mathcal{I})}^2 \right) + c'_\lambda \frac{\ln^2(n)}{n} \end{aligned}$$

with  $c'_\lambda = \frac{\Phi_1^2}{D_0^2} (4 + 2C_\lambda) (3 \|\nu\|_{L^2(\mathcal{I})}^2 + 12 \|\lambda\|_{L^2(\mathcal{I})}^2)$ .

#### 4.4 Proof of Theorem 10

We use the reduction scheme described in Tsybakov [30, chapter 2]. By Markov inequality,

$$C'^2 n^{-2\alpha/(2\alpha+1)} \mathbb{E}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \geq \mathbb{P}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})} \geq C' n^{-\alpha/(2\alpha+1)} \right).$$

Our aim is to show that

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{P}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})} \geq C' n^{-\alpha/(2\alpha+1)} \right) > 0.$$

Instead of searching an infimum on the whole class  $\mathcal{E}(\mathfrak{s}, b, \alpha)$ , we can limit ourselves to the finite set  $\{\lambda_0, \dots, \lambda_{P_n}\} \in \mathcal{E}(\mathfrak{s}, b, \alpha)$ , such that

$$\|\lambda_i - \lambda_j\|_{L^2(\mathcal{I})} \geq 2C' n^{-\alpha/(2\alpha+1)} \mathbb{1}_{\{i \neq j\}}. \quad (22)$$

Then

$$\begin{aligned} E_2 &:= \inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{P}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})} \geq C' n^{-\frac{\alpha}{2\alpha+1}} \right) \\ &\geq \inf_{\hat{\lambda}_n} \max_j \mathbb{P}^{\lambda_j} \left( \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})} \geq C' n^{-\frac{\alpha}{2\alpha+1}} \right). \end{aligned}$$

We note  $\psi^*$  the predictor

$$\psi^* := \arg \min_{0 \leq j \leq P_n} \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})}.$$

By the triangular inequality,  $\left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})} \geq \left\| \lambda_{\psi^*} - \lambda_j \right\|_{L^2(\mathcal{I})} - \left\| \lambda_{\psi^*} - \hat{\lambda}_n \right\|_{L^2(\mathcal{I})}$ .

Consequently, as  $\left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})} \geq \left\| \hat{\lambda}_n - \lambda_{\psi^*} \right\|_{L^2(\mathcal{I})}$ ,

$$\left\{ \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})} \geq A_n \right\} \supseteq \left\{ \left\| \lambda_{\psi^*} - \hat{\lambda}_n \right\|_{L^2(\mathcal{I})} \geq A_n \right\} \cup \left\{ \left\| \lambda_{\psi^*} - \lambda_j \right\|_{L^2(\mathcal{I})} \geq 2A_n \right\}.$$

By (22),  $\left\| \lambda_{\psi^*} - \lambda_j \right\|_{L^2(\mathcal{I})} \geq 2C'n^{-\alpha/(2\alpha+1)} \mathbb{1}_{\{\psi^* \neq j\}}$ . Then setting  $A_n = C'n^{-\alpha/(2\alpha+1)}$ ,

$\left\{ \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2(\mathcal{I})} \geq C'n^{-\alpha/(2\alpha+1)} \right\} \supseteq \{\psi^* \neq j\}$  and therefore:

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{P}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \geq C'n^{-\alpha/(2\alpha+1)} \right) \geq \inf_{\hat{\lambda}_n} \max_j \mathbb{P}^{\lambda_j} (\psi^* \neq j).$$

We denote by  $\mathbf{P}^{\lambda_j}$  the law of  $(Z_0, Y_1, Z_1, \dots, Y_n, Z_n)$  under  $\lambda_j$ . The following lemma is exactly Theorem 2.5 of Tsybakov [30].

**Lemma 12.** *Let us consider a series of functions  $\lambda_0, \dots, \lambda_{P_n}$  such that:*

a. *The functions  $\lambda_i$  are sufficiently apart:  $\forall i \neq j$*

$$\left\| \lambda_i - \lambda_j \right\|_{L^2(\mathcal{I})} \geq 2C'n^{-\alpha/(2\alpha+1)}.$$

b. *For all  $i$ , the function  $\lambda_i$  belongs to the subspace  $\mathcal{E}(\mathfrak{s}, b, \alpha)$ .*

c. *Absolute continuity:  $\forall 1 \leq j \leq P_n, \mathbf{P}^{\lambda_j} \ll \mathbf{P}^{\lambda_0}$ .*

d. *The distance between the measures of probabilities is not too large:*

$$\frac{1}{P_n} \sum_{j=1}^{P_n} \chi^2(\mathbf{P}^{\lambda_j}, \mathbf{P}^{\lambda_0}) \leq c \ln(P_n)$$

*with  $0 < c < 1/8$ , and  $\chi^2(., .)$  the  $\chi$ -square divergence.*

Then

$$\begin{aligned} \inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} C'^2 n^{-2\alpha/(2\alpha+1)} \mathbb{E}^\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) &\geq \inf_{\hat{\lambda}_n} \max_j \mathbb{P}^{\lambda_j} (\psi^* \neq j) \\ &\geq \frac{\sqrt{P_n}}{1 + \sqrt{P_n}} \left( 1 - 2c - 2\sqrt{\frac{c}{\ln(P_n)}} \right) \\ &> 0. \end{aligned}$$

**Step 1: Construction of  $(\lambda_0, \dots, \lambda_{P_n})$ .** Let us set

$$\lambda_0(x) = \varepsilon \mathbb{1}_{\{i_1 \leq x \leq j_2\}} + \mathbf{a} \frac{x^b}{(b+1)\mathbf{m}(x)} \mathbb{1}_{\{x > j_2\}} \quad \text{where} \quad \varepsilon = \max_{x \in [i_1, j_2]} \mathbf{a} \frac{x^b}{(b+1)\mathbf{m}(x)}$$

with  $\mathcal{J} = [j_1, j_2]$  defined in (7). As  $\lambda_0$  is constant on  $\mathcal{J}$ , this function belongs to the Hölder space  $H^\alpha(\mathcal{J})$  and  $\|\lambda_0\|_{H^\alpha(\mathcal{J})} = \varepsilon$  (see Appendix A.4 for the definition of the Hölder space). It remains to ensure that it belongs to  $\mathcal{E}(\mathfrak{s}, b, \alpha)$ . If  $\varepsilon > \mathbf{L}$ , then  $\mathcal{E}(\mathfrak{s}, b, \alpha) = \emptyset$ . If  $\mathbf{L} = \varepsilon$ , then any function  $\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)$  satisfies:  $\forall x \in [i_1, j_2], \lambda(x) = \lambda_0(x)$ . Let us assume that  $\varepsilon < \mathbf{L}$ : in that case, there exists  $\delta > 0$  such that  $\|\lambda_0\|_{H^\alpha(\mathcal{J})} \leq \mathbf{L} - \delta$ .

We consider a non-negative function  $K \in H^\alpha(\mathbb{R})$ , bounded, with support in  $[0, j_2 - i_1[$  and such that  $\|K\|_{L^1} \leq 1$ . We set  $h_n = n^{-1/(2\alpha+1)}$ ,  $p_n = \lceil 1/h_n \rceil$  and, for  $0 \leq k \leq p_n - 1$ ,  $x_k = i_1 + h_n k(j_2 - i_1)$ . We consider the functions  $\varphi_k(x) := ah_n^\alpha K((x - x_k)/h_n)$  with  $a < 1$ . The functions  $\varphi_k$  have support in  $[x_k, x_{k+1}) \subset \mathcal{J}$ . Moreover, by a change of variable  $y = (x - x_k)/h_n$ ,  $\|\varphi_k\|_{L^1} = ah_n^{\alpha+1} \|K\|_{L^1} \leq ah_n^{\alpha+1}$  and  $\|\varphi_k\|_{L^2}^2 = a^2 h_n^{2\alpha+1} \|K\|_{L^2}^2$ . We consider the set of functions

$$\mathcal{G}_n := \left\{ \lambda_\epsilon := \lambda_0 + \sum_{k=0}^{p_n-1} \epsilon_k \varphi_k, \quad (\epsilon_k) \in \{0, 1\}^{p_n} \right\}.$$

The cardinal of  $\mathcal{G}_n$  is  $2^{p_n}$ . For two vectors  $(\epsilon, \eta)$  with values in  $\{0, 1\}^{p_n}$ , the distance between two functions  $\lambda_\epsilon$  and  $\lambda_\eta$  is:

$$\|\lambda_\epsilon - \lambda_\eta\|_{L^2}^2 = a^2 h_n^{2\alpha+1} \|K\|_{L^2}^2 \sum_{k=1}^{p_n} (\epsilon_k - \eta_k)^2. \quad (23)$$

As the series  $\epsilon_k$  and  $\eta_k$  have values in  $\{0, 1\}$ , the quantity

$$\rho(\epsilon, \eta) := \sum_{k=1}^{p_n} \mathbb{1}_{\{\epsilon_k \neq \eta_k\}} = \sum_{k=1}^{p_n} (\epsilon_k - \eta_k)^2$$

is the Hamming distance between  $\eta$  and  $\epsilon$ . To apply Lemma 12, we need that,  $\forall \eta \neq \epsilon$ ,

$$\|\lambda_\epsilon - \lambda_\eta\|_{L^2}^2 \geq 4C'^2 h_n^{2\alpha} \text{ and consequently } \rho(\epsilon, \eta) \geq Ch_n^{-1}.$$

This is not the case if we take the whole  $\mathcal{G}_n$  (the minimal Hamming distance between two vectors  $\epsilon$  and  $\eta$  is 1). We need to extract a sub-series of functions. According to Tsybakov [30, Lemma 2.7] (bound of Varshamov-Gilbert), it is possible to extract a family  $(\epsilon_{(0)}, \dots, \epsilon_{(P_n)})$  of the set  $\Omega = \{0, 1\}^{p_n}$  such that  $\epsilon_{(0)} = (0, \dots, 0)$  and

$$\forall 0 \leq j < k \leq P_n, \quad \rho(\epsilon_{(j)}, \epsilon_{(k)}) \geq p_n/8, \quad \text{and} \quad P_n \geq 2^{p_n/8}.$$

As  $p_n \geq n^{1/(2\alpha+1)}$ ,

$$\ln(P_n) \geq \ln(2)n^{1/(2\alpha+1)}/8. \quad (24)$$

We define

$$\lambda_j := \lambda_{\epsilon_{(j)}} \quad \text{and} \quad \mathcal{H}_n = \{\lambda_0, \lambda_1, \dots, \lambda_{P_n}\}.$$

Then, for any  $\lambda_j, \lambda_k \in \mathcal{H}_n$ , if  $j \neq k$ , as  $p_n = \lceil 1/h_n \rceil$ , by (23),

$$\|\lambda_j - \lambda_k\|_{L^2}^2 \geq a^2 \|K\|_{L^2}^2 h_n^{2\alpha+1} p_n / 8 \geq a^2 \|K\|_{L^2}^2 h_n^{2\alpha} / 8.$$

This is exactly the expected lower bound if we take  $C' = a \|K\|_{L^2} / (4\sqrt{2})$ .

**Step 2: Functions  $\lambda_j$  belong to  $\mathcal{E}(\mathfrak{s}, b, \alpha)$ .** We already know that  $\lambda_0$  belongs to  $\mathcal{E}(\mathfrak{s}, b, \alpha)$ . Let us first compute the norm of  $\lambda_j$  on  $H^\alpha(\mathcal{J})$ . Let us set  $\mathbf{r} = \lfloor \alpha \rfloor$ . We have that  $(K(\cdot/h_n))^{(\mathbf{r})} = h_n^{-\mathbf{r}} K^{(\mathbf{r})}(\cdot/h_n)$ . We compute the modulus of smoothness:

$$\begin{aligned} \omega(\varphi_k^{(\mathbf{r})}, t)_\infty &= a \omega \left( h_n^\alpha \left( K \left( \frac{\cdot - x_k}{h_n} \right) \right)^{(\mathbf{r})}, t \right)_\infty = a h_n^\alpha \omega \left( h_n^{-\mathbf{r}} K^{(\mathbf{r})} \left( \frac{\cdot - x_k}{h_n} \right), t \right)_\infty \\ &= a h_n^{\alpha-\mathbf{r}} \omega \left( K^{(\mathbf{r})}, \frac{t}{h_n} \right)_\infty \end{aligned}$$

and

$$\begin{aligned} |\varphi_k|_{H^\alpha} &= \sup_{t>0} t^{\mathbf{r}-\alpha} \omega(\varphi_k^{(\mathbf{r})}, t)_\infty = a \sup_{t>0} t^{\mathbf{r}-\alpha} h_n^{\alpha-\mathbf{r}} \omega \left( K^{(\mathbf{r})}, \frac{t}{h_n} \right)_\infty \\ &= a \sup_{z>0} z^{\mathbf{r}-\alpha} \omega(K^{(\mathbf{r})}, z) = a |K|_{H^\alpha} \end{aligned}$$

by the change of variable  $z = t/h_n$ . The functions  $\varphi_k$  have disjoint supports. For any  $(x, y) \in \mathcal{J}$ , there exists  $(i, j)$  such that  $x \in [x_i, x_{i+1}(\cdot)$  and  $y \in [x_j, x_{j+1}(\cdot)$ . Then

$$\lambda_k^{(\mathbf{r})}(x) - \lambda_k^{(\mathbf{r})}(y) = \varepsilon_i \left( \varphi_i^{(\mathbf{r})}(x) - \varphi_i^{(\mathbf{r})}(y) \right) + \varepsilon_j \left( \varphi_j^{(\mathbf{r})}(x) - \varphi_j^{(\mathbf{r})}(y) \right).$$

Therefore

$$\omega(\lambda_k^{(\mathbf{r})}, t)_\infty \leq \sup_{i,j} \left( \omega(\varphi_j^{(\mathbf{r})}, t)_\infty + \omega(\varphi_i^{(\mathbf{r})}, t)_\infty \right) \leq 2\omega(\varphi_1^{(\mathbf{r})}, t)_\infty$$

and  $|\lambda_k|_{H^\alpha(\mathcal{J})} \leq 2a|K|_{H^\alpha}$ . Moreover,

$$\|\lambda_k\|_{L^\infty(\mathcal{J})} \leq \|\lambda_0\|_{L^\infty(\mathcal{J})} + a h_n^\alpha \|K\|_{L^\infty} \leq \|\lambda_0\|_{L^\infty(\mathcal{J})} + 2a \|K\|_{L^\infty}$$

and consequently  $\|\lambda_k\|_{H^\alpha(\mathcal{J})} \leq \|\lambda_0\|_{H^\alpha(\mathcal{J})} + 2a \|K\|_{H^\alpha}$ . Then  $\lambda_k \in H^\alpha(\mathcal{J}, \mathbf{L})$  for  $a$  sufficiently small. It remains to check that  $\lambda_k \in \mathcal{E}(\mathfrak{s}, b, \alpha)$ . For any  $0 \leq k \leq P_n$ :

a. As  $K$  is non-negative,  $\forall x \geq i_1$ ,  $\lambda_k(x) \geq \mathbf{a} \frac{x^b}{(b+1)\mathbf{m}(x)}$ .

b.  $\|\lambda_k\|_{H^\alpha(\mathcal{J})} \leq \mathbf{L}$  for  $a$  small enough.

c.  $\int_0^{i_1} \lambda_k(u) \mathbf{M}(u) du = 0 \leq \mathbf{1}$ .

Therefore  $\lambda_k \in \mathcal{E}(\mathfrak{s}, b, \alpha)$  for  $a$  small enough.

**Step 3: Absolute continuity.** We denote by  $\mathcal{P}_j$  the transition densities  $\mathcal{P}_{\lambda_j}$ . As  $(Z_0, Y_1, Z_1, \dots, Y_n, Z_n)$  is a Markov process,

$$\mathbf{P}^{\lambda_j}(z_0, dy_1, dz_1, \dots, dy_n, dz_n) = \mathcal{P}_j(z_0, y_1)Q(y_1, dz_1) \dots \mathcal{P}_j(z_{n-1}, y_n)Q(y_n, dz_n)dy_1 \dots dy_n.$$

By (3), we can rewrite:  $\mathcal{P}_0(x, y) = A_{x,y} \exp(-\tilde{A}_{x,y})$  where

$$A_{x,y} := \lambda_0(y)(\phi_x^{-1})'(y)\mathbb{1}_{\{y \geq x\}}, \quad \tilde{A}_{x,y} := \int_x^y \lambda_0(u)(\phi_x^{-1})'(u)du \quad (25)$$

and  $\mathcal{P}_j(x, y) = (A_{x,y} + B_{x,y}) \exp(-\tilde{A}_{x,y} - \tilde{B}_{x,y})$  where  $B_{x,y} = \sum_{k=1}^m \epsilon_k B_{x,y}^k$ ,  $\tilde{B}_{x,y} = \sum_{k=1}^{p_n} \epsilon_k \tilde{B}_{x,y}^k$  and

$$B_{x,y}^k := \varphi_k(x)(\phi_x^{-1})'(y)\mathbb{1}_{\{y \geq x\}}, \quad \tilde{B}_{x,y}^k := \int_x^y \varphi_k(u)(\phi_x^{-1})'(u)du. \quad (26)$$

The probability density  $\mathbf{P}^{\lambda_0}$  is null if one of the  $Q(y_i, dz_i)$  is null, if one of the indicator function  $\mathbb{1}_{\{y_{i+1} \geq z_i\}} = 0$ , or if one  $y_i$  is smaller than  $i_1$ ; then  $\mathbf{P}^{\lambda_j}$  is absolutely continuous with respect to  $\mathbf{P}^{\lambda_0}$ .

**Step 4: The  $\chi^2$  divergence.** As  $\mathbf{P}^{\lambda_0}, \mathbf{P}^{\lambda_j}$  are equivalent measures, we have:

$$\chi^2(\mathbf{P}^{\lambda_j}, \mathbf{P}^{\lambda_0}) = \int \left( \frac{d\mathbf{P}^{\lambda_j}}{d\mathbf{P}^{\lambda_0}} \right)^2 d\mathbf{P}^{\lambda_0} - 1.$$

Let us set  $E_3 := \chi^2(\mathbf{P}^{\lambda_j}, \mathbf{P}^{\lambda_0}) + 1$ . We can write:

$$E_3 = \int_{(\mathbb{R}^+)^n} \left( \frac{\mathcal{P}_j(z_0, y_1) \dots \mathcal{P}_j(z_{n-1}, y_n)}{\mathcal{P}_0(z_0, y_1) \dots \mathcal{P}_0(z_{n-1}, y_n)} \right)^2 \mathcal{P}_0(z_0, y_1) \dots \mathcal{P}_0(z_{n-1}, y_n) \\ \times \int_{(\mathbb{R}^+)^n} Q(y_1, dz_1) \dots Q(y_n, dz_n) dy_1 \dots dy_n.$$

As  $Q$  is the transition density, for any  $y_n$ ,  $\int_{\mathbb{R}^+} Q(y_n, dz_n) = 1$ . Moreover, as  $\int_{\mathbb{R}^+} \mathcal{P}_0(x, y) dy = \int_{\mathbb{R}^+} \mathcal{P}_j(x, y) dy = 1$ ,

$$E_3 = \int_{(\mathbb{R}^+)^{2(n-1)}} \frac{(\mathcal{P}_j(z_0, y_1) \dots \mathcal{P}_j(z_{n-2}, y_{n-1}))^2}{\mathcal{P}_0(z_0, y_1) \dots \mathcal{P}_0(z_{n-2}, y_{n-1})} Q(y_1, dz_1) \dots Q(y_{n-1}, dz_{n-1}) dy_1 \dots dy_{n-1} \\ \times \int_{\mathbb{R}^+} \frac{(\mathcal{P}_j(z_{n-1}, y_n))^2}{\mathcal{P}_0(z_{n-1}, y_n)} dy_n. \quad (27)$$

This expression of the  $\chi^2$  divergence enables us to approximate it more closely. Let us set

$$DP := \int_{\mathbb{R}^+} \frac{(\mathcal{P}_j(x, y))^2}{\mathcal{P}_0(x, y)} dy - 1 = \int_{\mathbb{R}^+} \left( \frac{\mathcal{P}_j(x, y)}{\mathcal{P}_0(x, y)} - 1 \right)^2 \mathcal{P}_0(x, y) dy \\ = \int_{\mathbb{R}^+} \left( \left( 1 + \frac{B_{x,y}}{A_{x,y}} \right) \exp(-\tilde{B}_{x,y}) - 1 \right)^2 A_{x,y} \exp(-\tilde{A}_{x,y}) dy. \quad (28)$$

As the support of  $\varphi_k$  is included in  $\mathcal{J}$ , we can remark that  $B_{x,y}$  is null on  $\mathcal{J}^c$  and

$$\tilde{B}_{x,y}^k = \int_{[x,y] \cap \mathcal{J}} \varphi_k(u) (\phi_x^{-1})'(u) du.$$

We bound the  $\chi^2$ -divergence differently on  $\mathcal{J}$  and  $\mathcal{J}^c$ :  $DP = R_1 + R_2$  where

$$\begin{aligned} R_1 &:= \int_{\mathcal{J}} \left( \left( 1 + \frac{B_{x,y}}{A_{x,y}} \right) \exp(-\tilde{B}_{x,y}) - 1 \right)^2 A_{x,y} \exp(-\tilde{A}_{x,y}) dy, \\ R_2 &:= \int_{\mathcal{J}^c} \left( \exp(-\tilde{B}_{x,y}) - 1 \right)^2 A_{x,y} \exp(-\tilde{A}_{x,y}) dy. \end{aligned}$$

We have that  $B_{x,y}^k \leq \mathbf{M}(y) \|\varphi_k\|_{\infty} \mathbb{1}_{\{y \geq x\}} \mathbb{1}_{\{y \in [x_k, x_{k+1}]\}}$  and therefore, as  $\|\varphi_k\|_{\infty} = ah_n^{\alpha} \|K\|_{\infty}$ ,

$$B_{x,y} \leq \sup_{y \in \mathcal{J}} \mathbf{M}(y) ah_n^{\alpha} \|K\|_{\infty} \mathbb{1}_{\{y \in \mathcal{J}\}} \leq Cah_n^{\alpha} \mathbb{1}_{\{y \in \mathcal{J}\}}. \quad (29)$$

By (26), we obtain, as the functions  $\varphi_k$  are supported in  $\mathcal{J}$ :

$$\tilde{B}_{x,y}^k = \int_x^y \varphi_k(z) (\phi_x^{-1})'(z) dz \leq \sup_{z \in \mathcal{J}} (\mathbf{M}(z)) \sup_k \|\varphi_k\|_{L^1} \leq ah_n^{\alpha+1} \sup_{z \in \mathcal{J}} (\mathbf{M}(z))$$

and, as  $p_n = \lceil 1/h_n \rceil$ ,

$$\tilde{B}_{x,y} \leq \sum_{k=1}^{p_n} \tilde{B}_{x,y}^k \leq C' ap_n h_n^{\alpha+1} \leq C' ah_n^{\alpha}. \quad (30)$$

Then by (30) and as  $\int_{\mathbb{R}^+} A_{x,y} \exp(-\tilde{A}_{x,y}) dy = 1$

$$R_2 \leq \int_{\mathbb{R}^+} O(a^2 h_n^{2\alpha}) A_{x,y} \exp(-\tilde{A}_{x,y}) dy = O(a^2 h_n^{2\alpha}).$$

As  $\lambda_0 = \varepsilon$  on  $\mathcal{J}$ , we get by (25) that

$$\varepsilon \mathbb{1}_{\{y \geq x\}} \inf_{y \in \mathcal{J}} \mathbf{m}(y) \leq \sup_{y \in \mathcal{J}} A_{x,y} \leq \sup_{y \in \mathcal{J}} \mathbf{M}(y) \varepsilon \mathbb{1}_{\{y \geq x\}}.$$

Moreover, on  $\mathbb{R}^+$ ,  $\exp(-\tilde{A}_{x,y}) \leq 1$ . Then by (29) and (30), we get that

$$R_1 = \int_{\mathcal{J}} \left( (1 + O(ah_n^{\alpha}) \exp(-O(ah_n^{\alpha})) - 1)^2 O(1) \right) dy = \int_{\mathcal{J}} O(a^2 h_n^{2\alpha}) dy = O(a^2 h_n^{2\alpha}).$$

Therefore  $DP = O(a^2 h_n^{2\alpha})$  and, by (27) and (28), we get by recurrence

$$\begin{aligned} E_3 &= \int_{(\mathbb{R}^+)^{2(n-1)}} \frac{(\mathcal{P}_j(z_0, y_1) \dots \mathcal{P}_j(z_{n-2}, y_{n-1}))^2}{\mathcal{P}_0(z_0, y_1) \dots \mathcal{P}_0(z_{n-2}, y_{n-1})} Q(y_1, dz_1) \dots Q(Y_{n-1}, dz_{n-1}) dy_1 \dots dy_{n-1} \\ &\times (O(a^2 h_n^{2\alpha}) + 1) \\ &= \prod_{i=1}^n (O(a^2 h_n^{2\alpha}) + 1) = 1 + a^2 n O(h_n^{2\alpha}). \end{aligned}$$

As  $h_n = n^{-\frac{1}{2\alpha+1}}$ ,

$$\chi^2(\mathbf{P}^{\lambda_0}, \mathbf{P}^{\lambda_j}) = E_3 - 1 \leq a^2 O(n^{1/(2\alpha+1)}).$$

By (24),  $\ln(P_n) \geq \ln(2)n^{1/(2\alpha+1)}/8$  and therefore,

$$\frac{1}{P_n} \sum_{k=1}^{P_n} \chi^2(\mathbf{P}^{\lambda_0}, \mathbf{P}^{\lambda_j}) = a^2 O(n^{1/(2\alpha+1)}) = a^2 O(\ln(P_n)) \leq \ln(P_n)/8$$

for  $a$  small enough, which concludes the proof.

## Acknowledgements

N. Krell was partly supported by the Agence Nationale de la Recherche PIECE 12-JS01-0006-01. The research of E. Schmisser was supported in part by the Labex CEMPI (ANR-11-LABX-0007-01)

## References

- [1] S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10:245–279, 2009.
- [2] R. Azaïs and A. Genadot, A new characterization of the jump rate for piecewise-deterministic Markov processes with discrete transitions, *Comm. Statist. Theory Methods*, 47, (8):1812–1829, 2018
- [3] R. Azaïs and A. Muller-Gueudin. Optimal choice among a class of nonparametric estimators of the jump rate for piecewise-deterministic Markov processes. *Electron. J. Stat.*, 10(2):3648–3692, 2016. ISSN 1935-7524.
- [4] R. Azaïs, J.-B. Bardet, A. Génadot, N. Krell, and P.-A. Zitt. Piecewise deterministic Markov process—recent results. In *Journées MAS 2012*, volume 44 of *ESAIM Proc.*, pages 276–290. EDP Sci., Les Ulis, 2014.
- [5] R. Azaïs, F. Dufour and A. Gégout-Petit, Non-parametric estimation of the conditional distribution of the interjumping times for piecewise-deterministic Markov processes, *Scandinavian Journal of Statistics. Theory and Applications*, 41, (4):950–969, 2014.
- [6] A. Barron, L. Birgé, and P. Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999. ISSN 0178-8051.
- [7] P. Baxendale Renewal theory and computable convergence rates for geometrically ergodic Markov chains, *The Annals of Applied Probability*, 15, (1B):700–738, 2005.

- [8] L. Birgé and P. Massart. Minimum contrast estimators on sieves: exponential bounds and rates of convergence. *Bernoulli*, 4(3):329–375, 1998. ISSN 1350-7265.
- [9] P. Blanchard and A. Jadczyk, Events and piecewise deterministic dynamics in event-enhanced quantum theory, *Physics Letters. A*, 203(5-6):260–266, 1995
- [10] B. Cloez, R. Dossalles, A. Genadot, F. Malrieu, A. Marguet and R. Yvinec, Probabilistic and piecewise deterministic models in biology, *Journées MAS 2016 de la SMAI—Phénomènes complexes et hétérogènes, ESAIM Proc. Surveys*, (60):225–245, 2017.
- [11] F. Comte and F. Merlevède. Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist.*, 6:211–238 (electronic), 2002. New directions in time series analysis (Luminy, 2001).
- [12] F. Comte, V. Genon-Catalot, and Y. Rozenholc. Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543, 2007.
- [13] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993. ISBN 0-412-31410-X.
- [14] B. de Saporta, F. Dufour and H. Zhang, *Numerical methods for simulation and optimization of piecewise deterministic Markov processes*, Mathematics and Statistics Series, Application to reliability, ISTE, London; John Wiley & Sons, Inc., Hoboken, NJ, 2016.
- [15] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. ISBN 3-540-50627-6.
- [16] M. Doumic, M. Hoffmann, N. Krell, and L. Robert. Statistical estimation of a growth-fragmentation model observed on a genealogical tree. *Bernoulli*, 21(3):1760–1799, 2015. ISSN 1350-7265.
- [17] V. Dumas, F. Guillemin, and P. Robert. A Markovian analysis of additive-increase multiplicative-decrease algorithms. *Adv. in Appl. Probab.*, 34(1):85–111, 2002. ISSN 0001-8678.
- [18] M. Farid and M. Davis, Optimal consumption and exploration: a case study in piecewise-deterministic Markov modelling, Optimal control and differential games (Vienna, 1997), *Annals of Operations Research*, 88:121–137, 1999.
- [19] T. Fujii. Nonparametric estimation for a class of piecewise-deterministic Markov processes. *J. Appl. Probab.*, 50(4):931–942, 2013. ISSN 0021-9002.
- [20] F. Guillemin, P. Robert, and B. Zwart. AIMD algorithms and exponential functionals. *Ann. Appl. Probab.*, 14(1):90–117, 2004. ISSN 1050-5164.

- [21] P. Hodara, Krell, and E. Löcherbach, Non-parametric estimation of the spiking rate in systems of interacting neurons, *Statistical Inference for Stochastic Processes An International Journal Devoted to Time Series Analysis and the Statistics of Continuous Time Processes and Dynamical Systems*, 21,(1):–111, 2018
- [22] , R. Höpfner, and K. Brodda, A stochastic model and a functional central limit theorem for information processing in large systems of neurons, *Journal of Mathematical Biology*, 52(4):439–457, 2006.
- [23] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005. ISSN 0091-1798.
- [24] N. Krell. Statistical estimation of jump rates for a piecewise deterministic Markov processes with deterministic increasing motion and jump mechanism. *ESAIM Probab. Stat.*, 20:196–216, 2016. ISSN 1292-8100.
- [25] P. Laurençot and B. Perthame. Exponential decay for the growth-fragmentation/cell-division equation. *Commun. Math. Sci.*, 7(2):503–510, 2009. ISSN 1539-6746.
- [26] L. Robert, M. Hoffmann, N. Krell, S. Aymerich, and J. Robert, and M. Doumic, . Division Control in Escherichia coli is Based on a Size-sensing rather than Timing Mechanism. *MC Biology, BioMedCentral.*, 12:17, 2015.
- [27] Y. Meyer. *Ondelettes et opérateurs. I.* Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, 1990. ISBN 2-7056-6125-0. Ondelettes. [Wavelets].
- [28] , V. Renault, M. Thieullen and E. Trélat, *Optimal control of infinite-dimensional piecewise deterministic Markov processes and application to the control of neuronal dynamics via optogenetics*, *Networks and Heterogeneous Media*, 12(2):417–459, 2017.
- [29] R. Rudnicki and M. Tyran-Kamińska. Piecewise deterministic Markov processes in biological models. In *Semigroups of operators—theory and applications*, volume 113 of *Springer Proc. Math. Stat.*, pages 235–255. Springer, Cham, 2015.
- [30] A. B. Tsybakov. *Introduction à l’estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004. ISBN 3-540-40592-5.
- [31] G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 1997.

# A Technical proofs and results

## A.1 Proof of Lemma 1

We consider a function  $s$  such that  $\|s\|_\infty = 1$ ; we obtain the expected result by dividing  $s$  by its  $L^\infty$ -norm. According to Assumption A2,

$$\left| \mathbb{E}_{z_0} \left( \frac{1}{n} \sum_{k=1}^n s(Y_k, Z_k) - \int s(y, z) \rho(dy, dz) \right) \right| \leq \frac{1}{n} \sum_{k=1}^n R\mathbf{V}_\lambda(z_0) \gamma^k \leq \frac{R\mathbf{V}_\lambda(z_0)}{n(1-\gamma)}$$

which proves the first inequality. Let us set  $\tilde{s}(Y, Z) = s(Y, Z) - \mathbb{E}_{z_0}(s(Y, Z))$ . We have:

$$\mathbb{E}_{z_0} \left( \left( \frac{1}{n} \sum_{k=1}^n \tilde{s}(Y_k, Z_k) \right)^2 \right) = \frac{1}{n^2} \sum_k \mathbb{E}_{z_0} (\tilde{s}^2(Y_k, Z_k)) + \frac{2}{n^2} \sum_{k < k'} \mathbb{E}_{z_0} (\tilde{s}(Y_k, Z_k) \tilde{s}(Y_{k'}, Z_{k'})). \quad (31)$$

We notice that:

$$\begin{aligned} \mathbb{E}_{z_0} (\tilde{s}^2(Y_k, Z_k)) &= \mathbb{E}_{z_0} (s^2(Y_k, Z_k)) - (\mathbb{E}_{z_0} (s(Y_k, Z_k)))^2 \leq \mathbb{E}_{z_0} (s^2(Y_k, Z_k)) \\ &\leq \int s^2(y, z) \rho(dy, dz) + \mathbb{E}_{z_0} (s^2(Y_k, Z_k)) - \int s^2(y, z) \rho(dy, dz) \\ &\leq \int s^2(y, z) \rho(dy, dz) + R\mathbf{V}_\lambda(z_0) \gamma^k \end{aligned}$$

by Assumption A2. Therefore

$$\mathbb{E}_{z_0} \left( \frac{1}{n^2} \sum_{k=1}^n \tilde{s}^2(Y_k, Z_k) \right) \leq \frac{1}{n} \int s^2(y, z) \rho(dy, dz) + \frac{R\mathbf{V}_\lambda(z_0)}{n^2(1-\gamma)}. \quad (32)$$

Let us bound the last term of (31). We can remark that  $(Z_0, Y_1, Z_1, \dots, Y_k, Z_k, \dots)$  is an inhomogeneous Markov chain. Therefore, for any  $(k < k')$ ,  $\mathbb{E}(s(Y_{k'}, Z_{k'}) | Y_k, Z_k) = \mathbb{E}(s(Y_{k'}, Z_{k'}) | Z_k)$  and by Assumption A2,

$$\begin{aligned} |\mathbb{E}(\tilde{s}(Y_{k'}, Z_{k'}) | Y_k, Z_k)| &\leq \left| \mathbb{E}(s(Y_{k'}, Z_{k'}) | Y_k, Z_k) - \int s(y, z) \rho(dy, dz) \right| \\ &\quad + \left| -\mathbb{E}_{z_0}(s(Y_{k'}, Z_{k'})) + \int s(y, z) \rho(dy, dz) \right| \\ &\leq R\gamma^{k'-k} \mathbf{V}_\lambda(Z_k) + R\mathbf{V}_\lambda(z_0) \gamma^{k'}. \end{aligned} \quad (33)$$

Then

$$\begin{aligned}
E_k &:= \sum_{k'=k+1}^n |\mathbb{E}_{z_0} (\tilde{s}(Y_k, Z_k) \tilde{s}(Y_{k'}, Z_{k'}))| = \sum_{k'=k+1}^n |\mathbb{E}_{z_0} (\tilde{s}(Y_k, Z_k) \mathbb{E}(\tilde{s}(Y_{k'}, Z_{k'}) | Y_k, Z_k))| \\
&\leq \sum_{k'=k+1}^n \mathbb{E}_{z_0} \left( |\tilde{s}(Y_k, Z_k)| (R\gamma^{k'-k} \mathbf{V}_\lambda(Z_k) + R\mathbf{V}_\lambda(z_0)\gamma^{k'}) \right) \\
&\leq \frac{R}{1-\gamma} \mathbb{E}_{z_0} (|\tilde{s}(Y_k, Z_k)| (\mathbf{V}_\lambda(Z_k) + \gamma^k \mathbf{V}_\lambda(z_0))).
\end{aligned}$$

As  $|\tilde{s}(Y_k, Z_k)| \leq |s(Y_k, Z_k)| + \mathbb{E}_{z_0} (|s(Y_k, Z_k)|)$ ,

$$\begin{aligned}
E_k &\leq \frac{R}{1-\gamma} (\mathbb{E}_{z_0} (|s(Y_k, Z_k)| \mathbf{V}_\lambda(Z_k)) + \mathbb{E}_{z_0} (|s(Y_k, Z_k)|) \mathbb{E}_{z_0} (\mathbf{V}_\lambda(Z_k))) \\
&\quad + \frac{R\mathbf{V}_\lambda(z_0)}{1-\gamma} \gamma^k 2\mathbb{E}_{z_0} (|s(Y_k, Z_k)|).
\end{aligned}$$

By Assumption A2, for any function  $|\psi| \leq \mathbf{V}_\lambda$ ,

$$\mathbb{E}_{z_0} (\psi(Y_k, Z_k)) \leq \int_0^\infty \psi(y, z) \rho(dy, dz) + R\mathbf{V}_\lambda(z_0)\gamma^k.$$

Then

$$\begin{aligned}
\sum_k E_k &\leq \sum_k \frac{R}{1-\gamma} \left( \int |s(y, z)| \mathbf{V}_\lambda(z) \rho(dy, dz) + R\mathbf{V}_\lambda(z_0)\gamma^k \right) \\
&\quad + \sum_k \frac{R}{1-\gamma} \left( \int |s(y, z)| \rho(dy, dz) + R\mathbf{V}_\lambda(z_0)\gamma^k \right) \\
&\quad \times \left( \int \mathbf{V}_\lambda(z) \rho(dy, dz) + R\mathbf{V}_\lambda(z_0)\gamma^k \right) \\
&\quad + \sum_k \frac{R\mathbf{V}_\lambda(z_0)}{1-\gamma} 2\gamma^k \left( \int |s(y, z)| \rho(dy, dz) + R\mathbf{V}_\lambda(z_0)\gamma^k \right) \\
&\leq n \frac{R}{1-\gamma} \left( \int |s(y, z)| \mathbf{V}_\lambda(z) \rho(dy, dz) + \int |s(y, z)| \rho(dy, dz) \int \mathbf{V}_\lambda(z) \mu(dz) \right) \\
&\quad + \frac{R\mathbf{V}_\lambda(z_0)}{(1-\gamma)^2} \left( R + R \int \mathbf{V}_\lambda(z) \mu(dz) + (R+2) \int |s(y, z)| \rho(dy, dz) \right) \\
&\quad + \frac{R^2(R+2)\mathbf{V}_\lambda^2(z_0)}{(1-\gamma)(1-\gamma^2)}.
\end{aligned}$$

By (31) and (32), we get:

$$\mathbb{E}_{z_0} \left( \left( \frac{1}{n} \sum_{k=1}^n \tilde{s}(Y_k, Z_k) \right)^2 \right) \leq \frac{1}{n} \left( \int s^2(y, z) \rho(dy, dz) + \int |s(y, z)| G_\lambda(y, z) \rho(dy, dz) \right) + \frac{C}{n^2}$$

where  $C$  depends only on  $R$ ,  $\mathbf{V}_\lambda$  and  $\gamma$  and we recall that  $G_\lambda(z) = \frac{R}{1-\gamma} (\mathbf{V}_\lambda(z) + \int \mathbf{V}_\lambda(u) \mu(du))$ .

## A.2 Proof of Lemma 3

Let  $G$  be an event of  $\mathcal{O}_0^k \times \mathcal{O}_{t+k}^\infty$ . Then  $G$  is a disjoint reunion of events  $E^i \cap F^{i,j}$  where

$$E^i = \{(Y_1, Z_1) \in J_1^i, \dots, (Y_k, Z_k) \in J_k^i\},$$

$$F^{i,j} = \{(Y_{t+k}, Z_{t+k}) \in I_0^{i,j}, \dots, (Y_{t+k+\mathbf{n}}, Z_{t+k+\mathbf{n}}) \in I_{\mathbf{n}}^{i,j}\}$$

with  $J_j^i$  and  $I_l^{i,j}$  subsets of  $(\mathbb{R}^+)^2$  and  $1 \leq \mathbf{n} < \infty$ . Then

$$D_G := P_{\mathcal{O}_0^k, \mathcal{O}_{t+k}^\infty}(G) - P_{\mathcal{O}_0^k} \otimes P_{\mathcal{O}_{t+k}^\infty}(G) = \sum_{i,j} \mathbb{P}(E^i \cap F^{i,j}) - \mathbb{P}(E^i) \mathbb{P}(F^{i,j}).$$

As  $(Y_k, Z_k)_{k \in \mathbb{N}}$  is a Markov chain,

$$\begin{aligned} A_{i,j} &:= \mathbb{P}(E^i \cap F^{i,j}) - \mathbb{P}(E^i) \mathbb{P}(F^{i,j}) \\ &= \int_{J_1^i \times \dots \times J_k^i} \mathcal{P}(z_0, dy_1) Q(y_1, dz_1) \dots \mathcal{P}(z_{k-1}, dy_k) Q(y_k, dz_k) \\ &\quad \times \int_{I_0^{i,j}} (\mathbb{P}(Y_{t+k} \in dy'_0, Z_{t+k} \in dz'_0 | Z_k = z_k) - \mathbb{P}(Y_{t+k} \in dy'_0, Z_{t+k} \in dz'_0 | Z_0 = z_0)) \\ &\quad \times \int_{I_1^{i,j} \times \dots \times I_{\mathbf{n}}^{i,j}} \mathcal{P}(z'_0, dy'_1) Q(y'_1, dz'_1) \dots \mathcal{P}(z'_{\mathbf{n}-1}, dy'_{\mathbf{n}}) Q(y'_{\mathbf{n}}, dz'_{\mathbf{n}}). \end{aligned}$$

To simplify the notations, let us set

$$\mathcal{R}_k(x, dy_1, dz_1, \dots, dy_k, dz_k) := \mathcal{P}(x, dy_1) Q(y_1, dz) \dots \mathcal{P}(z_{k-1}, y_k) Q(y_k, z_k).$$

and  $\mathcal{R}^t(x, dy, dz) = \mathbb{P}(Y_t \in dy, Z_t \in dz | Z_0 = x)$ . Then

$$\begin{aligned} A_{i,j} &= \int_{J_1^i \times \dots \times J_k^i} \mathcal{R}_k(z_0, dy_1, dz_1, \dots, dy_k, dz_k) \\ &\quad \times \int_{I_0^{i,j}} (\mathcal{R}^t(z_k, dy'_0, dz'_0) - \mathcal{R}^{t+k}(z_0, dy'_0, dz'_0)) \mathbb{1}_{y'_0, z'_0 \in I_0^{i,j}} \\ &\quad \times \int_{I_1^{i,j} \times \dots \times I_{\mathbf{n}}^{i,j}} \mathcal{R}_{\mathbf{n}}(z'_0, dy'_1, dz'_1, \dots, dy'_{\mathbf{n}}, dz'_{\mathbf{n}}). \end{aligned}$$

We regroup the  $F^{i,j}$ :

$$\begin{aligned} A_i &:= \sum_j A_{i,j} = \sum_j \mathbb{P}(E^i \cap F^{i,j}) - \mathbb{P}(E^i) \mathbb{P}(F^{i,j}) \\ &= \int_{J_1^i \times \dots \times J_k^i} \mathcal{R}_k(z_0, dy_1, dz_1, \dots, dy_k, dz_k) (\mathbb{E}_{z_k}(\psi(Y_t, Z_t)) - \mathbb{E}_{z_0}(\psi(Y_{t+k}, Z_{t+k}))) \end{aligned}$$

where  $\psi(y, z) := \sum_j \mathbb{1}_{(y,z) \in I_0^{i,j}} \int_{I_1^{i,j} \times \dots \times I_n^{i,j}} \mathcal{R}_n(z, dy'_1, dz'_1, \dots, dy'_n, dz'_n)$ . We can remark that  $\psi(y, z) = \sum_j \mathbb{1}_{y,z \in I_0^{i,j}} \mathbb{P}_z((Y_1, Z_1) \in I_1^{i,j}, \dots, (Y_n, Z_n) \in I_n^{i,j})$  and by the law of total probability,  $\psi(y, z) \leq 1$ . We can apply Assumption A2 to the function  $\psi$ :

$$|\mathbb{E}_{z_k}(\psi(Y_t, Z_t)) - \mathbb{E}_{z_0}(\psi(Y_{t+k}, Z_{t+k}))| \leq R\gamma^t \mathbf{V}_\lambda(z_k) + R\gamma^{t+k} \mathbf{V}_\lambda(z_0).$$

Then

$$\begin{aligned} |D_G| &= \left| \sum_i A_i \right| \leq R\gamma^t \sum_i \int_{J_1^i \times \dots \times J_k^i} \mathcal{R}_k(z_0, dy_1, dz_1, \dots, dy_k, dz_k) (\mathbf{V}_\lambda(z_k) + \mathbf{V}_\lambda(z_0)) \\ &\leq R\gamma^t \sum_i \mathbb{E}_{z_0} \left( (\mathbf{V}_\lambda(Z_k) + \mathbf{V}_\lambda(z_0)) \mathbb{1}_{(Y_1, Z_1) \in J_1^i, \dots, (Y_k, Z_k) \in J_k^i} \right) \\ &\leq R\gamma^t (\mathbb{E}_{z_0}(\mathbf{V}_\lambda(Z_k)) + \mathbf{V}_\lambda(z_0)). \end{aligned}$$

By Lemma 1,

$$\mathbb{E}_{z_0}(\mathbf{V}_\lambda(Z_k)) \leq \int \mathbf{V}_\lambda(z) \mu(dz) + R\gamma^k \mathbf{V}_\lambda(z_0).$$

Therefore

$$\beta_{Y,Z}(t) = \sup_k \sup_{G \in \mathcal{O}_0^k \times \mathcal{O}_{t+k}^\infty} |D_G| \leq R\gamma^t \left( \int \mathbf{V}_\lambda(z) \mu(dz) + (1 + R)\mathbf{V}_\lambda(z_0) \right).$$

As  $\gamma < 1$ ,

$$\beta_{Y,Z}(t) \leq ce^{-\beta t}$$

with  $\beta = -\ln(\gamma)$ ,  $c = R \left( \int \mathbf{V}_\lambda(z) \mu(dz) + (1 + R)\mathbf{V}_\lambda(z_0) \right)$ .

## A.3 Proof of Lemma 4

### A.3.1 Assumption A2 is satisfied

Assumption (S)d implies Assumption A2a. To prove Assumption A2b and c, in analogy with [24], we apply the following result, which is Theorem 1.1 of Baxendale [7] written for a Markov chain on  $\mathbb{R}^2$  and a finite measure instead of a probability.

**Result** (Sufficient conditions for ergodicity). *Let us consider  $(Y_k, Z_k)_{k \geq 1}$  an homogeneous Markov chain on  $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$  with transition probability  $\tilde{R}$ . Under the following three conditions,*

**Minorization condition** *There exist a set  $\mathcal{C} \subset \mathbb{R}^2$  and a finite measure  $\mathbf{s}$  such that  $\forall (y_1, z_1) \in \mathcal{C}, \forall A \in \mathcal{B}(\mathbb{R}^2)$ ,*

$$\int_A \tilde{\mathcal{R}}(y_1, z_1, dy, dz) \geq \mathbf{s}(A).$$

**Strong aperiodicity condition**  $\mathbf{s}(\mathcal{C}) > 0$ .

**Drift condition** *There exists a function  $\mathbf{V} : \mathbb{R}^2 \rightarrow [1, \infty[$  and two constants  $c < 1, K > 0$  such that*

$$\mathbb{E}_{y_1, z_1}(\mathbf{V}(Y_2, Z_2)) \leq c\mathbf{V}(y_1, z_1)\mathbb{1}_{\{(y_1, z_1) \in \mathcal{C}^c\}} + K\mathbb{1}_{\{(y_1, z_1) \in \mathcal{C}\}}.$$

*Then the process  $\{(Y_k, Z_k) : k \geq 0\}$  is recurrent positive and strongly ergodic, and has a unique stationary probability measure  $\rho$ .*

*Moreover, there exists  $\gamma$  and  $R$  depending only on  $\mathbf{s}, c$  and  $K$  such that, for any function  $\psi \leq \mathbf{V}$ ,*

$$|\mathbb{E}(\psi(Y_k, Z_k) | Z_0 = z_0) - \mathbb{E}_\rho(\psi(Y_1, Z_1))| \leq R\mathbf{V}(z_0)\gamma^k.$$

Then Assumptions A2b and c are satisfied.

Let us check that its three conditions (minorization, strong aperiodicity and drift) are satisfied. We need to control the transition density. As  $(Z_0, Y_1, Z_1, \dots)$  is an (inhomogeneous) Markov chain, let us note

$$\tilde{\mathcal{R}}(y_1, z_1, dy, dz) = \mathbb{P}(Y_2 \in dy, Z_2 \in dz | Z_1 = z_1, Y_1 = y_1) = \mathcal{P}(z_1, dy)Q(y, dz).$$

Let us set

$$i'_1 = \max \left( i_1, \left( \frac{1}{\mathbf{a}(1 - \kappa^{b+1})} \ln \left( \frac{2\kappa^{b+1}}{1 - \kappa^{b+1}} \right) \right)^{1/(b+1)} \mathbb{1}_{\{\kappa^{b+1} \geq 1/3\}} \right).$$

**Minorization condition** Let us set  $\mathcal{C} = \mathbb{R}^+ \times [0, i'_1]$ . For any  $(y_1, z_1) \in \mathcal{C}$ , any  $A \subseteq (\mathbb{R}^+)^2$ , by Assumption (S)b, we have that

$$\begin{aligned} \int_A \tilde{\mathcal{R}}(y_1, z_1, dy, dz) &= \int_A \lambda(y)(\phi_{z_1}^{-1})'(y) \exp \left( - \int_{z_1}^y \lambda(u)(\phi_{z_1}^{-1})'(u) du \right) \mathbb{1}_{\{z_1 \leq y\}} Q(y, dz) dy \\ &\geq \int_A \lambda(y)\mathbf{m}(y) \exp \left( - \int_0^y \lambda(u)\mathbf{M}(u) du \right) \mathbb{1}_{\{i'_1 \leq y\}} Q(y, dz) dy \end{aligned}$$

By Assumption (S)d and c, for any  $\lambda \in \mathcal{E}(\mathbf{s}, b, \alpha)$ ,

$$\begin{aligned} \int_A \tilde{\mathcal{R}}(y_1, z_1, dy, dz) &\geq \int_{A \cap [i'_1, i'_2] \times [0, i'_1]} \lambda(y)\mathbf{m}(y) \exp \left( - \int_0^y \lambda(u)\mathbf{M}(u) du \right) Q(y, dz) dy \\ &\geq \int_{A \cap [i'_1, i'_2] \times [0, i'_1]} \mathbf{a} \frac{y^b}{b+1} \exp \left( -1 - \mathbf{L} \int_{i_1}^y \mathbf{M}(u) du \right) Q(y, dz) dy \\ &=: \mathbf{s}_\mathbf{s}(A) \end{aligned}$$

and  $\mathbf{s}_\mathbf{s}(A)$  is a finite measure.

### Strong aperiodicity condition

$$\begin{aligned} \mathbf{s}_s(\mathcal{C}) &= \int_{i_1'}^{i_2'} \mathbf{a} \frac{y^b}{b+1} \exp\left(-1 - \mathbf{L} \int_{i_1}^y \mathbf{M}(u) du\right) \int_0^{i_1'} Q(y, dz) dy \\ &\geq \int_{i_1'}^{(\kappa^{-1}i_1') \wedge i_2'} \mathbf{a} \frac{y^b}{b+1} \exp\left(-1 - \mathbf{L} \int_{i_1}^{i_2'} \mathbf{M}(u) du\right) \int_0^{i_1'} Q(y, dz) dy. \end{aligned}$$

For any  $y \leq \kappa^{-1}i_1'$ , by Assumption (S) a,  $\int_0^{i_1'} Q(y, dz) = \mathbb{P}(Z_1 \leq i_1' | Y_1 = y) = 1$ . Therefore

$$\mathbf{s}_s(\mathcal{C}) \geq \int_{i_1'}^{(\kappa^{-1}i_1') \wedge i_2'} \mathbf{a} \frac{y^b}{b+1} \exp\left(-1 - \mathbf{L} \int_{i_1}^{i_2'} \mathbf{M}(u) du\right) dy.$$

Then  $\mathbf{s}_s(\mathcal{C}) > 0$ .

**Drift condition** For any  $(y_1, z_1)$ , as  $(Y_1, Z_1, Y_2, Z_2)$  is an (inhomogeneous) Markov chain,

$$\begin{aligned} \mathbb{E}(\mathbb{V}_b(Z_2) | Y_1 = y_1, Z_1 = z_1) &= \mathbb{E}_{z_1}(\mathbb{V}_b(Z_2)) \\ &= \int_0^\infty \mathbb{P}(\mathbb{V}_b(Z_2) > z | Z_1 = z_1) dz \end{aligned}$$

where  $\mathbb{V}_b(z) = \exp(\mathbf{a}z^{b+1})$ . By Assumption (S)a, as  $\mathbb{V}_b$  is an increasing function,  $\mathbb{V}_b(Z_2) > z \iff Z_2 \geq \mathbb{V}_b^{-1}(z) \implies Y_2 \geq \kappa^{-1}\mathbb{V}_b^{-1}(z)$ . Then by (2) and Assumption (S)b,

$$\begin{aligned} \mathbb{E}_{z_1}(\mathbb{V}_b(Z_2)) &\leq \int_0^\infty \mathbb{P}_{z_1}(Y_2 \geq \kappa^{-1}\mathbb{V}_b^{-1}(z)) dz \\ &= \int_0^\infty \exp\left(-\int_{z_1}^{\kappa^{-1}\mathbb{V}_b^{-1}(z)} (\phi_{z_1}^{-1})'(u) \lambda(u) du\right) \mathbb{1}_{\{\kappa^{-1}\mathbb{V}_b^{-1}(z) \geq z_1\}} dz \\ &\leq \int_0^\infty \exp\left(-\int_{z_1}^{\kappa^{-1}\mathbb{V}_b^{-1}(z)} \mathbf{m}(u) \lambda(u) du\right) \mathbb{1}_{\{\kappa^{-1}\mathbb{V}_b^{-1}(z) \geq z_1\}} dz =: \tilde{I}(z_1) \quad (34) \end{aligned}$$

Let us make the change of variable  $y = \kappa^{-1}\mathbb{V}_b^{-1}(z)$ , then  $dz = \kappa \mathbb{V}_b'(\kappa y) dy$  and

$$\tilde{I}(z_1) = \kappa \int_0^\infty \mathbb{V}_b'(\kappa y) \exp\left(-\int_{z_1}^y \mathbf{m}(u) \lambda(u) du\right) \mathbb{1}_{\{y \geq z_1\}} dy.$$

Let us first bound this quantity for  $z_1 \geq i_1$ . By Assumption (S)c, for any  $z_1 \geq i_1$ ,

$$\int_{z_1}^y \mathbf{m}(u) \lambda(u) du \geq \mathbf{a}(y^{b+1} - z_1^{b+1}).$$

As  $\mathbb{V}_b(y) = \exp(\mathbf{a}y^{b+1})$ ,  $\mathbb{V}'_b(\kappa y) = \mathbf{a}(b+1)\kappa^b y^b \exp(\mathbf{a}\kappa^{b+1}y^{b+1})$  and, for any  $z_1 \geq i_1$ ,

$$\begin{aligned} \tilde{I}(z_1) &\leq \int_{z_1}^{\infty} \mathbf{a}(b+1)\kappa^{b+1}y^b \exp(-\mathbf{a}y^{b+1}(1-\kappa^{b+1}) + \mathbf{a}z_1^{b+1}) dy \\ &\leq \mathbb{V}_b(z_1) \frac{\kappa^{b+1}}{1-\kappa^{b+1}} [-\exp(-\mathbf{a}(1-\kappa^{b+1})y^{b+1})]_{z_1}^{\infty} \\ &\leq \frac{\kappa^{b+1}}{1-\kappa^{b+1}} \mathbb{V}_b(z_1) \mathbb{V}_b(z_1)^{\kappa^{b+1}-1}. \end{aligned} \quad (35)$$

We have that

$$(i'_1)^{b+1} \geq \frac{1}{\mathbf{a}(1-\kappa^{b+1})} \ln\left(\frac{2\kappa^{b+1}}{1-\kappa^{b+1}}\right) \quad \text{then} \quad (\mathbb{V}_b(i'_1))^{1-\kappa^{b+1}} \geq \frac{2\kappa^{b+1}}{1-\kappa^{b+1}}.$$

Therefore, for any  $z_1 \geq i'_1$ , as  $\mathbb{V}_b$  is an increasing function,

$$\mathbb{V}_b(z_1)^{\kappa^{b+1}-1} \leq \mathbb{V}_b(i'_1)^{\kappa^{b+1}-1} \leq \frac{1-\kappa^{b+1}}{2\kappa^{b+1}}.$$

Then

$$\tilde{I}(z_1) \leq \frac{\mathbb{V}_b(z_1)}{2}. \quad (36)$$

Moreover, by (34),

$$\begin{aligned} \sup_{z_1 \leq i_1} \tilde{I}(z_1) &\leq \int_0^{\mathbb{V}_b(\kappa i_1)} \exp\left(-\int_0^{\kappa^{-1}\mathbb{V}_b^{-1}(z)} \mathbf{m}(u)\lambda(u)du\right) dz \\ &\quad + \int_{\mathbb{V}_b(\kappa i_1)}^{\infty} \exp\left(-\int_0^{i_1} \mathbf{m}(u)\lambda(u)du - \int_{i_1}^{\kappa^{-1}\mathbb{V}_b^{-1}(z)} \mathbf{m}(u)\lambda(u)du\right) dz \\ &\leq \mathbb{V}_b(\kappa i_1) + \int_0^{\infty} \exp\left(-\int_{i_1}^{\kappa^{-1}\mathbb{V}_b^{-1}(z)} \mathbf{m}(u)\lambda(u)du\right) dz \mathbb{1}_{\{\kappa^{-1}\mathbb{V}_b^{-1}(z) \geq i_1\}} \\ &\leq \mathbb{V}_b(\kappa i_1) + \tilde{I}(i_1) \end{aligned}$$

and by (35),

$$\sup_{z_1 \leq i'_1} \tilde{I}(z_1) \leq \mathbb{V}_b(i_1) + \frac{\kappa^{b+1}}{1-\kappa^{b+1}} \mathbb{V}_b(i'_1) \leq C\mathbb{V}_b(i'_1) < \infty$$

Therefore the three conditions (minorization, strong aperiodicity and drift) are satisfied, which gives Assumption A2.

### A.3.2 Assumption A3 is satisfied

It remains to prove that Assumption A3 is satisfied. We recall that

$$\Phi_0 = \inf_{x \in [0, i_2], y \in \mathcal{I}} (\phi_x^{-1})'(y) \quad \text{and} \quad \Phi_1 = \sup_{x \in [0, i_2], y \in \mathcal{I}} (\phi_x^{-1})'(y).$$

By equation (5), for any  $y \in \mathcal{I}$ ,

$$\mathbf{D}(y) \geq \Phi_0 \inf_{y' \in \mathcal{I}} \int_0^{y'} \mathbb{P}(Y_1 > y' | Z_0 = x) \mu(dx).$$

By equation (2), Assumption (S)b and d, for any  $y \in \mathcal{I}$ ,

$$\mathbb{P}(Z_1 > y | Z_0 = x) \geq \exp\left(-1 - \int_{i_1}^{i_2} \mathbf{L}(\phi_x^{-1})'(u) du\right) \geq e^{-1 - \mathbf{L}\Phi_1 i_2}.$$

Then

$$\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \inf_{y \in \mathcal{I}} \mathbf{D}(y) \geq \Phi_0 e^{-1 - \mathbf{L}\Phi_1 i_2} \mu([0, i_1]).$$

It remains to bound  $\mu([0, i_1])$  away from 0.

As  $\mu$  is the stationary density of  $(Z_k)$ ,  $\mu(\cdot | z, \infty) = \mathbb{P}_\mu(Z_1 > z)$ . Therefore, by Markov inequality, as  $\mathbb{V}_b$  is an increasing function,

$$\mu(\cdot | z, \infty) = \mathbb{P}_\mu(\mathbb{V}_b(Z_1) > \mathbb{V}_b(z)) \leq \mathbb{V}_b^{-1}(z) \mathbb{E}_\mu(\mathbb{V}_b(Z_1)).$$

By Lemma 4a,

$$\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{E}_\mu(\mathbb{V}_b(Z_1)) \leq \mathbb{V}_b(z_0)(1 + \gamma R) < \infty.$$

As  $\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mathbb{E}_\mu(\mathbb{V}_b(Z_1)) < \infty$ , and  $\mathbb{V}_b$  is an increasing function, there exists  $y_0 > 0$ ,  $\sup_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mu(\cdot | y_0, \infty) < 1$  and consequently,  $\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mu([0, y_0]) > 0$ . Let us consider the sequence

$$(z_0 := i_1, z_1 := z_0/\sqrt{\kappa}, \dots, z_j := z_{j-1}/\sqrt{\kappa} = \kappa^{-j/2}i_1, \dots, z_{k_n} := \kappa^{-k_n/2}i_1)$$

where  $z_{k_n-1} < y_0 \leq z_{k_n}$ . We can remark that

$$\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mu([0, z_{k_n}]) > 0. \tag{37}$$

As  $\mu$  is the stationary density, for any  $z > 0$ ,

$$\mu([0, z]) = \int_0^\infty \mathbb{P}(Z_1 \leq z | Z_0 = x) \mu(dx).$$

As  $\mathbb{P}(Z_1 \leq \kappa Y_1) = 1$ ,  $\mathbb{P}(Z_1 \leq z | Z_0 = x) \geq \mathbb{P}(Y_1 \leq \kappa^{-1}z | Z_0 = x)$  and by (2),

$$\begin{aligned} \mu([0, z_j]) &\geq \int_0^\infty \left( 1 - \exp \left( - \int_x^{\kappa^{-1}z_j} \lambda(u)(\phi_x^{-1})'(u) du \right) \mathbb{1}_{\{x \leq \kappa^{-1}z_j\}} \right) \mu(dx) \\ &\geq \int_{i_1}^{z_{j+1}} \left( 1 - \exp \left( - \int_x^{\kappa^{-1}z_j} \lambda(u)(\phi_x^{-1})'(u) du \right) \right) \mu(dx) \end{aligned}$$

as  $\kappa^{-1}z_j \geq z_{j+1}$ . By Assumption (S)b and c,  $\lambda$  and  $(\phi_x^{-1})'$  are bounded by below and there exists a constant  $\eta$  such that

$$\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \inf_{u \in [i_1, z_{k_n}]} \lambda(u)(\phi_x^{-1})'(u) \geq \eta.$$

Therefore, as  $\kappa^{-1}z_j = \kappa^{-1/2}z_{j+1}$ ,

$$\begin{aligned} \mu([0, z_j]) &\geq \int_{i_1}^{z_{j+1}} (1 - \exp(-\eta(\kappa^{-1}z_j - z_{j+1}))) \mu(dx) \\ &\geq (1 - \exp(-\eta(\kappa^{-1}z_j(1 - \sqrt{\kappa})))) \mu([i_1, z_{j+1}]). \end{aligned}$$

Let us set  $c_j = (1 - \exp(-\eta(\kappa^{-1}z_j(1 - \sqrt{\kappa}))))$ . We can note that

$$\mu([0, z_j]) \geq c_j (\mu([0, z_{j+1}]) - \mu([0, i_1]))$$

and in particular,  $\mu([0, i_1])(1 + c_0) \geq c_0 \mu([0, z_1])$ . By recurrence, we obtain:

$$\left( 1 + \sum_{j=0}^{k_n-1} \prod_{i=0}^j c_i \right) \mu([0, i_1]) \geq \left( \prod_{j=0}^{k_n-1} c_j \right) \mu([0, z_{k_n}]).$$

Then by (37)

$$\inf_{\lambda \in \mathcal{E}(\mathfrak{s}, b, \alpha)} \mu([0, i_1]) > 0$$

which concludes the proof.

## A.4 Besov and Hölder spaces

**Definition 13** (Modulus of continuity). *The modulus of continuity is defined by*

$$\omega(f, t) = \sup_{|x-y| \leq t} |f(x) - f(y)|.$$

If  $f$  is Lipschitz, the modulus of continuity is proportional to  $t$ . If  $\omega(f, t) = o(t)$ , then  $f$  is constant: the modulus of continuity can not measure higher smoothness.

**Definition 14** (Modulus of smoothness). *If  $f$  is a function on  $\mathcal{A}$ , we define its modulus of smoothness by*

$$\omega_{\mathbf{r}}(f, t)_p = \sup_{0 < h \leq t} \|\Delta_h^{\mathbf{r}}(f, \cdot)\|_{L^p(\mathcal{A})} \quad \text{where} \quad \Delta_h^{\mathbf{r}}(f, x) = \sum_{k=0}^{\mathbf{r}} (-1)^k C_{\mathbf{r}}^k f(x + kh).$$

We can remark that if  $f$  is  $C^{\mathbf{r}}$ , then

$$\lim_{t \rightarrow 0} t^{-\mathbf{r}} \omega_{\mathbf{r}}(f, t)_p = \|f^{(\mathbf{r})}\|_{L^p(\mathcal{A})} \quad \text{and} \quad \lim_{t \rightarrow 0} t^{-\alpha} \omega_{\mathbf{r}+1}(f, t)_p = \lim_{t \rightarrow 0} t^{-\alpha+\mathbf{r}} \omega(f^{(\mathbf{r})}, t)_p.$$

In particular, if  $f \in C^{\mathbf{r}}(\mathcal{A})$  with  $\mathcal{A}$  compact and if  $f^{(\mathbf{r}+1)}$  is Lipschitz, then  $\omega_{\mathbf{r}+1}(f, t)_p = O(t^{\mathbf{r}+1})$ . If  $f^{(\mathbf{r})}$  is  $(\alpha - \mathbf{r})$ -Hölder-continuous, that is if  $\forall x, y \in \mathcal{A}$ ,  $|f^{(\mathbf{r})}(x) - f^{(\mathbf{r})}(y)| \leq C|x - y|^{\alpha - \mathbf{r}}$ , then

$$\omega_{\mathbf{r}+1}(f, t)_p = O(t^{\alpha}).$$

If  $f^{(\mathbf{r})}$  is piecewise-continuous and  $(\alpha - \mathbf{r})$ -Hölder on the points of continuity, then

$$\omega_{\mathbf{r}+1}(f, t)_p = O(t^{1/p} + t^{\alpha}).$$

The modulus of continuity and the modulus of smoothness are sub-linear:

$$\omega_{\mathbf{r}}(f + g, t)_p \leq \omega_{\mathbf{r}}(f, t)_p + \omega_{\mathbf{r}}(g, t)_p \quad \text{and} \quad \omega_{\mathbf{r}}(af, t)_p = a\omega_{\mathbf{r}}(f, t)_p.$$

**Definition 15** (Besov space). *The Besov space  $B_{2,\infty}^{\alpha}(\mathcal{A})$  is the set of functions:*

$$B_{2,\infty}^{\alpha}(\mathcal{A}) = \{f \in L^2(\mathcal{A}), \sup_{t>0} t^{-\alpha} \omega_{\mathbf{r}+1}(f, t)_2 < \infty\}$$

where  $\mathbf{r} = \lfloor \alpha \rfloor$ . The norm is defined by:  $\|f\|_{B_{2,\infty}^{\alpha}} := \sup_{t>0} t^{-\alpha} \omega_{\mathbf{r}+1}(f, t)_2 + \|f\|_{L^2(\mathcal{A})}$ . We denote  $\mathcal{B}_{2,\infty}^{\alpha}(\mathcal{A}, M_1) = \{f \in B_{2,\infty}^{\alpha}(\mathcal{A}), \|f\|_{B_{2,\infty}^{\alpha}(\mathcal{A})} \leq M_1\}$ .

See DeVore and Lorentz [15] and Meyer [27] for more details. We use the Besov space to control the risk of the estimator of the stationary density  $\nu$ .

**Definition 16** (Hölder space). *The Hölder space is the set of functions:*

$$H^{\alpha}(\mathcal{A}) = \{f \in \mathcal{C}^{\mathbf{r}}(\mathcal{A}), t^{\mathbf{r}-\alpha} \omega(f^{(\mathbf{r})}, t)_{\infty} < \infty \quad \forall t > 0\}$$

where  $\mathbf{r} = \lfloor \alpha \rfloor$ . We note  $|f|_{H^{\alpha}(\mathcal{A})} := \sup_{t>0} t^{\mathbf{r}-\alpha} \omega(f^{(\mathbf{r})}, t)_{\infty}$  and define the norm of the Hölder space  $\|f\|_{H^{\alpha}(\mathcal{A})} = |f|_{H^{\alpha}(\mathcal{A})} + \|f\|_{L^{\infty}(\mathcal{A})}$  and  $H^{\alpha}(\mathcal{A}, M_1) = \{f \in H^{\alpha}(\mathcal{A}), \|f\|_{H^{\alpha}(\mathcal{A})} \leq M_1\}$ .

As noted before,  $t^{\mathbf{r}-\alpha} \omega(f^{(\mathbf{r})}, t)_{\infty} = t^{-\alpha} \omega_{\mathbf{r}}(f, t)_{\infty}$ : the Hölder space  $H^{\alpha}(\mathcal{A})$  is included in  $B_{\infty,\infty}^{\alpha}(\mathcal{A})$  which itself is included in  $B_{2,\infty}^{\alpha}(\mathcal{A})$ . We can remark that if a function is  $C^{\mathbf{r}}$  and piecewise  $C^{\mathbf{r}+1}$ , it belongs to  $B_{2,\infty}^{\mathbf{r}+1/2}$  but only to  $H^{\mathbf{r}}$ .

## A.5 Proof of Lemma 5

As  $\nu$  is the stationary distribution of  $(Y_k)$ , by (4) and (3), we have, for all  $y \in \mathcal{I}$ :

$$\begin{aligned}\nu(y) &= \int_{\mathbb{R}^+ \times \mathbb{R}^+} \nu(x) Q(x, dz) \mathcal{P}(z, y) dx \\ &= \int_{\mathbb{R}^+} \nu(x) \int_0^y Q(x, dz) \lambda(y) (\phi_z^{-1})'(y) e^{-\int_z^y \lambda(u) (\phi_z^{-1})'(u) du} dx \\ &= \int_{\mathbb{R}^+} \nu(x) \int_0^y Q(x, dz) \Lambda(z, y) dx\end{aligned}$$

with

$$\Lambda(z, y) = \lambda(y) (\phi_z^{-1})'(y) e^{-\int_z^y \lambda(u) (\phi_z^{-1})'(u) du}.$$

As the Hölder spaces are stables by multiplication, composition and integration,  $\Lambda$  has the same regularity than  $\lambda$  and  $(\phi_x^{-1})'$ . We have that

$$\begin{aligned}\nu(y+h) - \nu(y) &= \int_{\mathbb{R}^+} \int_0^{y+h} \nu(x) Q(x, dz) (\Lambda(z, y+h) - \Lambda(z, y)) dx \\ &\quad + \int_{\mathbb{R}^+} \int_y^{y+h} \nu(x) Q(x, dz) \Lambda(z, y) dx.\end{aligned}$$

Let us set  $\mathcal{Q}_g(y) = \int_{\mathbb{R}^+} \int_0^y g(x) Q(x, dz) dx$ . If  $\mathcal{Q}_\nu$  is differentiable, we get:

$$\nu'(y) = \int_{\mathbb{R}^+} \nu(x) \int_0^y Q(x, dz) \frac{\partial \Lambda}{\partial y}(z, y) dx + \Lambda(y, y) \mathcal{Q}'_\nu(y)$$

and if  $\mathcal{Q}_\nu$  belongs to  $C^{\mathbf{r}}$ , there exist  $(c_{k_1, k_2})_{k_1 + k_2 \leq \mathbf{r} - 1} \in \mathbb{R}$  such that :

$$\nu^{(\mathbf{r})}(y) = \int_{\mathbb{R}^+} \nu(x) \int_0^y Q(x, dz) \frac{\partial^k \Lambda}{\partial y^k}(z, y) dx + \sum_{k_1 + k_2 \leq \mathbf{r} - 1} c_{k_1, k_2} \frac{\partial^{k_1 + k_2} \Lambda}{\partial y^{k_1} \partial z^{k_2}}(y, y) \mathcal{Q}_\nu^{(\mathbf{r} - k_1 - k_2)}(y).$$

It remains to study the regularity of the function  $\mathcal{Q}_\nu$ .

We consider some particular transition measures  $Q$  in order to understand how the regularity of  $\mathcal{Q}_\nu$  (and  $\nu$ ) depends on the form and the regularity of  $Q$ .

**Continuous transition measure** There exists a function  $Q_1$  such that  $Q(x, dy) = Q_1(x, y) dy$ , and we can write

$$\mathcal{Q}_\nu(y) = \int_{\mathbb{R}^+} \int_0^y \nu(x) Q_1(x, z) dz dx \quad \text{and} \quad \mathcal{Q}'_\nu(y) = \int_{\mathbb{R}^+} \nu(x) Q_1(x, y) dx.$$

Moreover, as  $Q_1(x, y) = 0$  if  $x < y$ , with  $\mathcal{I} = [i_1, i_2]$ , we get

$$\|\mathcal{Q}_\nu\|_{L^\infty(\mathcal{I})} \leq i_2 \|Q_1\|_{L^\infty([i_1, \infty[ \times \mathcal{I})} \int_{\mathbb{R}^+} \nu(x) dx = i_2 \|Q_1\|_{L^\infty([i_1, \infty[ \times \mathcal{I})}.$$

Furthermore, by definition of the Hölder semi-norm, for  $\mathbf{r} = \lfloor \alpha \rfloor$

$$\begin{aligned}
\|\mathcal{Q}_\nu\|_{H^\alpha(\mathcal{I})} &= \sup_{t>0} t^{\mathbf{r}-\alpha-1} \sup_{y,y'\in\mathcal{I},|y-y'|\leq t} \left| \mathcal{Q}_\nu^{(\mathbf{r})}(y) - \mathcal{Q}_\nu^{(\mathbf{r})}(y') \right| \\
&\leq \sup_{t>0} t^{\mathbf{r}-\alpha-1} \int_{\mathbb{R}^+} \nu(x) \sup_{x\geq y} \sup_{y,y'\in\mathcal{I},|y-y'|\leq t} \left| \frac{\partial^{\mathbf{r}-1} \mathcal{Q}_1}{\partial y^{\mathbf{r}-1}}(x, y') - \frac{\partial^{\mathbf{r}-1} \mathcal{Q}_1}{\partial y^{\mathbf{r}-1}}(x, y) \right| dx \\
&\leq \sup_{t>0} t^{\mathbf{r}-\alpha-1} \sup_{z\geq i_1} \sup_{y,y'\in\mathcal{I},|y-y'|\leq t} \left| \frac{\partial^{\mathbf{r}-1} \mathcal{Q}_1}{\partial y^{\mathbf{r}-1}}(z, y) - \frac{\partial^{\mathbf{r}-1} \mathcal{Q}_1}{\partial y^{\mathbf{r}-1}}(z, y') \right| \int_{\mathbb{R}^+} \nu(x) dx \\
&= \sup_{z\geq i_1} \|\mathcal{Q}_1(z, \cdot)\|_{H^{\alpha-1}(\mathcal{I})}.
\end{aligned}$$

Then  $\|\mathcal{Q}_\nu\|_{H^\alpha(\mathcal{I})} \leq i_2 \|\mathcal{Q}_1\|_{H^{\alpha-1}([i_1, \infty[\times\mathcal{I})}$ .

**Deterministic transition measure** Let us assume that  $Q$  can be written  $Q(x, dy) = \delta_{f(x)}(dy)$  with  $f$  a bijection. As  $\mathbb{P}(Z \leq \kappa Y) = 1$ ,  $f(0) = 0$ . Then we have that

$$\nu(y) = \int_0^{f^{-1}(y)} \nu(x) \Lambda(f(x), y) dx \quad \text{and} \quad \mathcal{Q}_\nu(y) = \int_0^{f^{-1}(y)} \nu(x) dx.$$

If  $f^{-1}$  is differentiable:

$$\mathcal{Q}'_\nu(y) = (f^{-1})'(y) \nu(f^{-1}(y)).$$

So we get:

$$\nu'(y) = \int_0^{f^{-1}(y)} \nu(x) \frac{\partial \Lambda}{\partial y}(f(x), y) dx + \Lambda(y, y) (f^{-1})'(y) \nu(f^{-1}(y)).$$

The regularity of  $\nu'$  on  $\mathcal{I}$  depends on the regularity of  $\nu$  on  $f^{-1}(\mathcal{I})$  and of  $\Lambda$  and  $f^{-1}$  on  $\mathcal{I}$ . By recurrence, there exists a function  $\psi_2$  such that

$$\|\mathcal{Q}_\nu\|_{H^\alpha(\mathcal{I})} \leq \psi_2 \left( \|\lambda\|_{H^\alpha(\mathcal{J})}, \|(\phi^{-1})'\|_{H^\alpha([0, j_2] \times \mathcal{J})}, \|f^{-1}\|_{H^\alpha(\mathcal{J})} \right)$$

where

$$\mathcal{J}_0 = \mathcal{I} \quad , \quad \mathcal{J}_{k+1} = \text{Conv} \left( \mathcal{I} \cup \bigcup_{i=1}^j f_i^{-1}(\mathcal{J}_k) \right) \quad \text{and} \quad \mathcal{J} = \mathcal{J}_{\lfloor \alpha \rfloor} \cup [i_1, i_2].$$

If  $f$  is not a bijection (and  $f(x) \neq 0$ ), then  $\nu$  can be less regular than  $\lambda$ . Let us consider  $f(x) = \lfloor x/2 \rfloor$ . Then

$$\nu(y) = \int_{\mathbb{R}^+} \nu(x) \sum_{k=0}^{\lfloor y \rfloor} \mathbb{1}_{k=\lfloor x/2 \rfloor} \Lambda(k, y) dx = \sum_{k=0}^{\infty} \nu([2k, 2k+2]) \Lambda(k, y) \mathbb{1}_{k \leq y}$$

Then  $\nu$  is a piecewise constant function and is not differentiable. We can remark that

$$\mathcal{Q}_\nu(y) = \sum_{k=0}^{\infty} \nu([2k, 2k+2]) \mathbb{1}_{k \leq y}$$

is not differentiable.

If  $Q(x, dy) = \delta_0(y)$  (which implies that the vectors  $(Z_k, Y_k)$  are independent), then  $\nu(y) = \int_{\mathbb{R}^+} \nu(x) \Lambda(0, y) dx = \Lambda(0, y)$  has the same regularity as  $\Lambda$ . We can remark that  $\mathcal{Q}_\nu(y) = \int_{\mathbb{R}^+} \nu(x) = 1$  is  $C^\infty$ .

**General case** Under Assumption (S),

$$Q(x, dy) = Q_1(x, y) dy + p_0(x) \delta_0(dy) + \sum_{i=1}^{j_Q} p_i(x) \delta_{f_i(x)}(dy)$$

with  $(f_i)$  invertible, therefore

$$\mathcal{Q}_\nu(y) = \int_{\mathbb{R}^+} \int_0^y \nu(x) Q_1(x, z) dz + \int_{\mathbb{R}^+} \nu(x) p_0(x) dx + \sum_{i=1}^{j_Q} \int_0^{f_i^{-1}(y)} p_i(x) \nu(x) dx$$

and

$$\mathcal{Q}'_\nu(y) = \int_{\mathbb{R}^+} \nu(x) \frac{\partial Q_1}{\partial y}(x, y) dx + \sum_{i=1}^{j_Q} p_i(f_i^{-1}(y)) \nu(f_i^{-1}(y)) (f_i^{-1})'(y).$$

Therefore, there exists a function  $\psi_2$  such that

$$\|\nu\|_{H^\alpha(\mathcal{I})} \leq \psi_2 \left( \|\Lambda\|_{H^\alpha(\mathcal{J})}, (\|f_i\|_{H^\alpha(\mathcal{J})})_{1 \leq i \leq j_Q}, (\|p_i\|_{H^{\alpha-1}(\mathcal{J})})_{1 \leq i \leq j_Q}, \|Q_1\|_{H^{\alpha-1}(\mathcal{J})} \right).$$

As  $\lambda \in H^\alpha(\mathcal{J})$  and  $\forall x, (\phi_x^{-1})' \in H^\alpha(\mathcal{J})$ , then  $\forall x, \Lambda(z, \cdot) \in H^\alpha(\mathcal{J})$  and there exists a continuous function  $\psi_1$  such that

$$\|\Lambda(\cdot, \cdot)\|_{H^\alpha([0, j_2] \times \mathcal{J})} \leq \psi_1 \left( \|(\phi^{-1})'\|_{H^\alpha([0, j_2] \times \mathcal{I})}, \|\lambda\|_{H^\alpha(\mathcal{J})} \right)$$

which ends the proof.

## A.6 Proof of Talagrand's inequality for beta-mixing variables

The following lemma is very useful to replace weak dependent variables by variables which are independent by blocks. It is proved by Viennet [31, proof of Proposition 5.1].

**Lemma 17** (Berbee's coupling lemma). *The random variables  $\{Y_k\}_{k \in \mathbb{N}}$  are exponentially  $\beta$ -mixing. Let us set  $q_n = \lfloor (r+1) \ln(n)/\beta \rfloor$  where  $\beta$  characterizes the  $\beta$ -mixing coefficient (see Definition 2). We have that  $\beta(q_n) \leq 1/n^{r+1}$ . We set  $p_n = n/(2q_n)$ . There exist random vectors  $(Y_1^*, \dots, Y_n^*)$  such that:*

- $Y_i$  and  $Y_i^*$  have same law.
- The random vectors  $(Y_{2kq_n+1}^*, \dots, Y_{(2k+1)q_n}^*)_{0 \leq k < p_n}$  are independent, as the random vectors  $(Y_{(2k+1)q_n+1}^*, \dots, Y_{(2k+2)q_n}^*)_{0 \leq k < p_n}$ .
- For any integer  $k$ ,  $0 \leq k \leq 2p_n - 1$ ,  $\mathbb{P} \left( Y_{kq_n+1}, \dots, Y_{(k+1)q_n} \neq (Y_{kq_n+1}^*, \dots, Y_{(k+1)q_n}^*) \right) \leq \beta_Y(q_n) \leq n^{-(r+1)}$ .

Let us set  $\Omega^* = \{\omega, \forall k, Y_k = Y_k^*\}$ . Then

$$\mathbb{P}(\Omega^{*c}) \leq n\beta_Y(q_n) \leq \frac{1}{n^r}.$$

This following inequality comes from Talagrand's inequalities (see Birgé and Massart [8, Corollary 2 p354]).

**Lemma 18** (Talagrand's inequality). *Let  $X_1, \dots, X_n$  be independent random variables and  $S$  a vectorial subspace of finite dimension  $D$  satisfying Assumption 4. We denote by  $\mathcal{F}$  a countable family of  $S$ . Let us set*

$$F_n(u) = \frac{1}{n} \sum_{k=1}^n u(X_k) - \mathbb{E}_{z_0}(u(X_k))$$

with  $u \in L^2$ . If

$$\sup_{u \in \mathcal{F}} \|u\|_\infty \leq M_2, \quad \mathbb{E}_{z_0} \left( \sup_{u \in \mathcal{F}} |F_n(u)| \right) \leq H, \quad \sup_{u \in \mathcal{F}} \text{Var}_{z_0}(u(X_k)) \leq V,$$

then

$$\mathbb{E}_{z_0} \left( \sup_{u \in \mathcal{F}} F_n^2(u) - 6H^2 \right)_+ \leq C \left( \frac{V}{n} \exp \left( -\frac{nH^2}{6V} \right) + \frac{M_2^2}{n^2} \exp \left( -k_2 \frac{nH}{M_2} \right) \right)$$

where  $C$  is a universal constant and  $k_2 = (\sqrt{2} - 1)/(21\sqrt{2})$ .

*Proof of lemma 18.* We apply Theorem 1.1 of Klein and Rio [23] to the functions  $s^i(u) = \frac{u(Y_i) - \mathbb{E}_{z_0}(u(Y_i))}{2M_2}$  (notation used in Theorem 1.1 of Klein and Rio [23]). We obtain that

$$\mathbb{P} \left( \sup_{u \in \mathcal{F}} |F_n(u)| \geq H + x \right) \leq \exp \left( -\frac{nx^2}{2(V + 4HM_2) + 6M_2x} \right).$$

We modify this inequality following Corollary 2 of Birgé and Massart [8]. It gives:

$$\mathbb{P} \left( \sup_{u \in \mathcal{F}} |F_n(u)| \geq (1 + \eta)H + x \right) \leq \exp \left( -\frac{n}{3} \min \left( \frac{x^2}{2V}, \frac{\min(\eta, 1)x}{7M_2} \right) \right).$$

The end of the proof is done in Comte and Merlevède [11, p222-223].  $\square$

*Proof of lemma 11.* To deduce lemma 11, we simply apply the Berbee's coupling lemma to exponential  $\beta$ -mixing variables, and then the Talagrand's inequality. Indeed, by Berbee's coupling lemma, as  $Y_k^*$  and  $Y_k$  have same law:

$$I_n(s) = \frac{1}{n} \sum_{k=1}^n s(Y_k^*) - \mathbb{E}_{z_0}(s(Y_k^*)) + s(Y_k) - s(Y_k^*).$$

We first bound the second part of the sum  $I_2(s) := \frac{1}{n} \sum_{k=1}^n s(Y_k) - s(Y_k^*)$ . We have:

$$I_2^2(s) = \frac{1}{n^2} \left( \sum_{k=1}^n (s(Y_k) - s(Y_k^*)) \mathbb{1}_{Y_k \neq Y_k^*} \right)^2 \leq \frac{4M_2^2}{n^2} \left( \sum_{k=1}^n \mathbb{1}_{\{Y_k \neq Y_k^*\}} \right)^2$$

By Cauchy-Schwartz,  $I_2^2(s) \leq \frac{4M_2^2}{n} \sum_{k=1}^n \mathbb{1}_{\{Y_k \neq Y_k^*\}}$  and by Berbee's coupling lemma,  $\mathbb{E}_{z_0}(\sup_{s \in \mathcal{B}} I_2(s)) \leq \frac{4M_2^2}{n^2}$ .

Let us now bound the first term  $I_1(s) := \frac{1}{n} \sum_{k=1}^n s(Y_k^*) - \mathbb{E}_{z_0}(s(Y_k^*))$ . We have

$$I_1 = \frac{1}{p_n} \sum_{j=0}^{p_n-1} u_s(X_{j,0}) - \mathbb{E}_{z_0}(u_s(X_{j,0})) + \frac{1}{p_n} \sum_{j=0}^{p_n-1} u_s(X_{j,1}) - \mathbb{E}_{z_0}(u_s(X_{j,1}))$$

where  $X_{j,i} := \left( Y_{2(j+i)q_n+1}^*, \dots, Y_{(2(j+i)+1)q_n}^* \right)$  and  $u_s(x_1, \dots, x_{q_n}) := \frac{1}{q_n} \sum_{k=1}^{q_n} s(x_k)$ . The random variables  $X_{j,0}$  are independent, the same can be said for  $X_{j,1}$ . Moreover,  $|X_{j,i}| \leq M_2$  and  $\text{Var}_{z_0}(X_{j,i}) \leq V$ . Let us set

$$I_{n,i}^*(s) := \frac{1}{p_n} \sum_{j=0}^{p_n-1} u_s(X_{j,i}) - \mathbb{E}_{z_0}(u_s(X_{j,i})).$$

We have:  $I_1(s) := (I_{n,0}^*(s) + I_{n,1}^*(s))/2$ . Then,

$$\begin{aligned} \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}} I_1^2(s) - 6H^2 \right)_+ &\leq \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}} \frac{1}{4} (2(I_{n,0}^*(s))^2 + 2(I_{n,1}^*(s))^2) - 6H^2 \right) \\ &\leq \sum_{i=0}^1 \mathbb{E}_{z_0} \left( \sup_{s \in \mathcal{B}} (I_{n,i}^*(s))^2 - 6H^2 \right)_+. \end{aligned}$$

As the dimension of  $S$  is finite, we can find a countable family  $\mathcal{F}$  dense in  $\mathcal{B}$  and we can then apply the Talagrand's inequality to  $I_{n,0}^*$  and  $I_{n,1}^*$  which concludes the proof.  $\square$

## Appendix B: Simulations

For the simulations, two very classical PDMP processes are considered: the TCP and the size of a marked bacteria.

**TCP protocol.** The transmission control protocol (TCP) is one of the main data transmission protocols in the Internet. The maximum number of packets that can be sent at time  $t_k$  in a round is a random variable  $X_{t_k}$ . If the transmission is successful, then the maximum number of packets is increased by one:  $X_{t_{k+1}} = X_{t_k} + 1$ . If the transmission fails, then  $X_{t_{k+1}} = \kappa X_{t_k}$  with  $\kappa \in (0, 1)$ . A correct scaling of this process leads to a piecewise deterministic Markov process  $(X_t)$  with the characteristics:

$$\phi(x, t) = x + ct, \quad Q(x, \{y\}) = \mathbb{1}_{\{y=\kappa x\}}, \quad \lambda.$$

Then the function  $(\phi_x^{-1})'$  is constant:  $(\phi_x^{-1})' = 1/c$ . Let us denote by  $\Lambda$  a primitive of  $\lambda$ . By (2), we have:

$$\mathbb{P}(Y_1 > y | Z_0 = x) = \exp\left(-\frac{1}{c}(\Lambda(y) - \Lambda(x))\right) \mathbb{1}_{\{y \geq x\}}.$$

As  $\lambda$  is positive, its primitive is invertible and by a change of variable:

$$\mathbb{P}(\Lambda(Y_1) > v | \Lambda(Z_0) = u) = \exp\left(-\frac{1}{c}(v - u)\right) \mathbb{1}_{\{v \geq u\}}.$$

Then  $\Lambda(Y_j) | \Lambda(Z_{j-1})$  follows an exponential law translated by  $\Lambda(Z_{j-1})$  and of parameter  $1/c$ . Therefore, if we can find the inverse of the function  $\Lambda$ , we can construct the sequence  $(Y_j, Z_j)$  by recurrence:

$$\Lambda(Y_j) = \Lambda(Z_{j-1}) + cE_j, \quad Z_j = \kappa Y_j \tag{38}$$

where  $E_j$  are i.i.d. of law  $\mathcal{E}(1)$ .

If  $\lambda(x) = \lambda x^\delta$  with  $\delta > -1$ , then  $Y_j^{\delta+1} = Z_{j-1}^{\delta+1} + c(\delta+1)/\lambda E_j$  and we obtain

$$Y_j = \sqrt[\delta+1]{Z_{j-1}^{\delta+1} + \frac{(\delta+1)c}{\lambda} E_j}.$$

This model satisfies Assumption (S). In order to have a model with a non-increasing function  $\lambda$ , we also consider the function  $\lambda(x) = (x-a)^2 + b$  with  $a > 0$ ,  $b \geq 0$ . In that case, by (38),

$$(Y_j - a)^3 + 3b(Y_j - a) = (Z_{j-1} - a)^3 + 3b(Z_{j-1} - a) + 3cE_j$$

and, by Cardan's formula, this equation has a unique real solution, which is

$$Y_j = a + \frac{\sqrt[3]{Q + \sqrt{4b^3 + Q^2}} + \sqrt[3]{Q - \sqrt{4b^3 + Q^2}}}{2}$$

where  $Q = 3cE_j + (Z_{j-1} - a)^3 + 3b(Z_{j-1} - a)$ . This model also satisfies Assumption (S).

**Bacterial growth.** We choose randomly a bacteria, and follow its growth, until it divides in two parts more or less equal. Then we choose randomly one of its daughter, and so on. Between the jumps, the bacteria grows exponentially. During a jump, the size of the bacteria is more or less divided by two. We model this by setting  $Z_k = Y_k \times U_k$ , where  $U_k$  is a random variable independent of  $Y_k$ , in  $(0, 1)$ , and centered in  $1/2$ . The Beta distribution  $\beta(\alpha, \alpha)$  satisfies these conditions. For  $\alpha = 1$ , it is the uniform distribution, and when  $\alpha$  increases, the distribution is more concentrated around  $1/2$ . We choose  $\alpha = 20$ . Then

$$\phi(x, t) = xe^{ct}, \quad Q(x, y) = \frac{1}{\beta(20, 20)} \frac{y^{19}(x-y)^{19}}{x^{38}} \mathbb{1}_{\{y \leq x\}}.$$

Then  $(\phi_x^{-1})'(y) = \frac{1}{y}$  and by (2),

$$\mathbb{P}(Y_1 > y | Z_0 = x) = \exp\left(-\frac{1}{c} \int_x^y \frac{\lambda(s)}{s} ds\right) \mathbb{1}_{\{y \geq x\}}.$$

We need to find a primitive of  $\lambda(x)/x$ . If  $\lambda(x) = \lambda x^\delta$ ,  $\delta > 0$ , then:

$$\mathbb{P}(Y_1 > y | Z_0 = x) = \exp\left(-\frac{\lambda}{\delta c} (y^\delta - x^\delta)\right) \mathbb{1}_{\{y \geq x\}}.$$

Therefore

$$\mathbb{P}(Y_1^\delta > y | Z_0^\delta \geq x) = \exp\left(-\frac{\lambda}{\delta c} (y - x)\right) \mathbb{1}_{\{y \geq x\}}$$

and the law of the random variable  $Y_k^\delta$  is an exponential translated by  $Z_{k-1}^\delta$  and of parameter  $\lambda/\delta c$ . Then

$$Y_k = \sqrt[\delta]{\frac{\delta c}{\lambda} E_k + Z_{k-1}^\delta}, \quad Z_k = Y_k U_k$$

with  $E_k \sim \mathcal{E}(1)$  i.i.d. and  $U_k \sim \beta(20, 20)$  i.i.d. All the conditions of Assumption (S) are satisfied, except point a. Indeed,  $\mathbb{P}(Z < Y) = 1$ , but there do not exists any  $\kappa < 1$  such that  $\mathbb{P}(Z \leq \kappa Y) = 1$ . However, in the simulations, it seems that the process is ergodic and that A3 is satisfied.

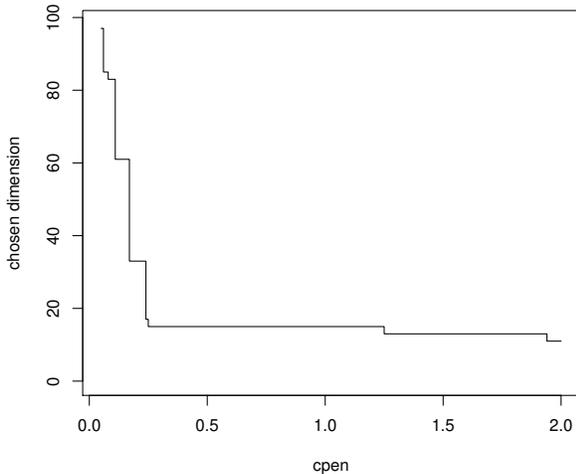
**Computations** For the two models,  $\nu$  has a density with respect to the Lebesgue measure on  $\mathbb{R}$ , so it can be estimated on any compact interval  $\mathcal{A}$ , here  $\mathcal{A} = [-1, 5]$  to avoid edge effects. The estimator is computed thanks to a projection on a trigonometric basis. The constant involved in  $pen(m)$ ,  $cpen$ , should be greater than  $\frac{3}{2}(\psi_1 + \psi_2 C_\lambda)$ , with  $\psi_1 = \psi_2 = \frac{1}{3}$ . The problem is that  $C_\lambda$ , a correlation term, is not easily tractable. We set  $cpen = \psi_1 + \psi_2 = 2/3$  for all models. This choice seems confirmed by the simulations results: the oracle *or* remains close to 1.

The constant  $cpen$  could be determined via the slope heuristic. Indeed, if the constant in the penalty is too small, the algorithm selects the maximal dimension. If the penalty is

large enough, it selects models of reasonable size. We then let the constant  $c$  in the penalty vary and note the dimension selected. For  $c$  smaller than a value  $c_{min}$ , the largest models are selected, and for  $c$  greater than  $c_{min}$ , smaller models are chosen. The "best" constant is  $c = 2c_{min}$ . See Arlot and Massart [1] for instance.

Figure 2 shows the selected dimension with respect to  $cpen$ , the constant in the penalty. When the constant in the penalty increases, the chosen dimension first decreases very rapidly, until  $cpen=0.24$ , then it decreases very slowly towards 1. Then  $2c_{min} = 0.48$ . Our chosen penalty constant,  $2/3$ , is a little greater than  $2c_{min}$ , and selects the same dimension (here 17).

Figure 2: Choice of the dimension



However, the slope heuristic involves quite a lot of computations, so it can not be used for every simulation, only to check that the penalty constant is coherent.

In Figures 3 and 4, for each graph, five simulations of the PDMP with  $n = 10^5$  are realized. For each simulation, the estimator  $\hat{\lambda}$ , the density  $\hat{\nu}_{\hat{m}}$  and  $\hat{\mathbf{D}}_n$  are drawn.

In the tables, 200 simulations for each 4-tuple  $(n, c, \kappa, \lambda)$  are computed. The estimation interval  $\mathcal{I} = [0.5, 2]$  is such that  $\mathbf{D}$  is greater than the threshold  $(\ln(n))^{-1}$  on  $\mathcal{I}$  for  $n = 10^{-5}$  for all our models. For each set of parameters, the mean of the selected dimension  $\hat{D}_m$ , the mean and the standard variation of the  $L^2$  error on  $\mathcal{I}$ , denoted by "risk" and "sd" are calculated. We also want to prove that our estimator is truly adaptive. As  $\nu$  is unknown, we can not check that  $\hat{m}$  is the better choice for estimating  $\nu$ . Instead, let us consider the estimator

$$\hat{\lambda}_m = \frac{\hat{\nu}_m}{\hat{\mathbf{D}}_n} \mathbb{1}_{\{\hat{\nu}_m \geq 0\}} \mathbb{1}_{\{\hat{\mathbf{D}}_n \geq (\ln(n))^{-1}\}}.$$

Then  $\hat{\lambda}_n = \hat{\lambda}_{\hat{m}}$ . The optimal dimension is

$$m_{opt} = \min_{m \in \mathcal{M}_n} \left\| \hat{\lambda}_m - \lambda \right\|_{L^2(\mathcal{I})}^2$$

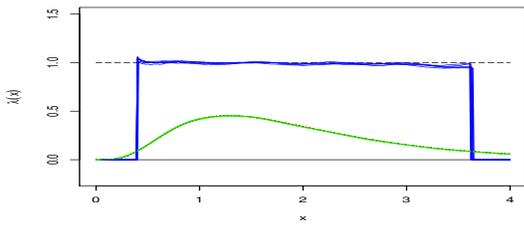
and the minimal risk  $\left\| \hat{\lambda}_{m_{opt}} - \lambda \right\|_{L^2(\mathcal{I})}^2$ . In the tables, we give the empirical means of  $D_{\hat{m}}$ ,  $D_{m_{opt}}$ , the empirical mean and standard deviation of the risk and the empirical mean of the oracle

$$or := \text{mean} \left( \left\| \hat{\lambda}_{\hat{m}} - \lambda \right\|_{L^2(\mathcal{I})}^2 / \left\| \hat{\lambda}_{m_{opt}} - \lambda \right\|_{L^2(\mathcal{I})}^2 \right).$$

In Figure 5, four simulations are realized, each for a different value of  $n$  ( $n = 10^2, 10^3, 10^4$  and  $10^5$ ) in order to show the convergence of our estimator.

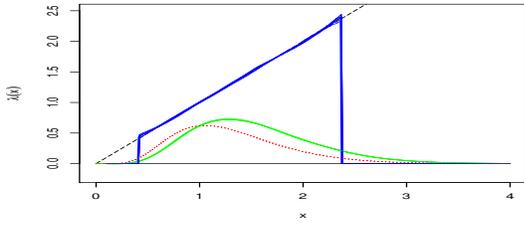
**Results** In Figures 3-4, the estimator  $\hat{\lambda}$  is very close to  $\lambda$ , at least when  $x$  is neither too small nor too large, that is when there are enough values to compute the estimator. The estimators  $\hat{\nu}_{\hat{m}}$  and  $\hat{\mathbf{D}}_n$  are quite smooth, whereas  $\hat{\lambda}$  tends to oscillate. This is due to the division of two estimators. In Tables 3-4, the risk decreases when  $n$  increases and seems to tend toward 0. The oracle remains close to 1, our estimator is really adaptive. When the number of observations is small, the risk may seem quite important (for instance, for figure 4 when  $\lambda(x) = x^2$ ). This is simply because  $\mathbf{D}$  is smaller than the threshold ( $1/\ln(10^2) = 0.2$ ), and the estimator  $\hat{\lambda}$  is set to 0 on some part of  $\mathcal{I}$ , or even on the whole interval. The estimation near 0 can be good for some models, for instance when  $\kappa = 1/5$  and  $\lambda(x) = x$ , because the random variables  $Z_k$  take smaller values (at a jump, we divide the process by 5 instead of by 2). The function  $\mathbf{D}$  then take higher values near 0, and the estimator  $\hat{\lambda}$  is positive even for small values of  $x$ . This problem is illustrated in Figure 5: when  $n$  increases, the estimator is better both because the support interval of  $\hat{\lambda}$  increases and because on the support interval, the estimator is closer to the true function.

Figure 3: TCP protocol:  $\phi(x, t) = x + t$ ,  $Q(x, y) = \delta_{y=\kappa x}$



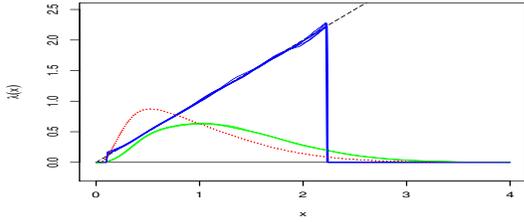
$$\kappa = 1/2, \lambda(x) = 1, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	5.1	6.7	0.12	0.06	1.27
$10^3$	8.3	10.1	$9.6\text{e-}3$	$9.6\text{e-}3$	1.72
$10^4$	12.3	14.4	$8.5\text{e-}4$	$5.5\text{e-}4$	1.64
$10^5$	17.1	18.6	$1.2\text{e-}4$	$6.6\text{e-}5$	1.47



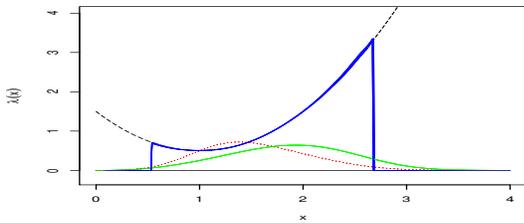
$$\kappa = 1/2, \lambda(x) = x, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	7.1	8.5	0.24	0.20	1.16
$10^3$	10.4	12.8	$9.5\text{e-}3$	$6.5\text{e-}3$	1.64
$10^4$	14.1	16.4	$1.1\text{e-}3$	$7.3\text{e-}4$	1.49
$10^5$	18.1	20.5	$1.2\text{e-}4$	$6.9\text{e-}5$	1.31



$$\kappa = 1/5, \lambda(x) = x, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	6.6	6.7	0.54	0.24	1.08
$10^3$	10.9	10.7	0.077	0.083	1.40
$10^4$	18.0	15.9	$1.5\text{e-}3$	$1.0\text{e-}3$	2.17
$10^5$	25.9	22.4	$2.0\text{e-}4$	$1.2\text{e-}4$	1.81



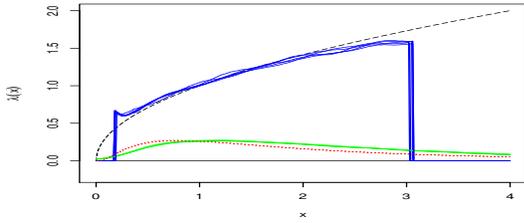
$$\kappa = 1/2, \lambda(x) = (x - 1)^2 + 1/2, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	5.4	6.6	0.080	0.014	1.06
$10^3$	7.3	8.9	0.045	$5.4\text{e-}3$	1.01
$10^4$	9.3	11.3	0.027	$2.3\text{e-}3$	1.004
$10^5$	11.7	16.3	0.014	$5.9\text{e-}4$	1.003

-- : true  $\lambda$     - : estimated  $\hat{\lambda}$     . . . : estimated  $D_n$     - : estimated  $\hat{v}_{\hat{m}}$

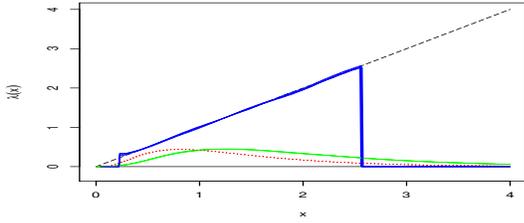
Figure 4: Bacterial growth

$$\phi(x, t) = x \exp(ct), \quad \lambda(x) = x^\delta, \quad Q(x, y) = \beta(20, 20) \frac{y^{19}(x - y)^{19}}{y^{39}} \mathbb{1}_{\{y \leq x\}}$$



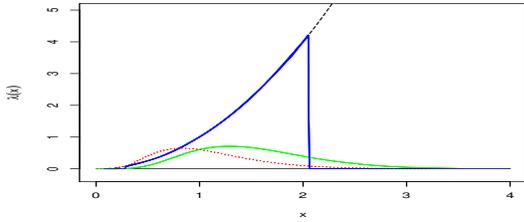
$$\lambda(x) = \sqrt{x}, c = 1, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	3.8	6.1	0.72	0.29	1.04
$10^3$	6.6	7.8	0.014	0.021	2.32
$10^4$	11.2	10.7	$2.4e-3$	$1.5e-3$	1.97
$10^5$	17.7	16.9	$9.0e-4$	$3.1e-4$	1.29



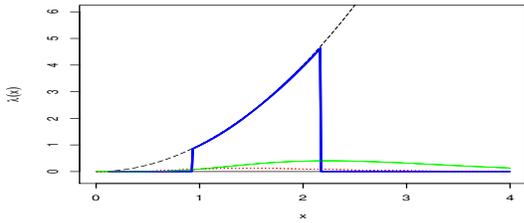
$$\lambda(x) = x, c = 1, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	5.2	6.3	0.59	0.26	1.07
$10^3$	8.2	9.1	$9.2e-3$	$9.0e-3$	2.05
$10^4$	12.0	13.2	$1.2e-3$	$8.1e-4$	1.60
$10^5$	17.3	17.1	$2.2e-4$	$1.4e-4$	1.54



$$\lambda(x) = x^2, c = 1, \mathcal{I} = [0.5, 2]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	7.1	7.9	2.64	0.35	1.01
$10^3$	10.1	11.7	1.48	0.21	1.003
$10^4$	13.5	14.2	0.41	0.094	1.002
$10^5$	17.4	17.5	$2.8e-4$	$2.5e-4$	1.62

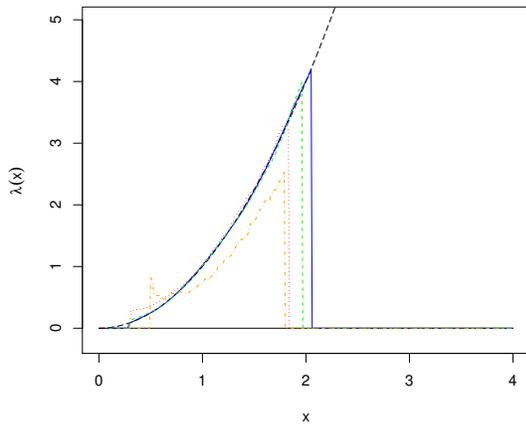


$$\lambda(x) = x^2, c = 3, \mathcal{I} = [0.5, 2]$$

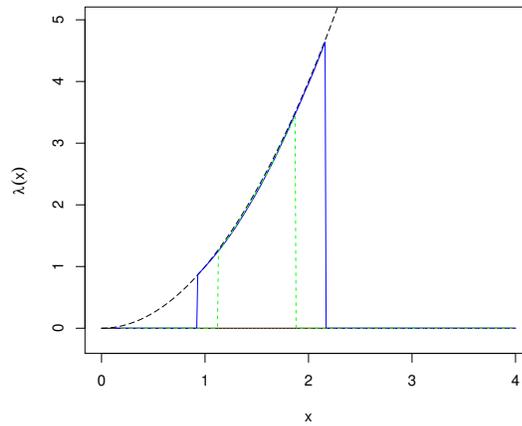
$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	sd	or
$10^2$	3.8	3.0	4.29	0	1
$10^3$	5.7	3.0	4.29	0	1
$10^4$	7.8	11.1	1.56	0.11	1.001
$10^5$	10.0	13.9	0.087	0.0026	1.001

-- : true  $\lambda$     - : estimated  $\hat{\lambda}$     ... : estimated  $D_n$     - : estimated  $\hat{\nu}_{\hat{m}}$

Figure 5: Convergence of the estimator (size of a marked bacteria)  
 $\phi(x, t) = xe^t, \quad \lambda(x) = x^2$                        $\phi(x, t) = xe^{3t}, \quad \lambda(x) = x^2$



Estimator's support increases with  $n$   
 $n = 10^5$ ,  $-\cdot-\cdot n = 10^4$ ,  $\cdot\cdot\cdot\cdot n = 10^3$ ,  $-\cdot n = 10^2$ .



The estimator is null if  $n \leq 10^3$