



**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. 2019. hal-01996064v1

**HAL Id: hal-01996064**

**<https://hal.science/hal-01996064v1>**

Preprint submitted on 28 Jan 2019 (v1), last revised 8 Dec 2020 (v2)

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

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

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

N. Krell\*, É. Schmisser†

January 28, 2019

## Abstract

In this paper, we consider a piecewise deterministic Markov process (PDMP), with known flow and deterministic transition measure, and unknown jump rate  $\lambda$ . 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  $\lambda$ . We provide uniform bounds for the risk of these estimators, and prove that the estimator of the jump rate is nearly minimax (up to a  $\ln^2(n)$  factor). Simulations illustrate the behavior of our estimator.

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

**Subject Classification:** 62M05, 62G05, 62G07, 62G20, 60J75, 60J25

## 1 Introduction

The PDMP is a large class of models, they are used to model deterministic phenomena in which randomness appears as point events, such as transmission control protocol (TCP) window size, the size of a marked bacteria, risk processes in mathematical insurance, stress release in seismology, . . . . See [Rudnicki and Tyran-Kamińska \[2015\]](#) for a nice presentation of biological problem in which PDMP appear. In [Azaïs et al. \[2014\]](#), you will find a presentation of the pdmp with a large field of application.

The TCP protocol (see [Dumas et al. \[2002\]](#), [Guillemin et al. \[2004\]](#) 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  $(X_t)$ . Another example of PDMP is the size of a marked bacteria (see [Doumic et al. \[2015\]](#), [Laurençot and Perthame \[2009\]](#)). We choose randomly a bacteria, and follow its growth, until it divides in two. Then we choose randomly one of its daughter, and so on. Between the jumps, the bacteria grows exponentially.

More generally, we consider a filtered piecewise deterministic Markov process (PDMP)  $(X_t)_{t \geq 0}$  with flow  $\phi$ , deterministic transition measure  $Q(x, \cdot) = \mathbb{1}_{\{f(x) \in \cdot\}}$  (with  $f$  a deterministic known function) and jump rate  $\lambda$ . 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  $\lambda(\phi(x, t))$ .

---

\*Univ Rennes, CNRS, IRMAR - UMR 6625, F-35000 Rennes, France **email:** nathalie.krell@univ-rennes1.fr

†Université de Lille 1, Laboratoire Paul Painlevé, CNRS UMR 8524. Bâtiment M2, Cité Scientifique, 59655 Villeneuve d'Ascq, France. **email:** emeline.schmisser@math.univ-lille1.fr

The process restart from  $f(X_{T_1})$  as before. For the TCP protocol,  $f(x) = \kappa x$  and the flow is additive ( $\phi(x, t) = x + ct$ ), for the marked bacteria,  $f(x) = x/2$  and the flow is multiplicative ( $\phi(x, t) = xe^{ct}$ ). As the process  $(X_t)$  is observed without errors, the flow  $\phi$  and the transition measure  $f$  are known. We denote  $(T_1, \dots, T_n)$  the times of jump and consider the Markov chain  $(Z_0 = x_0, Z_k = f(X_{T_k}), k > 0)$ . The aim of this paper is to construct an adaptive estimator of the jump rate  $\lambda$  on the compact interval  $\mathcal{I}$ . In the meantime, we also provide an adaptive estimator of  $\nu_\lambda$ .

As far as we know, there exist few results for nonparametric estimation for PDMPs. [Fujii \[2013\]](#) and [Azaïs and Muller-Gueudin \[2016\]](#) consider a very general model. The process  $(X_t)$  takes values in  $\mathbb{R}^d$ , and jumps can occur deterministically if the process reaches the boundary of an open set  $E$ . The transition measure is not deterministic (the deterministic case is even excluded):  $X_{T_1} | X_{T_1^-}$  is a random variable, its law is a known function of  $X_{T_1^-}$ . The process  $(Z_k)$  is assumed to be ergodic. [Fujii \[2013\]](#) constructs an estimator of  $\nu_\lambda$  thanks to local times, and pointwise kernel estimators of the jump rate and the transition measure. He proves the consistency of his estimators, but do not give any rate of convergence. [Azaïs and Muller-Gueudin \[2016\]](#) construct kernel pointwise estimators of the stationary density and the jump rate  $\lambda$ . They show that their estimator is consistent and prove its asymptotic normality. They choose an adaptive estimator of  $\lambda$  by taking the estimator of the minimal variance and bound its risk. In the contrary of the two previous works we give some explicit assumptions that ensure that the process  $(Z_k)$  is strongly ergodic and has a unique invariant density  $\nu_\lambda$ . In the same setting as ours, [Krell \[2016\]](#) construct a pointwise kernel estimator of  $\lambda$  and prove its normality, but do not provide a bound for the risk of the adaptive estimator. To our knowledge, there do not exist results for the  $L^2$ -risk in our settings, neither a minimax rate of convergence.

We construct an adaptive estimator of the jump rate  $\lambda$  and bounds its  $L^2$  risk uniformly. For this purpose, as in [Krell \[2016\]](#), we use the equality

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

where  $\nu_\lambda$  is the stationary density of the random variables  $Z_k$  and the function  $\mathbf{D}(x)$  can be estimated nearly like a cumulative distribution function, it converges with rate  $n^{1/2}$  ([Krell \[2016\]](#)). To estimate the density function  $\nu_\lambda$ , we use a projection method. We choose a series of increasing vectorial subspaces  $S_0, \dots, S_m$ . On each subspace, we construct a nonparametric estimator  $\hat{\nu}_m$  of  $\nu_\lambda$ , belonging to  $S_m$  and provide a uniform bound for its risk. Then we choose the "best" estimator by a penalisation method, in the same way as [Barron et al. \[1999\]](#), and give an oracle inequality for the adaptive estimator  $\hat{\nu}_{\hat{m}}$ . Afterwards we bound uniformly the risk of the resulting quotient estimator  $\hat{\lambda} = \hat{\nu}_{\hat{m}} / \hat{\mathbf{D}}_n$ . We finally prove that the estimator  $\hat{\lambda} := \hat{\nu}_{\hat{m}} / \hat{\mathbf{D}}_n$  is nearly minimax (up to a  $\ln^2(n)$  term) using the methodology of [Tsybakov \[2004\]](#).

In [Section 2](#), we specify the model and its assumptions. We focus on the estimation of the stationary density on the estimation of the jump rate  $\lambda$  in [3](#). [Section 4](#) is devoted to simulations for the TCP protocol and the bacterial growth, with various functions  $\lambda$ . The outcomes are consistent with the theoretical results. Proofs are gathered in [Section 5](#) and in the [Appendix](#) where we prove a technical result, a Talagrand inequality for exponentially  $\beta$ -mixing variables.

## 2 PDMP

In general a piecewise deterministic Markov process 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 at the jump time. In this article, we consider a specific class of PDMP for which the transition measure  $Q$  is a deterministic kernel  $Q(x, A) = \mathbb{1}_{\{f(x) \in A\}}$ . More precisely,

**Assumption A1.**

- a. The flow  $\phi : [0, \infty) \times [0, \infty) \mapsto [0, \infty)$  is a one-parameter group of homeomorphisms:  $\phi$  is  $\mathcal{C}^1$ ,  $\phi(\cdot, t)$  is an homeomorphism for each  $t \in [0, \infty)$ , satisfying the semigroup property:  $\phi(\cdot, t + s) = \phi(\phi(\cdot, s), t)$  and  $\phi_x(\cdot) := \phi(x, \cdot)$  is an  $\mathcal{C}^1$ -diffeomorphism. This implies that  $\phi(x, 0) = x$ .
- b. The jump rate  $\lambda : [0, \infty) \rightarrow \mathbb{R}_+$  is assumed to be a measurable function satisfying

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

that is, the jump rate does not explode.

- c. The Markov kernel  $Q(x, A) = \mathbb{1}_{\{f(x) \in A\}}$  where  $f : [0, \infty) \mapsto [0, \infty)$  is an increasing  $\mathcal{C}^1$ -diffeomorphism.

For instance, we can take  $\phi(x, t) = x + ct$  (linear flow) or  $\phi(x, t) = xe^{ct}$  (exponential flow).

Given these three characteristics, it can be shown (Davis [1993, 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  $\{X(t)\}$  starting from a point  $x_0 \in \mathbb{R}^+$  may be constructed as follows. Consider a random variable  $T_1$  with survival function

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

If  $T_1$  is equal to infinity, then the process  $X$  follows the flow, i.e. for  $t \in \mathbb{R}_+$ ,  $X(t) = \phi(x_0, t)$ . Otherwise let

$$Z_1 = f(\phi(x_0, T_1)). \quad (2)$$

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  $\{X(t)\}$  with jump times  $\{T_k\}_{k \in \mathbb{N}}$  (where  $T_0 = 0$ ). A very natural Markov chain is linked to  $\{X(t)\}$ , namely the jump chain  $(Z_n)_{n \in \mathbb{N}}$ .

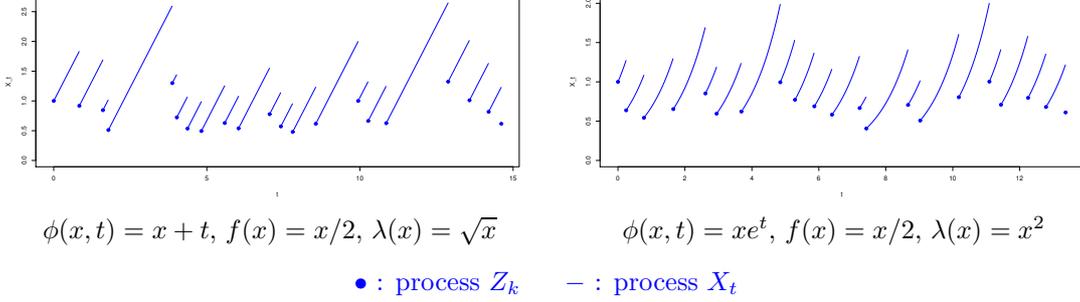
To simplify the notations, let us set  $\phi_x(t) = \phi(x, t)$ . By (1) and (2),

$$\begin{aligned} \mathbb{P}(Z_1 \geq y | Z_0 = x_0) &= \mathbb{P}(T_1 \geq (f \circ \phi_{x_0})^{-1}(y) | Z_0 = x_0) = \exp\left(-\int_0^{(f \circ \phi_{x_0})^{-1}(y)} \lambda(\phi_{x_0}(s)) ds\right) \\ &= \exp(-\Lambda((f \circ \phi_{x_0})^{-1}(y))) \end{aligned}$$

and by the change of variable  $u = f \circ \phi_{x_0}(s)$ , we get

$$\mathbb{P}(Z_1 \geq y | Z_0 = x_0) = \exp\left(-\int_{f(x_0)}^y \lambda(f^{-1}(u)) g_{x_0}(u) du\right) \mathbb{1}_{y \geq f(x_0)} \quad (3)$$

Figure 1: Exemples of simulations of processes  $(X_t)$  et  $(Z_k)$   
TCP protocol Bacterial growth



where  $g_x(y) = \left[ (f \circ \phi_x)^{-1} \right]'(y) \geq 0$ . By the monotonicity of  $f \circ \phi_x$ , we get the transition probability of the Markov chain:

$$\mathcal{P}_\lambda(x, dy) := \mathbb{P}(Z_1 \in dy | Z_0 = x) = \lambda(f^{-1}(y))g_x(y)e^{-\int_{f(x)}^y \lambda(f^{-1}(u))g_x(u)du} \mathbb{1}_{\{y \geq f(x)\}} dy. \quad (4)$$

We need some assumptions, in particular to ensure that the process is ergodic, which is often a keystone in statistical inference for Markov process. The Hölder space  $\mathcal{H}^\alpha(\mathcal{I})$  and the associated norm would be define in the appendix. We first give an assumption on the known functions  $f$  and  $\phi$ :

**Assumption A2.** We suppose that for a compact interval of  $(0, \infty)$ ,  $f : [0, \infty) \mapsto [0, \infty)$  and  $\phi_x : [0, \infty) \times [0, \infty) \mapsto [0, \infty)$  are known and follow, for  $M_2 > 0$ ,  $0 < \kappa < 1$ ,  $\alpha > 0$ , and two positive continuous functions  $\mathbf{m} : [0, \infty) \mapsto (0, \infty)$  and  $\mathbf{M} : [0, \infty) \mapsto (0, \infty)$  such that for all  $x \in [0, \infty) : \mathbf{M}(x) \geq \mathbf{m}(x) > 0$ ,

$$g_x \in \mathcal{H}^\alpha(\mathcal{I}), \quad \|f\|_{L^\infty(\mathcal{I})} \leq M_2, \quad \text{and} \quad f^{-1} \in \mathcal{H}^\alpha(\mathcal{I})$$

$$\forall x > 0, \quad 0 < f(x) \leq \kappa x, \quad (5)$$

$$\forall y > 0, \quad \forall x \geq 0, \quad \mathbf{m}(y) \leq g_x(y) \leq \mathbf{M}(y). \quad (6)$$

To obtain ergodicity and uniform bounds for the estimation of the jump rate  $\lambda$ , we consider the set of functions:

**Definition A3.** For  $b > 0$  and a vector of positive constants  $\mathbf{c} = (r, L, \mathbf{a})$ , such that  $r > i_2$  and

$$r \geq f^{-1} \left( \left[ \frac{-\ln(1 - \kappa^{b+1})}{1 - \kappa^{b+1}} \right]^{1/(b+1)} \right)$$

we introduce the class  $\mathcal{F}(\mathbf{c}, b)$  of continuous functions  $\lambda : [0, \infty) \mapsto [0, \infty)$ , such that

a.

$$\forall x \geq r, \quad \frac{(f(x))^b}{\mathbf{m}(f(x))} \leq \frac{\lambda(x)}{\mathbf{a}},$$

b.

$$\int_0^{f(r)} \mathbf{M}(u) \lambda(f^{-1}(u)) du \leq L. \quad (7)$$

Item A3a ensures that the right tail of the distribution of  $Z_1|Z_0$  is bounded:

$$\forall x \geq f^{-1}(r), \mathbb{P}(Z_1 \geq y|Z_0 = x) \leq \exp\left(-\frac{a}{b+1}(y^{b+1} - f(x)^{b+1})\right).$$

We get the following proposition, which is proved in Krell [2016].

**Proposition 1** (Ergodicity).

a. Under A1 - A3, if there exists  $(\mathbf{c}, b)$  such that  $\lambda \in \mathcal{F}(\mathbf{c}, b)$ , the Markov chain  $(Z_k)_{k \geq 0}$  is recurrent positive and strongly ergodic. There exists a unique invariant probability measure of the form

$$\nu_\lambda(dx) = \nu_\lambda(x)dx \text{ on } [0, \infty).$$

b. Moreover, for all  $y \in [0, \infty)$  we have the relation:

$$\lambda(y) = \frac{\nu_\lambda(f(y))}{\mathbf{D}(y)} \quad \text{where} \quad \mathbf{D}(y) := \mathbb{E}_{\nu_\lambda}(g_{Z_0}(f(y))1_{\{Z_0 \leq y \leq f^{-1}(Z_1)\}}).$$

*Remark.* The set  $\mathcal{F}(\mathbf{c}, b)$  is the same as in [Krell, 2016, Definition 3.2]. Indeed, we have that  $\int_{f(r)}^\infty \mathbf{m}(u) \lambda(f^{-1}(u)) du \geq \int_{f(r)}^\infty u^b = \infty$ , moreover,  $\int_{f(r)}^r \mathbf{M}(u) \lambda(f^{-1}(u)) du \geq \int_{f(r)}^r \frac{\mathbf{M}(u)}{\mathbf{m}(u)} u^b =: l$ .

The jump rate  $\lambda$  is very difficult to estimate directly, but it is related to the stationary density  $\nu_\lambda$ , which is simpler to estimate. To estimate the jump rate, we construct a quotient estimator, which is possible only if  $\mathbf{D}(y) > 0$  on the interval of estimation  $\mathcal{I} := [i_1, i_2]$ . We can remark that if  $\lambda > 0$  on the interval  $[i_1, \infty)$ , then  $\nu_\lambda(y) > 0$  on  $(f(i_1), \infty)$  and

$$\mathbf{D}(y) = \frac{\nu_\lambda(f(y))}{\lambda(y)} > 0 \quad \forall y > i_1.$$

Therefore, as  $\mathcal{I}$  is compact, there exists  $D_0 > 0$  such that  $\forall y \in \mathcal{I}, \mathbf{D}(y) \geq D_0$ .

*Remark.* In Krell [2016], the author bounds uniformly its estimator for any compact set  $\mathcal{D}$  included in  $[d(\mathbf{c}), \infty[$ , where  $d(\mathbf{c})$  is unknown and depends on the family  $\mathcal{F}(\mathbf{c}, b)$ . In this paper, we want uniform bounds on the chosen compact interval  $\mathcal{I}$ . By definition A3a, we already know that  $\forall x \geq r, \lambda(x) > 0$ , it remains to control what happens between  $i_1$  and  $r$ .

**Definition A4.** Let  $\varepsilon > 0$  and  $\bar{\mathbf{c}} = (\mathbf{c}, \varepsilon)$ . We introduce the set of positive functions  $\mathcal{E}(\bar{\mathbf{c}}, b)$   $\lambda : [0, \infty) \mapsto [0, \infty)$  such that  $\lambda \in \mathcal{F}(\mathbf{c}, b)$  and

$$\forall x \in [i_1, r], \lambda(x) \geq \varepsilon$$

Actually we have a precise result on the convergence to the unique invariant probability, which would be useful for the convergence result.

**Proposition 2.** Under A1-A3,

a. *Contraction. Let us set*

$$\mathbb{V}(x) := \exp\left(\frac{a}{b+1}(f(x))^{b+1}\right) \quad \forall x \in \mathbb{R}^+.$$

There exist  $0 < \gamma < 1$  and a constant  $R \in \mathbb{R}^{+*}$  such that, for any  $x \in \mathbb{R}^+$ :

$$\sup_{\lambda \in \mathcal{F}(\mathbf{c}, b)} \sup_{|\psi| \leq \mathbb{V}} \left| \mathcal{P}_\lambda^k \psi(x) - \int_0^\infty \psi(z) \nu_\lambda(z) dz \right| \leq R \mathbb{V}(x) \gamma^k$$

where  $|\psi| \leq \mathbb{V}$  means:  $\forall x, |\psi(x)| \leq \mathbb{V}(x)$ . That is,  $\forall \lambda \in \mathcal{F}(\mathbf{c}, b), \forall |\psi| \leq \mathbb{V}$ ,

$$\left| \mathbb{E}(\psi(Z_k) | Z_0 = x) - \int_0^\infty \psi(z) \nu_\lambda(z) dz \right| \leq R \mathbb{V}(x) \gamma^k.$$

The constants  $R$  and  $\gamma$  depends explicitly on  $(\mathbf{c}, b)$ . As  $\mathbb{V} \geq 1$ , the bound is true for any function  $\psi$  such that  $\|\psi\|_\infty \leq 1$ .

b. *The function  $\mathbb{V}(y) \nu_\lambda(y)$  is uniformly integrable on  $\mathcal{F}(\mathbf{c}, b)$ :*

$$\sup_{\lambda \in \mathcal{F}(\mathbf{c}, b)} \int \nu_\lambda(y) \mathbb{V}(y) dy < \infty.$$

This proposition is proved in [Krell, 2016, Section 4.1, Proof of Proposition 3.3 and Section 4.2, Speed of convergence for the empirical measure]. The following corollary is a consequence of Proposition 2 a.

**Corollary 3.** *Under A1-A3, for any bounded function  $s$ , if  $\lambda \in \mathcal{F}(\mathbf{c}, b)$ :*

$$\left| \mathbb{E} \left( \frac{1}{n} \sum_{k=1}^n s(Z_k) - \int s(x) \nu_\lambda(x) dx \right) \right| \leq \|s\|_\infty \frac{R \mathbb{V}(x_0)}{n(1-\gamma)}$$

and

$$\text{Var} \left( \frac{1}{n} \sum_{k=1}^n s(Z_k) \right) \leq \frac{1}{n} \int s^2(z) \nu_\lambda(z) dz + \frac{\|s\|_\infty}{n} \int |s(z)| \nu_\lambda(z) G_\lambda(z) dz + \frac{c_\lambda \|s\|_\infty^2}{n^2}$$

where  $G_\lambda(z) = \frac{R}{1-\gamma} (\mathbb{V}(z) + \int \mathbb{V}(x) \nu_\lambda(x) dx)$  and  $c_\lambda$  are uniformly bounded on  $\mathcal{F}(\mathbf{c}, b)$ .

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

The following lemma is proved in Section 5.

**Lemma 4.** *Under A1-A4, there exists  $\eta > 0, D_0 > 0$  such that:*

$$\inf_{\lambda \in \mathcal{E}(\bar{\mathbf{c}}, b)} \inf_{y \in f(\mathcal{I})} \nu_\lambda(y) \geq \eta \quad \text{and} \quad \inf_{\lambda \in \mathcal{E}(\bar{\mathbf{c}}, b)} \inf_{y \in \mathcal{I}} \mathbf{D}(y) \geq D_0.$$

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

**Definition 5.** Let us define the  $\sigma$ -algebra

$$\mathcal{O}_a^b = \sigma(\{Z_{j_1} \in I_1, \dots, Z_{j_n} \in I_n\}, a \leq j_1 \leq \dots \leq j_n \leq b, \mathbf{n} \in \mathbb{N}, I_k \in \mathcal{B}([0, \infty[))).$$

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

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

The  $\beta$ -mixing coefficient characterizes the dependence between what happens before  $T_k$  and what happens after  $T_{t+k}$ . The process  $(Z_k)_{k \geq 0}$  is  $\beta$ -mixing if  $\lim_{t \rightarrow \infty} \beta_Z(t) = 0$ . It is exponentially (or geometrically)  $\beta$ -mixing if there exists  $c, \beta$  such that  $\beta_Z(t) \leq ce^{-\beta t}$ .

The following corollary is a consequence of Proposition 2. It is proved in Section 5.

**Corollary 6.** We work under A1-A3, if there exists  $(\mathbf{c}, b)$  such that  $\lambda \in \mathcal{F}(\mathbf{c}, b)$ , the PDMP is geometrically  $\beta$ -mixing. Moreover, there exists  $c$  such that,  $\forall t > 0$ :

$$\sup_{\lambda \in \mathcal{F}(\mathbf{c}, b)} \beta_X(t) \leq c\gamma^t.$$

## 3 Estimation of the jump rate

### 3.1 The observation scheme

As in Krell [2016] 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:  $(X(0), X(T_i), 1 \leq i \leq n) = (Z_i, 0 \leq i \leq n)$  is sufficient, as one can remark that for all  $n \geq 1$ :  $T_n = (f \circ \phi_{Z_{n-1}})^{-1}(Z_n)$ .

### 3.2 Methodology

Krell [2016] constructed a pointwise kernel estimator of  $\nu_\lambda$  before deriving an estimator of  $\lambda$ . Densities are often approximated thanks by kernels methods (see Tsybakov [2004] for instance). Indeed, 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. Obtaining an adaptive estimator with kernel methods involves a double convolution, and therefore intensive computations. On the contrary, 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.

We first aim at estimating  $\nu_\lambda$  on a compact set  $\mathcal{A} \supseteq f(\mathcal{I})$  where  $\mathcal{I}$  is the estimation set of the jump rate  $\lambda$ . We construct  $L^2$  estimators by projection on an orthonormal basis. As usual in nonparametric estimation, the risk of our estimator can be decomposed in a variance term and a bias term which depends of the regularity of the density function  $\nu_\lambda$ . We choose to use the Besov spaces to characterize the regularity, which are well adapted to  $L^2$  estimation (particularly for the wavelets decomposition). See Appendix A for the definition of Besov spaces.

It is quite difficult to estimate a function nonparametrically. To do so, we introduce a sequence of vectorial subspaces  $S_m$ . We construct an estimator  $\hat{\nu}_m$  of  $\nu_\lambda$  on each subspace and then select the best estimator  $\hat{\nu}_{\hat{m}}$ .

**Assumption A5.**

- 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:

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

where  $s_m$  is the orthogonal projection of  $s$  on  $S_m$ .

The subspaces generated by wavelets, piecewise polynomials or trigonometric polynomials satisfy these assumption (see DeVore and Lorentz [1993] for trigonometric polynomials and Meyer [1990] for wavelets and piecewise polynomials). Conditions **a**, **b** and **d** are usual (see Comte et al. [2007, section 2.3] for instance). Condition **c** is necessary because we are not in the stationary case: it helps us to control some covariance terms.

### 3.3 Estimation of the stationary density

Let us now construct an estimator  $\hat{\nu}_m$  of  $\nu_\lambda$  on the vectorial subspace  $S_m$ . We consider an orthonormal basis  $(\varphi_l)$  of  $S_m$  satisfying Assumption A5. Let us set

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

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

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

**Proposition 7.** *If  $D_m^2 \leq n$ , under A1-A3 and A5, for any  $\lambda \in \mathcal{F}(\mathbf{c}, b)$ ,*

$$\mathbb{E} \left( \|\hat{\nu}_m - \nu_\lambda\|_{L^2(\mathcal{A})}^2 \right) \leq \|\nu_m - \nu_\lambda\|_{L^2(\mathcal{A})}^2 + C_{\mathbf{c},b} \frac{D_m}{n} + \frac{c_{\mathbf{c},b}}{n}$$

where  $C_{\mathbf{c},b} = \psi_1 + \psi_2 \int G_\lambda(z) \nu_\lambda(z) dz$  and  $G_\lambda(z) = \frac{R}{1-\gamma} (\mathbb{V}(z) + \int \mathbb{V}(z) \nu_\lambda(z) dz)$  (see Corollary 3).

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

$$\mathbb{E} \left( \|\hat{\nu}_{m_{opt}} - \nu_\lambda\|_{L^2(\mathcal{A})}^2 \right) \leq C \|\nu_\lambda\|_{B_{2,\infty}^\alpha} n^{-2\alpha/(2\alpha+1)}.$$

This is the usual nonparametric convergence rate (see [Tsybakov \[2004\]](#)).

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  $\nu_\lambda$ . Let us introduce the contrast function  $\gamma_n(s) = \|s\|_{L^2}^2 - \frac{2}{n} \sum_{k=1}^n s(Z_k)$ . If  $s \in S_m$ , then  $s = \sum_l b_l \varphi_l$  and

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

The minimum is obtained for  $\hat{a}_l = \frac{1}{n} \sum_{k=1}^n \varphi_l(Z_k)$ . Therefore  $\hat{\nu}_m = \arg \min_{s \in S_m} \gamma_n(s)$ . 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  $pen(m) = \frac{3C_\lambda D_m}{2n} + \frac{3c_\lambda}{2n}$  (or more generally  $pen(m) = \frac{\sigma D_m}{n} + \frac{\sigma'}{n}$ , with  $\sigma \geq \frac{3C_\lambda}{2}$ ,  $\sigma' \geq \frac{3c_\lambda}{2}$ ) and choose

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

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

**Theorem 8** (Risk of the adaptive estimator). *Under [A1-A3](#) and [A5](#), for any  $\lambda \in \mathcal{F}(c, b)$ ,  $\forall \sigma \geq \frac{3}{C} \lambda^2$ ,  $\sigma' \geq \frac{3c_\lambda}{2}$ ,  $pen(m) = \frac{\sigma D_m}{n} + \frac{\sigma'}{n}$ ,*

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

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  $\nu_\lambda$  belongs to some (unknown) Besov space  $B_{2,\infty}^\alpha$ :

$$\sup_{\lambda \in \mathcal{F}(c, b), \|\nu_\lambda\|_{B_{2,\infty}^\alpha} \leq C} \mathbb{E} \left( \|\nu_\lambda - \hat{\nu}_{\hat{m}}\|_{L^2(\mathcal{A})}^2 \right) \leq C' n^{-2\alpha/(2\alpha+1)}. \quad (8)$$

### 3.4 Estimation of the jump rate

According to [Proposition 1b](#),

$$\lambda(y) = \frac{\nu_\lambda(f(y))}{\mathbf{D}(y)} \quad \text{where} \quad \mathbf{D}(y) := \mathbb{E}_{\nu_\lambda} (g_{Z_0}(f(y)) \mathbf{1}_{Z_0 \leq y} \mathbf{1}_{Z_1 \geq f(y)}) .$$

To estimate the jump rate, we therefore use a quotient estimator. Unfortunately, the Besov spaces are not stable by product or convolution (as they are subspaces of  $L^2(\mathcal{A})$ ). To relate the risk of  $\hat{\lambda}$ , the estimator of the jump rate, to the regularity of  $\lambda$  (and not  $\nu_\lambda$ ), we need to consider a smaller class of functions, the Hölder spaces (see [Appendix A](#)).

With Hölder regularity, we can relate the regularity of  $\nu_\lambda$  to the regularity of  $\lambda$ . By [Krell, 2016, Lemma 3.5], if  $\lambda \in \mathcal{E}(\bar{c}, b) \cap H^\alpha(\mathcal{I}, M_1)$ , as  $g_x$  and  $f^{-1}$  belong to the Hölder space  $H^\alpha(\mathcal{I})$ , then

$$\|\nu_\lambda\|_{H^\alpha(f(\mathcal{I}))} \leq \psi(\mathcal{I}, \|\lambda\|_{H^\alpha(\mathcal{I})}, \|g_x\|_{H^\alpha(\mathcal{I})}, \|f^{-1}\|_{H^\alpha(\mathcal{I})}) \quad (9)$$

for some continuous function  $\psi$ . Let us now consider the quantity

$$\hat{\mathbf{D}}_n(y) := \frac{1}{n} \sum_{k=1}^n g_{Z_{k-1}}(f(y)) 1_{Z_k \geq f(y), y \geq Z_{k-1}}.$$

According to Krell [2016, proof of Proposition 3.2], we have that

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \sup_{y \in \mathcal{I}} \mathbb{E} \left( \left( \hat{\mathbf{D}}_n(y) - \mathbf{D}(y) \right)^2 \right) \lesssim \frac{1}{n}. \quad (10)$$

We can now consider the estimator

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

Then  $\hat{\lambda}_n$  will converge with nearly the same rate of convergence as  $\hat{\nu}_\lambda$ :

**Theorem 9.** *Under A1-A5, as soon as  $\ln(n)^{-1} \leq D_0/2$ , for any  $\lambda \in \mathcal{E}(\bar{c}, b)$ ,*

$$\mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \leq C_{\bar{c}, b} \ln^2(n) \left( \mathbb{E} \left( \|\hat{\nu}_\lambda - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 \right) + \frac{1}{n} \right).$$

As a consequence, by (8),

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b), \|\nu_\lambda\|_{B_{2, \infty}^\alpha} \leq C} \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \lesssim \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

To obtain a result related to the regularity of  $\lambda$  and not  $\nu_\lambda$ , we assume that  $\lambda$  belong to  $H^\alpha(M_1, \mathcal{I})$ .

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b) \cap H^\alpha(M_1, \mathcal{I})} \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \lesssim \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

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

We have proved that:

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b) \cap H^\alpha(\mathcal{I}, M_1)} \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \lesssim \ln(n)^2 n^{-2\alpha/(2\alpha+1)}.$$

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

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\bar{c}, b) \cap H^\alpha(\mathcal{I}, M_1)} \left( \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \right) \geq C \ln^2(n) n^{-2\alpha/(2\alpha+1)}.$$

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  $w(n)$  smaller 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-A5 are satisfied, then*

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\bar{\epsilon}, b) \cap \dot{H}^\alpha(\mathcal{I}, M_1)} \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \geq C n^{-2\alpha/(2\alpha+1)}.$$

where the infimum is taken among all estimators.

## 4 Simulations

**TCP protocol.** We consider the piecewise deterministic Markov process  $(X_t)$  with flow  $\phi(x, t) = x + ct$ ,  $f(x) = \kappa x$  and  $g_x(y) = 1/(\kappa c)$ . By (3), we have:

$$\mathbb{P}(Z_1 \geq y | Z_0 = x) = \exp \left( -\frac{1}{c} (\Lambda(y/\kappa) - \Lambda(x)) \right) 1_{y \geq \kappa x}$$

with  $\Lambda$  a primitive of  $\lambda$ . The function  $\Lambda$  is increasing and therefore invertible and

$$\mathbb{P}(\Lambda(Z_1/\kappa) \geq v | \Lambda(Z_0) = u) = \exp \left( -\frac{1}{c} (v - u) \right) 1_{v \geq u}.$$

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

$$\Lambda(Z_j/\kappa) = \Lambda(Z_{j-1}) + cE_j \tag{12}$$

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

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

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

This model satisfies A1, A2 and A4 if  $\delta > -1$ . A3 is fulfilled only if  $\delta > 0$  (indeed, if  $\delta \leq 0$ , the model does not comply with A3a).

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 (12),

$$(Z_j/\kappa - a)^3 + 3b(Z_j/\kappa - a) = (Z_{j-1} - a)^3 + 3b(Z_{j-1} - a) + 3cE_j$$

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

$$Z_j = \kappa \left( a + \frac{\sqrt[3]{Q + \sqrt{4b^3 + Q^2}} + \sqrt[3]{Q - \sqrt{4b^3 + Q^2}}}{2} \right)$$

where  $Q = 3cE_j + (Z_{j-1} - a)^3 + 3b(Z_{j-1} - a)$ . This model satisfies A1-A4.

**Bacterial growth.** We consider a PDMP with  $\phi(x, t) = xe^{ct}$  and  $f(x) = x/2$ . Then  $g_x(y) = \frac{1}{cy}$  and by (3),

$$\mathbb{P}(Z_1 \geq y | Z_0 = x) = \exp \left( -\frac{1}{c} \int_x^{2y} \frac{\lambda(s)}{s} ds \right) 1_{y \geq x/2}.$$

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

$$\mathbb{P}(Z_1 \geq y | Z_0 = x) = \exp\left(-\frac{2^\delta \lambda}{\delta c} \left(y^\delta - \frac{x^\delta}{2^\delta}\right)\right) 1_{y \geq x/2}.$$

Therefore

$$\mathbb{P}\left((2Z_1)^\delta \geq y | Z_0^\delta \geq x\right) = \exp\left(-\frac{\lambda}{\delta c}(y - x)\right) 1_{y \geq x}$$

and the law of the random variable  $(2Z_j)^\delta$  is an exponential translated by  $Z_{j-1}^\delta$  and of parameter  $\lambda/\delta c$ . Then

$$Z_j = \frac{1}{2} \sqrt[\delta]{\frac{\delta c}{\lambda} E_j + Z_{j-1}^\delta},$$

with  $E_j \sim \mathcal{E}(1)$  i.i.d. This model satisfies **A1**, **A2** and **A4** if  $\delta > 0$  and **Assumption A3** only if  $\delta > 1$  (point **a** is not verified if  $\delta \leq 1$ ).

**Computations** The estimator of  $\nu_\lambda$  is computed thanks to a projection on a trigonometric basis on the interval  $\mathcal{A} = [0, 6]$ . The constant involved in  $pen(m)$  is not easily tractable. To select it, one could use the dimension jump. Indeed, if the penalty  $pen(m) = cD_m/n$  is too small, we will always select the maximal dimension. If the penalty is large enough, we will select smaller models. We then let the constant  $c$  in the penalty vary and we note the dimension selected. For  $c$  smaller than a value  $c_{min}$ , we always select the largest models, and then it decreases rapidly. We set  $c = 2c_{min}$ . See [Arlot and Massart \[2009\]](#) for instance. However, this method involves quite a lot of computations. Instead, we always choose  $pen(m) = 2D_m/n$ , which is the constant used for independent random variables (here  $\phi_1 = 2/A = 1/3$ ). This choice seems confirmed by the simulations results: the oracle  $or$  remains close to 1.

In figures **2** and **3**, for each graph, we realize five simulations of the PDMP with  $n = 10^5$ . For each simulation, we draw the estimator  $\hat{\lambda}$ , the density  $\hat{\nu}_\lambda(f(x))$  and  $\hat{\mathbf{D}}_n(x)$ .

In the tables, we realize 50 simulations for each 4-uplet  $(n, c, \kappa, \lambda)$ . The estimation interval  $\mathcal{I}$  is chosen such that  $\mathbf{D}$  is greater than the threshold  $(\ln(n))^{-1}$  on  $\mathcal{I}$  for  $n = 10^{-5}$ . For each set of parameters, we compute the mean of the selected dimension  $\hat{D}_m$  and the mean of the  $L^2$  error on  $\mathcal{I}$  denoted "risk". We also want to prove that our estimator is truly adaptive. As  $\nu_\lambda$  is unknown, we can not check that  $\hat{m}$  is the better choice for estimating  $\nu_\lambda$ . Instead, let us consider the estimator

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

We have that  $\hat{\lambda}_n = \hat{\lambda}_{\hat{m}}$ . We compute

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

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 risk and the oracle

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

For the sake of completeness, we also give the mean of the computation time, denoted by  $T$ .

**Results** In Figures 2-3, we can see that  $\hat{\lambda}$  is quite close  $\lambda$ , at least when  $x$  is neither too small nor too large, that is when there is enough values to compute the estimator. For  $x$  too large, the estimator seems to disconnect before to be constant equal to 0. For  $x$  near 0, we can have a spike if  $\lambda(0) \neq 0$ . The estimators of  $\hat{\nu}$  and  $D_n$  are quite smooth, whereas  $\hat{\lambda}$  tends to oscillate. This is due to the division of two estimators. The estimation of  $\hat{\lambda}_n$  is not good for the bacterial growth when  $\lambda = \sqrt{x}$ , the estimator is biased. However, in that case, Assumption A3a is not satisfied and there can be a problem of convergence toward the stationary measure (or even of the existence of the stationary measure). On the contrary, for  $\lambda = x$  for the bacterial growth and  $\lambda = 1$  for the TCP, the convergence is good, even if Assumption A3a is not fulfilled (but we are just on the edge). In Tables 2-3, we can observe that the risk decreases when  $n$  increases and seems to tend toward 0. The oracle remains close to 1, our estimator is really adaptive.

## 5 Proofs

### 5.1 Proof of Corollary 3

We consider a function  $s$  such that  $\|s\|_\infty = 1$ ; we will obtain the expected result by dividing  $s$  by its  $L^\infty$ -norm. We set  $Z_0 = X_0 = x_0$ . According to Proposition 2a,

$$\left| \mathbb{E} \left( \frac{1}{n} \sum_{k=1}^n s(Z_k) - \int s(x) \nu_\lambda(x) dx \right) \right| \leq \frac{1}{n} \sum_{k=1}^n R\mathbb{V}(x_0) \gamma^k \leq \frac{R\mathbb{V}(x_0)}{n(1-\gamma)}$$

which proves a. Let us set  $\tilde{s}(Y) = s(Y) - \mathbb{E}(s(Y))$ . We have:

$$\mathbb{E} \left( \left( \frac{1}{n} \sum_{k=1}^n \tilde{s}(Z_k) \right)^2 \right) = \frac{1}{n^2} \sum_k \mathbb{E}(\tilde{s}^2(Z_k)) + \frac{2}{n^2} \sum_{k < k'} \mathbb{E}(\tilde{s}(Z_k) \tilde{s}(Z_{k'})). \quad (13)$$

Thanks to Proposition 2a, we can write:

$$\begin{aligned} \mathbb{E}(\tilde{s}^2(Z_k)) &\leq \mathbb{E}(s^2(Z_k)) \\ &\leq \int s^2(x) \nu_\lambda(x) dx + \mathbb{E}(s^2(Z_k)) - \int s^2(x) \nu_\lambda(x) dx \\ &\leq \int s^2(x) \nu_\lambda(x) dx + R\mathbb{V}(x_0) \gamma^k. \end{aligned}$$

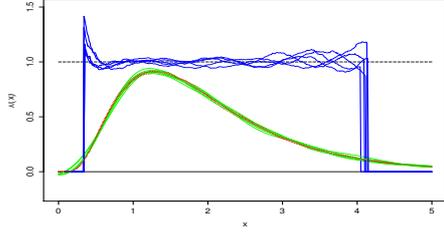
Therefore

$$\mathbb{E} \left( \frac{1}{n^2} \sum_{k=1}^n \tilde{s}^2(Z_k) \right) \leq \frac{1}{n} \int s^2(x) \nu_\lambda(x) dx + \frac{R\mathbb{V}(x_0)}{n^2(1-\gamma)}. \quad (14)$$

Let us bound the last term of (13). For any  $(k < k')$ , by Proposition 2a,

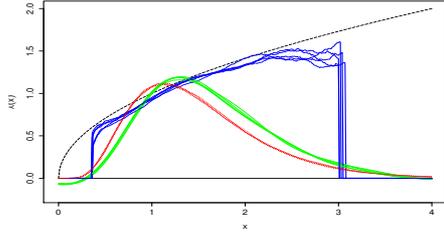
$$\begin{aligned} |\mathbb{E}(\tilde{s}(Z_{k'}) | Z_k)| &= \left| \mathbb{E}(s(Z_{k'}) | Z_k) - \int s(z) \nu_\lambda(z) dz - \mathbb{E}(s(Z_{k'})) + \int s(z) \nu_\lambda(z) dz \right| \\ &\leq R\gamma^{k'-k} \mathbb{V}(Z_k) + R\mathbb{V}(x_0) \gamma^{k'}. \end{aligned} \quad (15)$$

Figure 2: TCP protocol:  $\phi(x, t) = x + t$ ,  $f(x) = \kappa x$



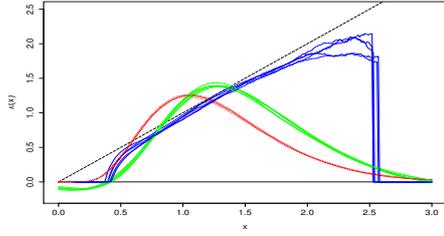
$$\kappa = 1/2, \lambda(x) = 1, \mathcal{I} = [0.2, 4]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	7.1	10	1.18	1.13	0.002
$10^3$	12.6	20.4	0.42	1.07	0.019
$10^4$	19.8	29.4	0.15	1.05	0.20
$10^5$	28.5	41.6	<b>0.12</b>	1.01	3.41



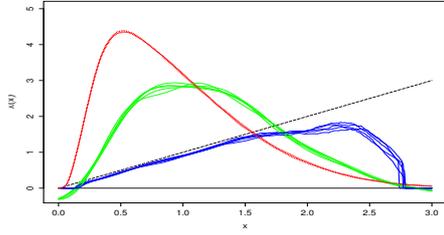
$$\kappa = 1/2, \lambda(x) = \sqrt{x}, \mathcal{I} = [0.2, 3]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	8.7	9.9	1.45	1.1	0.002
$10^3$	14.8	21.6	0.45	1.12	0.015
$10^4$	22	25.1	0.069	1.11	0.17
$10^5$	30	27	<b>0.056</b>	1.05	2.96



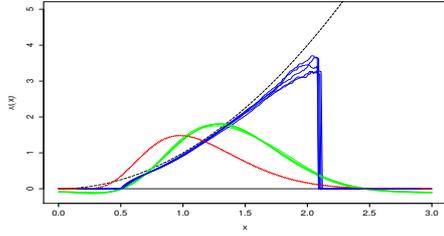
$$\kappa = 1/2, \lambda(x) = x, \mathcal{I} = [0.5, 2.5]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	10.3	10.2	1.51	1.04	0.0016
$10^3$	17	21.9	0.38	1.21	0.012
$10^4$	23.8	25.1	0.018	1.1	0.14
$10^5$	31.6	23	<b>0.024</b>	1.40	2.6



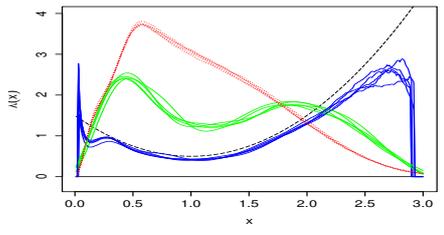
$$\kappa = 1/5, \lambda(x) = x, \mathcal{I} = [0.1, 2.5]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	11.0	9.1	1.66	1.14	0.0017
$10^3$	34.0	25.3	0.25	1.53	0.012
$10^4$	69.12	53.80	0.074	1.38	0.14
$10^5$	106.5	48.4	<b>0.065</b>	1.30	2.70



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

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	11.0	10.7	1.46	1.00	0.002
$10^3$	19.6	21.4	0.096	1.58	0.012
$10^4$	26.6	22.5	0.025	1.90	0.16
$10^5$	35.0	19.8	<b>0.016</b>	1.27	3.31



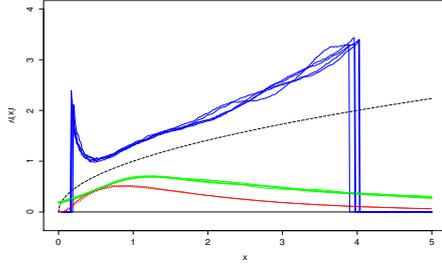
$$\kappa = 1/5, \lambda(x) = (x - 1)^2 + 1/2, \mathcal{I} = [0.1, 2.8]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	11.0	10.8	1.82	1.00	0.002
$10^3$	33.0	31.2	0.23	1.03	0.016
$10^4$	84.8	47	0.093	1.69	0.19
$10^5$	210	44.5	<b>0.085</b>	1.55	3.71

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

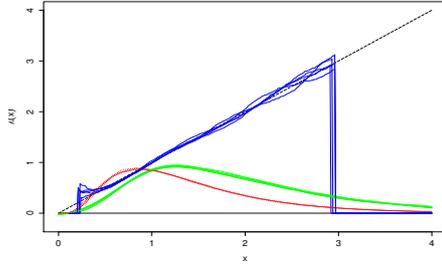
Figure 3: Bacterial growth

$$\phi(x, t) = x \exp(ct), \quad g(x) = 1/(cx), \quad \lambda(x) = x^\delta$$



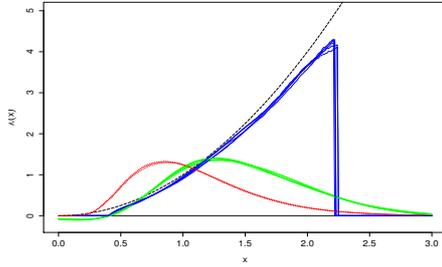
$$\lambda(x) = \sqrt{x}, c = 1, \mathcal{I} = [0.5, 3]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	4.0	3.96	1.11	1.10	0.002
$10^3$	9.1	3.0	0.43	1.91	0.016
$10^4$	16.4	3.0	0.41	1.79	0.19
$10^5$	25.8	3.0	<b>0.40</b>	1.78	3.80



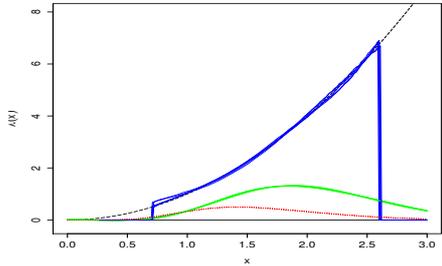
$$\lambda(x) = x, c = 1, \mathcal{I} = [0.5, 2.5]$$

$n$	$D_{hatm}$	$D_{m_{opt}}$	risk	or	T
$10^2$	6.5	9.5	0.96	1.74	0.003
$10^3$	12.4	16.2	0.039	2.67	0.014
$10^4$	20.1	22.4	0.0036	2.18	0.17
$10^5$	28.3	33.3	<b>0.00051</b>	1.82	3.51



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

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	10.5	10.1	1.32	1.09	0.002
$10^3$	16.7	19.3	0.025	1.89	0.012
$10^4$	24.0	20.8	0.017	1.97	0.16
$10^5$	31.4	21.9	<b>0.0097</b>	1.27	3.3



$$\lambda(x) = x^2, c = 3, \mathcal{I} = [1, 2.5]$$

$n$	$D_{\hat{m}}$	$D_{m_{opt}}$	risk	or	T
$10^2$	6.2	9.0	14.7	1.00	0.002
$10^3$	9.9	12.2	4.34	1.01	0.013
$10^4$	13.9	21.0	0.13	1.29	0.16
$10^5$	18.1	20.6	<b>0.11</b>	1.08	3.31

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

Then by (15),

$$\begin{aligned} E_k &:= \sum_{k'=k+1}^n |\mathbb{E}(\tilde{s}(Z_k)\tilde{s}(Z_{k'}))| = \sum_{k'=k+1}^n |\mathbb{E}(\tilde{s}(Z_k)|\mathbb{E}(\tilde{s}(Z_{k'})|Z_k))| \\ &\leq \sum_{k'=k+1}^n \mathbb{E}(|\tilde{s}(Z_k)|) (R\gamma^{k'-k}\mathbb{V}(Z_k) + R\mathbb{V}(x_0)\gamma^{k'}) \leq \frac{R}{1-\gamma} \mathbb{E}(|\tilde{s}(Z_k)| (\mathbb{V}(Z_k) + \gamma^k\mathbb{V}(x_0))). \end{aligned}$$

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

$$E_k \leq \frac{R}{1-\gamma} (\mathbb{E}(|s(Z_k)|\mathbb{V}(Z_k)) + \mathbb{E}(|s(Z_k)|) \mathbb{E}(\mathbb{V}(Z_k))) + \frac{R\mathbb{V}(x_0)}{1-\gamma} \gamma^k 2\mathbb{E}(|s(Z_k)|).$$

By Proposition 2a, for any function  $\psi \leq \mathbb{V}$ ,  $\mathbb{E}(\psi(Z_k)|Z_0 = x_0) \leq \int_0^\infty \psi(z)\nu_\lambda(z)dz + R\mathbb{V}(x_0)\gamma^k$ . Then

$$\begin{aligned} \sum_k E_k &\leq \sum_k \frac{R}{1-\gamma} \left( \int |s(z)|\mathbb{V}(z)\nu_\lambda(z)dz + R\mathbb{V}(x_0)\gamma^k \right) \\ &\quad + \sum_k \frac{R}{1-\gamma} \left( \int |s(z)|\nu_\lambda(z)dz + R\mathbb{V}(x_0)\gamma^k \right) \left( \int \mathbb{V}(z)\nu_\lambda(z)dz + R\mathbb{V}(x_0)\gamma^k \right) \\ &\quad + \sum_k \frac{R\mathbb{V}(x_0)}{1-\gamma} 2\gamma^k \left( \int |s(z)|\nu_\lambda(z)dz + R\mathbb{V}(x_0)\gamma^k \right) \\ &\leq n \frac{R}{1-\gamma} \left( \int |s(z)|\mathbb{V}(z)\nu_\lambda(z)dz + \int |s(z)|\nu_\lambda(z)dz \int \mathbb{V}(z)\nu_\lambda(z)dz \right) \\ &\quad + \frac{R\mathbb{V}(x_0)}{(1-\gamma)^2} \left( R + R \int \mathbb{V}(z)\nu_\lambda(z)dz + (R+2) \int |s(z)|\nu_\lambda(z)dz \right) + \frac{R^2(R+2)\mathbb{V}(x_0)}{(1-\gamma)(1-\gamma^2)}. \end{aligned}$$

By (13) and (14), we get:

$$\mathbb{E} \left( \left( \frac{1}{n} \sum_{k=1}^n \tilde{s}(Z_k) \right)^2 \right) \leq \frac{1}{n} \left( \int s^2(z)\nu_\lambda(z)dz + \int |s(z)|G_\lambda(z)\nu_\lambda(z)dz \right) + \frac{C}{n^2}$$

where  $C$  is uniformly bounded on  $\mathcal{F}(\bar{c}, b)$ .

## 5.2 Proof of Lemma 4

There exists  $(\bar{c}, b)$  such that  $\lambda \in \mathcal{E}(\bar{c}, b)$ . As  $\mathbf{D}(y) = \frac{\nu_\lambda(f(y))}{\lambda(y)}$  and  $\lambda(y)$  is strictly positive on  $\mathcal{I}$  by A4,

$$\left\{ \exists \eta, \sup_{y \in f(\mathcal{I})} \nu_\lambda(y) \geq \eta \right\} \Rightarrow \left\{ \exists D_0 > 0, \sup_{y \in \mathcal{I}} \mathbf{D}(y) \geq D_0 \right\}.$$

Therefore, we will just prove the first inequality.

**Bound of  $\nu(y)$  on  $f(\mathcal{I})$ .** By equation (4),

$$\mathcal{P}(x, y) = \lambda(f^{-1}(y))g_x(y) \exp \left( - \int_{f(x)}^y \lambda(f^{-1}(u))g_x(u)du \right) 1_{y \geq f(x)}. \quad (16)$$

By A4,  $\forall y \in f(\mathcal{I})$ ,  $\lambda(f^{-1}(y)) \geq \varepsilon$ . Moreover, by A3b, for any  $y \in f(\mathcal{I})$ :

$$\int_{f(x)}^y \lambda(f^{-1}(u))g_x(u)du \leq \int_0^{f(i_2)} \mathbf{M}(u)\lambda(f^{-1}(u))du \leq L.$$

We replace in (16) and obtain that  $\mathcal{P}(x, y) \geq \varepsilon \mathbf{m}(y)e^{-L}1_{y \geq f(x)}$ . Then, for any  $y \in f(\mathcal{I})$ , as  $\nu_\lambda$  is the stationary density:

$$\begin{aligned} \nu_\lambda(y) &= \int \nu_\lambda(x)\mathcal{P}(x, y)dx \geq \int_0^{f^{-1}(y)} \varepsilon \mathbf{m}(y)e^{-L}\nu_\lambda(x)dx \\ &\geq \varepsilon e^{-L}\mathbf{m}(y) \int_0^{i_1} \nu_\lambda(u)du \geq \varepsilon e^{-L}\mathbf{m}(y)\nu_\lambda([0, i_1]). \end{aligned}$$

It remains to bound  $\nu_\lambda([0, i_1])$  from below.

**Bound of  $\nu_\lambda([0, i_1])$ .** As  $\nu_\lambda$  is the stationary density,  $\nu_\lambda([z, \infty]) = \mathbb{P}_{\nu_\lambda}(Z_1 \geq z)$ . Therefore, by Markov inequality, as  $\mathbb{V}$  is an increasing function,

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \nu_\lambda([z, \infty]) = \sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \mathbb{P}_{\nu_\lambda}(\mathbb{V}(Z_1) \geq \mathbb{V}(z)) \leq \mathbb{V}^{-1}(z) \sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \mathbb{E}_{\nu_\lambda}(\mathbb{V}(Z_1)).$$

By Proposition 2b, as  $\mathcal{E}(\bar{c}, b) \subseteq \mathcal{F}(c, b)$ ,

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \mathbb{E}_{\nu_\lambda}(\mathbb{V}(Z_1)) \leq C.$$

Therefore, there exists  $y_0 > 0$ ,  $\sup_{\lambda \in \mathcal{E}(\bar{c}, b)} \nu_\lambda([y_0, \infty]) < 1$  and consequently,  $\inf_{\lambda \in \mathcal{E}(\bar{c}, b)} \nu_\lambda([0, y_0]) > 0$ . Now as  $f(x) \leq \kappa x$ , 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

$$\inf_{\lambda \in \mathcal{E}(\bar{c}, b)} \nu_\lambda([0, z_{k_n}]) > 0. \quad (17)$$

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

$$\nu_\lambda([0, z]) = \int_0^z \nu_\lambda(y)dy = \int_0^z \int_0^\infty \mathcal{P}(x, y)\nu_\lambda(x)dx dy = \int_0^\infty \nu_\lambda(x)\mathbb{P}(Z_1 \leq z|Z_0 \in dx)$$

and by (3),

$$\nu_\lambda([0, z]) = \int_0^{f^{-1}(z)} \nu_\lambda(x) \left( 1 - \exp\left(-\int_{f(x)}^z \lambda(f^{-1}(u))g_x(u)du\right) \right) dx.$$

As  $\lambda(f^{-1}(u))$  is bounded by below on  $(f(i_1), \infty)$  and  $g_x(u) \geq \mathbf{m}(u)$ , there exists a constant  $\eta$  such that

$$\inf_{\lambda \in \mathcal{E}(\bar{c}, b)} \inf_{u \in [i_1, z_{k_n}]} \lambda(f^{-1}(u))g_x(u) \geq \eta.$$

As  $f(u) \leq \kappa u$  and  $\kappa < 1$ ,  $f^{-1}(z_j) \geq z_j/\kappa \geq z_j/\sqrt{\kappa} = z_{j+1}$ . Therefore

$$\begin{aligned} \nu_\lambda([0, z_j]) &\geq \int_{i_1}^{f^{-1}(z_j)} \nu_\lambda(x) \left( 1 - \exp\left(-\int_{f(x)}^{z_j} \lambda(f^{-1}(u))g_x(u)du\right) \right) dx \\ &\geq \int_{i_1}^{z_{j+1}} \nu_\lambda(x) (1 - \exp(-\eta(z_j - f(x)))) dx \\ &\geq \int_{i_1}^{z_{j+1}} \nu_\lambda(x) (1 - \exp(-\eta(z_j - \kappa z_{j+1}))) dx \\ &\geq (1 - \exp(-\eta(z_j(1 - \sqrt{\kappa})))) \nu_\lambda([i_1, z_{j+1}]). \end{aligned}$$

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

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

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

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

Then by (17)

$$\inf_{\lambda \in \mathcal{E}(\bar{\epsilon}, b)} \nu_\lambda([0, i_1]) > 0$$

which concludes the proof.

### 5.3 Proof of Corollary 6

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

$$E^i = \{Z_1 \in J_1^i, \dots, Z_k \in J_k^i\}, \quad F^{i,j} = \{Z_{t+k} \in I_0^{i,j}, \dots, Z_{t+k+n} \in I_n^{i,j}\}$$

with  $J_i$  and  $I_i$  are subsets of  $\mathbb{R}^+$  and  $1 \leq n < \infty$ . Then

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

For a vector  $(a_1, \dots, a_k) \in (\mathbb{R}^+)^k$ , let us set  $\mathcal{P}(a_1, \dots, a_k) = \mathcal{P}(a_1, a_2) \dots \mathcal{P}(a_{k-1}, a_k)$  where the transition probability  $\mathcal{P}(x, y) = \mathcal{P}_\lambda(x, y)$  is defined in equation (4). Then, as  $(Z_k)_{k \in \mathbb{N}}$  is a Markov chain,

$$\begin{aligned} \mathbb{P}(E^i \cap F^{i,j}) - \mathbb{P}(E^i)\mathbb{P}(F^{i,j}) &= \int_{J_1^i \times \dots \times J_k^i} \mathcal{P}(x_0, x_1, \dots, x_k) \\ &\quad \times \int_0^\infty (\mathcal{P}^t(x_k, y_0) - \mathcal{P}^{t+k}(x_0, y_0)) 1_{y_0 \in I_0^{i,j}} \\ &\quad \times \int_{I_1^{i,j} \times \dots \times I_n^{i,j}} \mathcal{P}(y_0, \dots, y_n) dy_0 \dots dy_n dx_1 \dots dx_k. \end{aligned}$$

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

$$\begin{aligned} &\sum_j \mathbb{P}(E^i \cap F^{i,j}) - \mathbb{P}(E^i)\mathbb{P}(F^{i,j}) \\ &= \int_{J_1^i \times \dots \times J_k^i} \mathcal{P}(x_0, x_1, \dots, x_k) \int_0^\infty (\mathcal{P}^t \psi(x_k) - \mathcal{P}^{t+k} \psi(x_0)) dx_1 \dots dx_k \end{aligned}$$

where  $\psi(x) := \sum_j 1_{x \in I_0^{i,j}} \int_{I_1^{i,j} \times \dots \times I_n^{i,j}} \mathcal{P}(x, y_1, \dots, y_n) dy_1 \dots dy_n$ . We can remark that  $\psi(x) = \sum_j 1_{x \in I_0^{i,j}} \mathbb{P}_x \left( Z_1 \in I_1^{i,j}, \dots, Z_n \in I_n^{i,j} \right)$  and by the law of total probability,  $\psi(x) \leq 1$ . We can apply Proposition 2a to the function  $\psi$ :

$$\begin{aligned} |\mathcal{P}^t \psi(x_k) - \mathcal{P}^{t+k} \psi(x_0)| &= \left| \mathcal{P}^t \psi(x_k) - \int_0^\infty \psi(z) \nu_\lambda(z) dz + \int_0^\infty \psi(z) \nu_\lambda(z) dz - \mathcal{P}^{t+k} \psi(x_0) \right| \\ &\leq R\gamma^t \mathbb{V}(x_k) + R\gamma^{t+k} \mathbb{V}(x_0). \end{aligned}$$

Then

$$\begin{aligned} |D_G| &\leq R\gamma^t \sum_i \int_{J_1^i \times \dots \times J_k^i} \mathcal{P}(x_0, \dots, x_k) (\mathbb{V}(x_k) + \mathbb{V}(x_0)) dx_1 \dots dx_k \\ &\leq R\gamma^t \sum_i \mathbb{E}_{x_0} \left( (\mathbb{V}(Z_k) + \mathbb{V}(x_0)) 1_{Z_1 \in J_1^i, \dots, Z_k \in J_k^i} \right) \leq R\gamma^t (\mathbb{E}_{x_0} (\mathbb{V}(Z_k)) + \mathbb{V}(x_0)). \end{aligned}$$

By Proposition 2,

$$\mathbb{E}_{x_0} (\mathbb{V}(Z_k)) \leq \int \mathbb{V}(y) \nu_\lambda(y) dy + R\gamma^k \mathbb{V}(x_0)$$

and  $\int \mathbb{V}(y) \nu_\lambda(y) dy$  is uniformly bounded on  $\mathcal{F}(c, b)$  by Proposition 2b. Therefore

$$\sup_{\lambda \in \mathcal{F}(c, b)} \beta_Z(t) = \sup_k \sup_{G \in \mathcal{O}_0^k \times \mathcal{O}_{t+k}^\infty} |D_G| \leq R\gamma^t \left( \int \mathbb{V}(y) \nu_\lambda(y) dy + (1 + R)\mathbb{V}(x_0) \right).$$

As  $\gamma < 1$ ,

$$\sup_{\lambda \in \mathcal{F}(c, b)} \beta_Z(t) \leq ce^{-\beta t}$$

with  $\beta = -\ln(\gamma)$ ,  $c = R \left( \sup_{\lambda \in \mathcal{F}(c, b)} \int \mathbb{V}(y) \nu_\lambda(y) dy + (1 + R)\mathbb{V}(x_0) \right)$ .

## 5.4 Proof of Proposition 7

We have the following bias-variance decomposition:

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

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

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

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  $(\varphi_l)$  form an orthonormal basis of  $S_m$ , we have

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

By Corollary 3,

$$|\mathbb{E}(\hat{a}_l) - a_l| \leq \|\varphi_l\|_\infty \frac{R\mathbb{V}(x_0)}{n(1-\gamma)}.$$

Therefore, thanks to A5b,  $\|\varphi_l\|_\infty^2 \leq \psi_1 D_m$  and, as  $D_m^2 \leq n$ , we get:

$$\sup_{\lambda \in \mathcal{F}(c,b)} \|\mathbb{E}(\hat{\nu}_m) - \nu_m\|_{L^2(\mathcal{A})}^2 \leq \sum_{l=1}^{D_m} \frac{\|\varphi_l\|_\infty^2}{n^2} \frac{R^2 \mathbb{V}^2(x_0)}{(1-\gamma)^2} \leq \psi_1 \frac{D_m^2}{n^2} \frac{R^2 \mathbb{V}^2(x_0)}{(1-\gamma)^2} \leq \psi_1 \frac{1}{n} \frac{R^2 \mathbb{V}^2(x_0)}{(1-\gamma)^2}.$$

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  $\hat{a}_\lambda$ :

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

By Corollary 3,

$$\begin{aligned}\text{Var}(\hat{a}_l) &= \text{Var}\left(\frac{1}{n} \sum_{k=1}^n \varphi_l(Z_k)\right) \\ &\leq \frac{1}{n} \int_{\mathcal{A}} \varphi_l^2(x) \nu_\lambda(x) dx + \frac{\|\varphi_l\|_\infty}{n} \int_{\mathcal{A}} |\varphi_l(x)| G_\lambda(x) \nu_\lambda(x) dx + \frac{c \|\varphi_l\|_\infty^2}{n^2}.\end{aligned}\quad (18)$$

By Assumption A5b and c,  $\sum_{l=1}^{D_m} \int_{\mathcal{A}} \varphi_l^2(x) \nu_\lambda(x) dx \leq \psi_1 D_m$ ,  $\sum_{l=1}^{D_m} \|\varphi_l\|_\infty \int_{\mathcal{A}} |\varphi_l(x)| \nu_\lambda(x) dx \leq \psi_2 D_m$  and  $\sum_{l=1}^{D_m} \|\varphi_l\|_\infty^2 \leq \psi_1 D_m^2 \leq \psi_1 n$ . Therefore:

$$\sup_{\lambda \in \mathcal{F}(c,b)} \int_{\mathcal{A}} \text{Var}(\hat{\nu}_m(x)) dx = \sup_{\lambda \in \mathcal{F}(c,b)} \sum_l \text{Var}(\hat{a}_l) \leq C_\lambda \frac{D_m}{n} + \frac{c}{n}\quad (19)$$

where  $C_\lambda = \psi_1 + \psi_2 \int_{\mathcal{A}} G_\lambda(z) \nu_\lambda(z) dz$ .

## 5.5 Proof of Theorem 8

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  $m \in \mathcal{M}_n$ , 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  $\gamma_n(s) = \|s\|_{L^2(\mathcal{A})}^2 - 2n^{-1} \sum_{k=1}^n s(Z_k)$ . Then

$$\|\hat{\nu}_{\hat{m}}\|_{L^2(\mathcal{A})}^2 \leq \|\nu_m\|_{L^2(\mathcal{A})}^2 + \text{pen}(m) - \text{pen}(\hat{m}) + \frac{2}{n} \sum_{k=1}^n \hat{\nu}_{\hat{m}}(Z_k) - \nu_m(Z_k).\quad (20)$$

We have that, for any function  $s \in L^2(\mathcal{A})$ :

$$\|s\|_{L^2(\mathcal{A})}^2 = \|s - \nu_\lambda\|_{L^2(\mathcal{A})}^2 - \|\nu_\lambda\|_{L^2(\mathcal{A})}^2 + 2 \int_{\mathcal{A}} s(x)\nu_\lambda(x)dx.$$

We apply this equality to  $\hat{\nu}_{\hat{m}}$  and  $\nu_m$ . Equation (20) becomes:

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

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

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

where  $\mathcal{B}_{m, m'} = \{s \in S_m + S_{m'}, \|s\|_{L^2(\mathcal{A})} = 1\}$ . 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,

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

As  $\|\hat{\nu}_{\hat{m}} - \nu_m\|_{L^2(\mathcal{A})}^2 \leq 2 \|\hat{\nu}_{\hat{m}} - \nu_\lambda\|_{L^2(\mathcal{A})}^2 + 2 \|\nu_m - \nu_\lambda\|_{L^2(\mathcal{A})}^2$ , we get:

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

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

$$I_n(s) := \frac{1}{n} \sum_{k=1}^n s(Z_k) - \mathbb{E}(s(Z_k)), \quad J_n(s) := \frac{1}{n} \sum_{k=1}^n \left( \mathbb{E}(s(Z_k)) - \int_{\mathcal{A}} s(x)\nu_\lambda(x)dx \right) \quad (21)$$

and  $p(m, m') := (\text{pen}(m) + \text{pen}(m'))/4$ . We can remark that  $\mathbb{E}(I_n J_n) = 0$ . Then:

$$\mathbb{E} \left( \|\hat{\nu}_{\hat{m}} - \nu_\lambda\|_{L^2(\mathcal{A})}^2 \right) \leq 3 \|\nu_m - \nu_\lambda\|_{L^2(\mathcal{A})}^2 + 4\text{pen}(m) + 8\mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) + J_n^2(s) \right) - 8p(m, \hat{m}). \quad (22)$$

By Assumption A5b,  $s \in \mathcal{B}_{m, \hat{m}}$  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 Corollary 3,

$$\sup_{\lambda \in \mathcal{F}(c, b)} \sup_{s \in \mathcal{B}_{m, \hat{m}}} J_n^2(s) \leq \sup_{s \in \mathcal{B}_{m, \hat{m}}} \frac{R^2 \mathbb{V}^2(x_0) \|s\|_\infty^2}{n^2(1-\gamma)^2} \leq \frac{4\psi_1^2 R^2 \mathbb{V}^2(x_0) n}{n^2(1-\gamma)^2} \leq \frac{C}{n}. \quad (23)$$

It remains to bound  $\mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right)_+$ . 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:

$$\mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right)_+ \leq \sum_{m, m' \in \mathcal{M}_n} \mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, m'}} I_n^2(s) - p(m, m') \right)_+. \quad (24)$$

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). *The random variables  $Z_1, \dots, Z_n$  are exponentially  $\beta$ -mixing. Let us set  $b_0 \geq -1/\ln(\gamma)$ . We define  $q_n := 2b_0 \ln(n)$ ,  $p_n = n/(2q_n)$ . We have that  $\beta(q_n) \leq c\gamma^{2b_0 \ln(n)} \lesssim n^{-2}$ . Let us consider*

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

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

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

then we have:

$$\mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, m'}} |I_n^2(s) - 6H^2| \right)_+ \leq C \frac{V}{n} \exp \left( -k_1 \frac{H^2}{12V} \right) + C' \frac{M_2^2}{p_n^2} \exp \left( -k_2 \frac{\sqrt{p_n} H}{\sqrt{q_n} M_2} \right) + 2 \frac{M_2^2}{n^2}$$

where  $C, C', k_1$  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 A5b,

$$\sup_{\lambda \in \mathcal{F}(c, b)} \sup_{s \in \mathcal{B}} \|s\|_\infty \leq \sup_{s \in \mathcal{B}} \psi_1^{1/2} D^{1/2} \|s\|_{L^2(\mathcal{A})} = \psi_1^{1/2} D^{1/2} := M_2.$$

By Corollary 3,

$$\text{Var} \left( \frac{1}{q_n} \sum_{k=1}^{q_n} s(Z_k) \right) \leq \frac{1}{q_n} \int s^2(z) \nu_\lambda(z) dz + \frac{\|s\|_\infty}{q_n} \int |s(z)| \nu_\lambda(z) G_\lambda(z) dz + \frac{c_\lambda \|s\|_\infty^2}{q_n^2}.$$

By Cauchy-Schwarz,

$$\|s^2 \nu_\lambda\|_{L^1(\mathcal{A})} \leq \|s\|_{L^2(\mathcal{A})} \|s \nu_\lambda\|_{L^2(\mathcal{A})} \leq \|s\|_{L^2(\mathcal{A})} \|s\|_\infty \|\nu_\lambda\|_{L^2(\mathcal{A})}$$

and

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

By Assumption A5b,  $\|s\|_\infty \leq \psi_1^{1/2} D^{1/2}$  and then

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

It remains to find  $H$  such that  $\mathbb{E}(\sup_{s \in \mathcal{B}} |I_n(s)|) \leq H/\sqrt{n}$ . Let us introduce  $(\varphi_l)_{1 \leq l \leq D}$  an orthonormal basis of  $S_m + S_{m'} = S_{\max(m, m')}$  satisfying Assumption A5. Then we can write  $s = \sum_l b_l \varphi_l$ . As the function  $s \rightarrow I_n(s)$  is linear:

$$\sup_{s \in \mathcal{B}} I_n^2(s) = \sup_{\sum b_l^2 = 1} \left( \sum_{l=1}^D b_l I_n(\varphi_l) \right)^2 \leq \sup_{\sum b_l^2 = 1} \left( \sum_{l=1}^D b_l^2 \right) \left( \sum_{l=1}^D I_n^2(\varphi_l) \right) = \sum_{l=1}^D I_n^2(\varphi_l). \quad (25)$$

We can remark that  $I_n(\varphi_l) = \hat{a}_l - \mathbb{E}(\hat{a}_l)$  (see equation (21)) and by consequence,  $\mathbb{E}(I_n^2(\varphi_l)) = \text{Var}(\hat{a}_l)$ . By (19):

$$\sup_{\lambda \in \mathcal{F}(c, b)} \mathbb{E} \left( \sup_{s \in \mathcal{B}} I_n^2(s) \right) \leq \sum_{l=1}^D \mathbb{E}(I_n^2(\varphi_l)) \leq \frac{C_\lambda D}{n} + \frac{c_\lambda}{n} := \frac{H^2}{n}$$

where  $C_\lambda = \psi_1 + \psi_2 \int_{\mathcal{A}} G_\lambda(z) \nu_\lambda(z) dz$ . We can now apply Lemma 11 with

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

As  $2/(x+y) \geq \min(1/x, 1/y)$ , for  $p(m, m') \geq 6C_\lambda D/n + 6c_\lambda/n$ ,

$$\begin{aligned} \mathbb{E} \left( \sup_{s \in \mathcal{B}} I_n^2(s) - p(m, m') \right)_+ &\leq C \frac{V}{n} \exp \left( -k_1 \frac{H^2}{12V} \right) + C' \frac{M_2^2}{p_n^2} \exp \left( -k_2 \frac{\sqrt{p_n} H}{\sqrt{q_n M_2}} \right) + 2 \frac{M_2^2}{n^2} \\ &\leq C \left( \frac{c_1 D^{1/2}}{n} + \frac{cD}{nq_n} \right) \left( \exp \left( -\frac{C_\lambda D}{12} \frac{1}{c_1 D^{1/2}} \right) + \exp \left( -\frac{C_\lambda D}{12} \frac{q_n}{cD} \right) \right) \\ &\quad + \frac{C\psi_1 D}{p_n^2} \exp \left( -k_2 p_n^{1/2} \frac{C_\lambda^{1/2} D^{1/2}}{q_n^{1/2}} \frac{1}{\psi_1^{1/2} D^{1/2}} \right) + 2 \frac{\psi_1 D}{n^2} \\ &\lesssim \frac{D}{n} \exp(-c_0 D^{1/2}) + \frac{D}{n} \exp(-c_2 q_n) + \frac{D \ln^2(n)}{n^2} \exp \left( -c_3 \frac{n^{1/2}}{\ln(n)} \right) + \frac{D}{n^2} \end{aligned}$$

where the constants are uniform on the set  $\{\lambda \in \mathcal{F}(c, b), \|\nu_\lambda\|_{L^2(\mathcal{A})} \leq c\}$ . The second term can be made smaller than  $n^{-2}$  for  $q_n (= 2b_0 \ln(n))$  large enough. The third is also smaller to  $n^{-2}$  thanks to the exponential term. Then

$$\sup_{\lambda \in \mathcal{F}(c, b), \|\nu_\lambda\|_{L^2} \leq C} \mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, m'}} I_n^2(s) - p(m, m') \right) \leq C_1 \frac{D_{m, m'}}{n} \exp(-c_0 D_{m, m'}^{1/2}) + C_2 \frac{D_{m, m'}}{n^2}.$$

As  $\sum_{k=1}^{\infty} k e^{-ck^{1/2}} < \infty$  and  $\sum_{m' \in \mathcal{M}_n} D_{m, m'} \leq \max_{m' \in \mathcal{M}_n} D_{m, m'}^2 \leq n$ , by (24),

$$\sup_{\lambda \in \mathcal{F}(c, b), \|\nu_\lambda\|_{L^2(\mathcal{A})} \leq c} \mathbb{E} \left( \sup_{s \in \mathcal{B}_{m, \hat{m}}} I_n^2(s) - p(m, \hat{m}) \right) \lesssim \frac{1}{n}. \quad (26)$$

Collecting (22), (23) and (26), for any  $m \in \mathcal{M}_n$ :

$$\sup_{\lambda \in \mathcal{F}(c, b), \|\nu_\lambda\|_{L^2(\mathcal{A})} \leq C} \mathbb{E} \left( \|\hat{\nu}_{\hat{m}} - \nu_\lambda\|_{L^2}^2 \right) \leq 3 \|\nu_m - \nu_\lambda\|_{L^2}^2 + 4pen(m) + \frac{c}{n}$$

which concludes the proof.

## 5.6 Proof of Theorem 9

For  $n$  big enough,  $1/\ln(n)$  is smaller than  $D_0/2$  and then by Markov inequality and (10),

$$\mathbb{P}(\hat{\mathbf{D}}_n(y) \leq 1/\ln(n)) \leq \mathbb{P}(|\hat{\mathbf{D}}_n(y) - \mathbf{D}(y)| \leq D_0/2) \leq \frac{4}{D_0^2} \mathbb{E} \left( (\hat{\mathbf{D}}_n(y) - \mathbf{D}(y))^2 \right) \leq \frac{4c}{nD_0^2}. \quad (27)$$

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

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

We can write:

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

As  $\mathbf{D} \geq D_0$  by Lemma 4:

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

By (10) and (27),

$$\begin{aligned} \mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) &\leq 3 \ln^2(n) \mathbb{E} \left( \left\| \hat{\nu}_\lambda \circ f - \nu_\lambda \circ f \right\|_{L^2(\mathcal{I})}^2 \right) + 3cD_0^{-2} \frac{\ln^2(n) \int_{\mathcal{I}} \nu_\lambda^2(f(y)) dy}{n} \\ &\quad + 12cD_0^{-2} \frac{\int_{\mathcal{I}} \lambda^2(y) dy}{n}. \end{aligned}$$

By A2 and Proposition 1b,  $\nu_\lambda(f(y)) \leq \mathbb{E}_{\nu_\lambda}(g_{Z_0}(f(y))) \lambda(y) \leq \mathbf{M}(f(y)) \lambda(y)$ . We obtain:

$$\mathbb{E} \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2(\mathcal{I})}^2 \right) \leq C \left( \ln(n)^2 \mathbb{E} \left( \left\| \hat{\nu}_\lambda \circ f - \nu_\lambda \circ f \right\|_{L^2(\mathcal{I})}^2 \right) + \frac{\ln^2(n) \|\lambda\|_{L^2(\mathcal{I})}^2}{n} \right) \quad (28)$$

where  $C$  is uniformly bounded on  $\mathcal{E}(\bar{c}, b)$ . As  $f$  is invertible,

$$\begin{aligned} \mathbb{E} \left( \left\| \hat{\nu}_\lambda \circ f - \nu_\lambda \circ f \right\|_{L^2(\mathcal{I})}^2 \right) &= \mathbb{E} \left( \left\| (\hat{\nu}_\lambda - \nu_\lambda) \sqrt{(f^{-1})'} \right\|_{L^2(f(\mathcal{I}))}^2 \right) \\ &\leq \sup_{y \in f(\mathcal{I})} (f^{-1})'(y) \mathbb{E} \left( \|\hat{\nu}_\lambda - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 \right) \end{aligned}$$

which, with (28), give the first bound.

By Theorem 8,

$$\sup_{\lambda \in \mathcal{E}(\bar{c}, b), \|\nu_\lambda\|_{L^2(f(\mathcal{I}))} \leq C} \mathbb{E} \left( \|\nu_\lambda - \hat{\nu}_m\|_{L^2(f(\mathcal{I}))}^2 \right) \leq \min_{m \in \mathcal{M}_n} \left( 3 \|\nu_m - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 + 24 \frac{C_\lambda D_m}{n} \right) + \frac{c}{n}.$$

If  $\lambda$  belongs to  $H^\alpha(M_1, \mathcal{I})$ , as  $g_x$  and  $f^{-1}$  also belongs to  $H^\alpha(M_1, \mathcal{I})$  by Assumption A2, then by (9),  $\nu_\lambda \in H^\alpha(\psi(M_1), f(\mathcal{I})) \subseteq \mathcal{B}_{2,\infty}^\alpha(\psi(M_1), f(\mathcal{I}))$  and in that case,

$$\|\nu_m - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 \leq \|\nu_\lambda\|_{B_{2,\infty}^\alpha(f(\mathcal{I}))}^2 D_m^{-2\alpha}.$$

The quantity  $3 \|\nu_m - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 + 24 \frac{C_\lambda D_m}{n}$  is minimum when  $D_m \propto n^{1/(2\alpha+1)}$  and therefore

$$\sup_{\lambda \in \mathcal{E}(\bar{\epsilon}, b, H)} \min_{m \in \mathcal{M}_n} \left( 3 \|\nu_m - \nu_\lambda\|_{L^2(f(\mathcal{I}))}^2 + 24 \frac{C_\lambda D_m}{n} \right) \leq C n^{-2\alpha/(2\alpha+1)}$$

which concludes the proof.

## 5.7 Proof of Theorem 10

To simplify the notations, we denote by  $\mathcal{E}(\bar{\epsilon}, b, H)$  the set  $\mathcal{E}(\bar{\epsilon}, b) \cap H^\alpha(\mathcal{I}, M_1)$  in this proof. We use the reduction scheme described in [Tsybakov \[2004, chapter 2\]](#). By Markov inequality,

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

Our aim is to show that

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

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

$$\|\lambda_i - \lambda_j\|_{L^2} \geq 2C' n^{-\alpha/(2\alpha+1)} \delta_{i,j}. \quad (29)$$

Then

$$\inf_{\hat{\lambda}_n} \sup_{\lambda \in \mathcal{E}(\bar{\epsilon}, b, H)} \mathbb{P}_\lambda \left( \left\| \hat{\lambda}_n - \lambda \right\|_{L^2} \geq C' n^{-\alpha/(2\alpha+1)} \right) \geq \inf_{\hat{\lambda}_n} \max_j \mathbb{P}_{\lambda_j} \left( \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2} \geq C' n^{-\alpha/(2\alpha+1)} \right).$$

We note  $\psi^*$  the predictor

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

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

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

By (29),  $\|\lambda_{\psi^*} - \lambda_j\|_{L^2} \geq 2C' n^{-\alpha/(2\alpha+1)} 1_{\psi^* \neq j}$ . Then setting  $A_n = C' n^{-\alpha/(2\alpha+1)}$ ,  $\left\{ \left\| \hat{\lambda}_n - \lambda_j \right\|_{L^2} \geq C' n^{-\alpha/(2\alpha+1)} \right\} \supseteq \{\psi^* \neq j\}$  and therefore:

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

We denote by  $\mathbf{P}_{\lambda_j}$  the law of  $(Z_0, Z_1, \dots, Z_n)$  under  $\lambda_j$ . The following lemma is exactly Theorem 2.5 of [Tsybakov \[2004\]](#).

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

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

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

b. For all  $i$ , the function  $\lambda_i$  belongs to the subspace  $\mathcal{E}(\bar{c}, b, H)$ .

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:

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

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

Then

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

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

$$\lambda_0(x) = \varepsilon 1_{[0, r[} + \frac{\mathbf{a}(f(x))^b}{\mathbf{m}(f(x))} 1_{x \geq r}.$$

As  $\lambda_0$  is constant on  $\mathcal{I}$ , this function belongs to the Hölder space  $H^\alpha(\mathcal{I})$  and  $\|\lambda_0\|_{H^\alpha(\mathcal{I})} = \varepsilon$ . The set  $\mathcal{E}(\bar{c}, b, H)$  is not empty if and only if  $\varepsilon \leq \min \left( M_1, \frac{L}{\int_0^r \mathbf{M}(u) du} \right)$ . In that case,  $\lambda_0 \in \mathcal{E}(\bar{c}, b, H)$ . Moreover, we suppose that the previous inequality is strict such that  $\mathcal{E}(\bar{c}, b, H)$  is not reduced to a point: there exists  $\delta > 0$  such that  $\|\lambda_0\|_{H^\alpha(\mathcal{I}, M_1)} \leq M_1 - \delta$  and  $\int_0^{f(r)} M(u) \lambda_0(u) du \leq L - \delta$ .

We consider a non-negative function  $K \in H^\alpha(\mathbb{R})$ , bounded, with support in  $[0, 1[$  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  $0 \leq k \leq p_n - 1$ ,  $x_k = i_1 + h_n k(i_2 - i_1)$ . We consider the functions

$$\varphi_k(x) := ah_n^\alpha K \left( \frac{x - x_k}{h_n} \right)$$

with  $a < 1$ . The functions  $\varphi_k$  have support in  $[x_k, x_{k+1}) \subset \mathcal{I}$ . 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:

$$\|\lambda_\epsilon - \lambda_\eta\|_{L^2}^2 = a^2 h_n^{2\alpha+1} \|K\|_{L^2}^2 \sum_{k=1}^{p_n} (\epsilon_k - \eta_k)^2. \quad (30)$$

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

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

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-serie of functions. According to Tsybakov [2004, 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

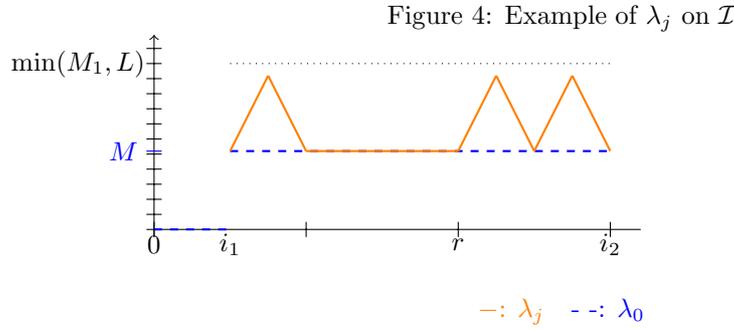
$$\forall 0 \leq j < k \leq P_n, \quad \rho(\epsilon_{(j)}, \epsilon_{(k)}) \geq p_n/8, \quad \text{and} \quad P_n \geq 2^{p_n/8}.$$

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

$$\ln(P_n) \geq \ln(2)n^{1/(2\alpha+1)}/8. \quad (31)$$

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 (30),

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

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

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

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

and by the change of variable  $z = t/h_n$ ,

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

The functions  $\varphi_k$  have disjoint supports. For any  $(x, y) \in \mathcal{I}$ , there exists  $(i, j)$  such that  $x \in [x_i, x_{i+1})$  and  $y \in [x_j, x_{j+1})$ . 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

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

and  $|\lambda_j|_{H^\alpha(\mathcal{I})} \leq 2a\|K\|_{H^\alpha}$ . Moreover,  $\|\lambda_j\|_{L^\infty(\mathcal{I})} \leq \|\lambda_0\|_{L^\infty(\mathcal{I})} + ah_n^\alpha \|K\|_{L^\infty} \leq \|\lambda_0\|_{L^\infty(\mathcal{I})} + 2a\|K\|_{L^\infty}$  and consequently  $\|\lambda_j\|_{H^\alpha(\mathcal{I})} \leq \|\lambda_0\|_{H^\alpha(\mathcal{I})} + 2a\|K\|_{H^\alpha}$ . Then  $\lambda_j \in H^\alpha(\mathcal{I}, M_1)$  for  $a$  sufficiently small. It remains to check that  $\lambda_j \in \mathcal{E}(\bar{c}, b)$ . For any  $0 \leq j \leq P_n$ :

- a. As  $K$  is nonnegative,  $\forall x \geq r$ ,  $\lambda_j(x) \geq a \frac{(f(x))^b}{\mathbf{m}(f(x))}$ .
- b. In the same way,  $\forall x, i_1 \leq x \leq r$ ,  $\lambda_j(x) \geq \varepsilon$ .
- c. Control of the integral:

$$\begin{aligned} \int_0^{f(r)} \mathbf{M}(u) \lambda_j(f^{-1}(u)) du &\leq \int_0^{f(r)} \mathbf{M}(u) \left( \lambda_0(f^{-1}(u)) + ah_n^\alpha K(f^{-1}(u)) \sum_j \varepsilon_j \right) du \\ &\leq L - \delta + ah_n^{\alpha-1} \int_0^{f(r)} \mathbf{M}(u) K(f^{-1}(u)) du \leq L \end{aligned}$$

for  $a$  small enough.

Therefore  $\lambda_j \in \mathcal{E}(\bar{c}, b, H)$  for  $a$  small enough.

**Step 3: Absolute continuity.** We denote by  $\mathcal{P}_j$  the transition densities induced by  $\mathbf{P}_{\lambda_j}$ . As  $(Z_n)_{n \in \mathbb{N}}$  is a Markov process,

$$\mathbf{P}_{\lambda_j}(x_0, \dots, x_n) = \mathcal{P}_j(x_0, x_1) \dots \mathcal{P}_j(x_{n-1}, x_n).$$

By (4), we can rewrite:  $\mathcal{P}_0(x, y) = A_{x,y} \exp(-\tilde{A}_{x,y})$  where

$$A_{x,y} := \lambda_0(f^{-1}(y)) g_x(y) \mathbb{1}_{\{y \geq f(x)\}}, \quad \tilde{A}_{x,y} := \int_{f(x)}^y \lambda_0(f^{-1}(u)) g_x(u) du$$

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

$$B_{x,y}^k := \varphi_k(f^{-1}(y)) g_x(y) \mathbb{1}_{\{y \geq f(x)\}}, \quad \tilde{B}_{x,y}^k := \int_{f(x)}^y \varphi_k(f^{-1}(u)) g_x(u) du.$$

The probability density  $\mathbf{P}_{\lambda_0}$  is null if one of the  $\mathcal{P}_0(x_i, x_{i+1})$  is null. This is the case if and only if  $A_{x_i, x_{i+1}} = 0$ , which implies either  $g_{x_i}(x_{i+1}) \mathbb{1}_{x_{i+1} \geq f(x_i)} = 0$  and therefore  $B_{x_i, x_{i+1}} = 0$ , either  $\lambda_0(f^{-1}(x_{i+1})) = 0$ , and therefore  $x_{i+1} \in [0, f(i_1)]$  and  $B_{x_i, x_{i+1}} = 0$ . 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:

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

We can write:

$$\chi^2(\mathbf{P}_{\lambda_j}, \mathbf{P}_{\lambda_0}) + 1 = \int_{(\mathbb{R}^+)^n} \left( \frac{\mathcal{P}_j(x_0, x_1) \dots \mathcal{P}_j(x_{n-1}, x_n)}{\mathcal{P}_0(x_0, x_1) \dots \mathcal{P}_0(x_{n-1}, x_n)} \right)^2 \mathcal{P}_0(x_0, x_1) \dots \mathcal{P}_0(x_{n-1}, x_n) dx_1 \dots dx_n.$$

Therefore, as  $\int_{\mathbb{R}^+} \mathcal{P}_0(x, y) dy = \int_{\mathbb{R}^+} \mathcal{P}_j(x, y) dy = 1$ ,

$$\begin{aligned} \chi^2(\mathbf{P}_{\lambda_j}, \mathbf{P}_{\lambda_0}) + 1 &= \int_{(\mathbb{R}^+)^{n-1}} \frac{(\mathcal{P}_j(x_0, x_1) \dots \mathcal{P}_j(x_{n-2}, x_{n-1}))^2}{\mathcal{P}_0(x_0, x_1) \dots \mathcal{P}_0(x_{n-2}, x_{n-1})} dx_1 \dots dx_{n-1} \int_{\mathbb{R}^+} \frac{(\mathcal{P}_j(x_{n-1}, x_n))^2}{\mathcal{P}_0(x_{n-1}, x_n)} dx_n \\ &= \int_{(\mathbb{R}^+)^{n-1}} \frac{(\mathcal{P}_j(x_0, x_1) \dots \mathcal{P}_j(x_{n-2}, x_{n-1}))^2}{\mathcal{P}_0(x_0, x_1) \dots \mathcal{P}_0(x_{n-2}, x_{n-1})} dx_1 \dots dx_{n-1} \\ &\quad \times \left( \int_{\mathbb{R}^+} \left( \frac{\mathcal{P}_j(x_{n-1}, x_n)}{\mathcal{P}_0(x_{n-1}, x_n)} - 1 \right)^2 \mathcal{P}_0(x_{n-1}, x_n) dx_n + 1 \right). \end{aligned} \quad (32)$$

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

$$\begin{aligned} DP &:= \int_{\mathbb{R}^+} \left( \frac{\mathcal{P}_j(x, y)}{\mathcal{P}_0(x, y)} - 1 \right)^2 \mathcal{P}_0(x, y) dy \\ &= \int_{\mathbb{R}^+} \left( \left( 1 + \frac{B_{x,y}}{A_{x,y}} \right) \exp(-\tilde{B}_{x,y}) - 1 \right)^2 A_{x,y} \exp(-\tilde{A}_{x,y}) dy. \end{aligned} \quad (33)$$

We can remark that if  $y \in f(\mathcal{I})^c$ ,  $\mathcal{P}_j(x, y) = \mathcal{P}_0(x, y)$ . Therefore

$$DP = \int_{f(\mathcal{I})} \left( \left( 1 + \frac{B_{x,y}}{A_{x,y}} \right) \exp(-\tilde{B}_{x,y}) - 1 \right)^2 A_{x,y} \exp(-\tilde{A}_{x,y}) dy.$$

By Assumption A2 and as  $\lambda_0 = \varepsilon$  on  $\mathcal{I}$ , we get that on  $f(\mathcal{I})$ ,

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

Moreover, on  $\mathbb{R}^+$ ,  $\exp(-\tilde{A}_{x,y}) \leq 1$ . We have that

$$B_{x,y}^k \leq \mathbf{M}(y) \|\varphi_k\|_\infty 1_{y \geq f(x)} 1_{f^{-1}(y) \in [x_k, x_{k+1}(x)}}$$

and therefore, as  $\|\varphi_k\|_\infty = ah_n^\alpha \|K\|_\infty$

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

By a change of variable,

$$\begin{aligned} \tilde{B}_{x,y}^k &= \int_x^{f^{-1}(y)} \varphi_k(z) g_x(f(z)) f'(z) dz \leq \sup_{z \in f(\mathcal{I})} (\mathbf{M}(z)) \sup_{z \in \mathcal{I}} f'(z) \|\varphi_k\|_{L^1} \\ &\leq ah_n^{\alpha+1} \sup_{z \in f(\mathcal{I})} (\mathbf{M}(z)) \sup_{z \in \mathcal{I}} f'(z) \end{aligned}$$

and

$$\tilde{B}_{x,y} \leq \sum \tilde{B}_{x,y}^k \leq C' a p_n h_n^{\alpha+1} \leq C' a h_n^\alpha.$$

Then

$$DP \leq C \int_{f(\mathcal{I})} a^2 h_n^{2\alpha} \leq C a^2 h_n^{2\alpha}.$$

Therefore, by (32) and (33), we get by recurrence

$$\begin{aligned} \chi^2(\mathbf{P}_{\lambda_0}, \mathbf{P}_{\lambda_j}) + 1 &= \int_{(\mathbb{R}^+)^{n-1}} \frac{(\mathcal{P}_j(x_0, x_1) \dots \mathcal{P}_j(x_{n-2}, x_{n-1}))^2}{\mathcal{P}_0(x_0, x_1) \dots \mathcal{P}_0(x_{n-2}, x_{n-1})} dx_1 \dots dx_{n-1} \times (\mathcal{O}(a^2 h_n^{2\alpha}) + 1) \\ &= \prod_{i=1}^n (\mathcal{O}(a^2 h_n^{2\alpha}) + 1) = 1 + a^2 n \mathcal{O}(h_n^{2\alpha}). \end{aligned}$$

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

$$\chi^2(\mathbf{P}_{\lambda_0}, \mathbf{P}_{\lambda_j}) \leq a^2 \mathcal{O}(n^{1/(2\alpha+1)})$$

and by (31),  $\ln(P_n) \geq \ln(2)n^{1/(2\alpha+1)}/8$  and therefore,

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

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

## A 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 cannot measure higher smoothness.

**Definition 14** (Modulus of smoothness). *We define the 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  $t^{\alpha-\mathbf{r}}\omega(f^{(\mathbf{r})}, t)_{\infty} = t^{\alpha}\omega_{\mathbf{r}}(f, t)_{\infty}$  and if  $f^{(\mathbf{r})}$  is Lipschitz, then  $\omega_{\mathbf{r}}(f, t)_{\infty} = \mathcal{O}(t^{\mathbf{r}})$ . 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:*

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

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

See [DeVore and Lorentz \[1993\]](#) and [Meyer \[1990\]](#) for more details. We use the Besov space to control the risk of the estimator of the stationary density  $\nu_\lambda$ .

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

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

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

As noted before,  $t^{\mathbf{r}-\alpha}\omega(f^{(\mathbf{r})}, t)_\infty = t^{-\alpha}\omega_{\mathbf{r}}(f, t)_\infty$ : the Hölder space  $H^\alpha(\mathcal{I})$  is included in  $B_{\infty, \infty}^\alpha$  which itself is included in  $B_{2, \infty}^\alpha(\mathcal{I})$ .

## B Proof of Lemma 11

The following lemma is very useful to replace weak dependent variables by independent variables. It is proved by [Viennet \[1997\]](#), proof of Proposition 5.1].

**Lemma 17** (Berbee's coupling lemma). *The random variables  $Z_{k\Delta}$  are exponentially  $\beta$ -mixing. Let  $q_n = \lfloor (r+1)\ln(n)/\beta \rfloor$  where  $\beta$  characterizes the  $\beta$ -mixing coefficient (see [Definition 5](#)). We have that  $\beta(q_n) \leq 1/n^{r+1}$ . We set  $p_n = n/(2q_n)$ . There exist random variables  $(Z_1^*, \dots, Z_n^*)$  such that:*

- $Z_i$  et  $Z_i^*$  have same law.
- $(Z_1^*, \dots, Z_{q_n}^*), (Z_{2q_n+1}^*, \dots, Z_{3q_n}^*), \dots$  are independent, as the random variables  $(Z_{q_n+1}^*, \dots, Z_{2q_n}^*), (Z_{3q_n+1}^*, \dots, Z_{4q_n}^*), \dots$
- $\mathbb{P}\left(Z_{kq_n+1}, \dots, Z_{(k+1)q_n} \neq (Z_{kq_n+1}^*, \dots, Z_{(k+1)q_n}^*)\right) \leq \beta(q_n) \leq n^{-(r+1)}$ .

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

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

This following inequality comes from Bernstein inequalities (see [Birgé and Massart \[1998\]](#), Corollary 2 p354]).

**Lemma 18** (Talagrand's inequality). *Let  $Y_1, \dots, Y_n$  be independent random variables and  $S$  a vectorial subspace of finite dimension  $D$  satisfying [Assumption 5](#). We denote by  $\mathcal{F}$  a countable family of  $S$ . Let us set*

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

with  $u \in L^2$ . If

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

then

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

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

*Proof.* We apply Theorem 1.1 of Klein and Rio [2005] to the functions  $s^i(x) = \frac{u(x) - \mathbb{E}(u(Y_i))}{2M_2}$ . 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 [1998]. 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 [2002, p222-223].  $\square$

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  $Z_k^*$  and  $Z_k$  have same law:

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

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

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

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

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

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

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

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

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

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

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.

## 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

- S. Arlot and P. Massart. Data-driven calibration of penalties for least-squares regression. *Journal of Machine Learning Research*, 10:245–279, 2009.
- 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. doi: 10.1214/16-EJS1207.
- 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. doi: 10.1051/proc/201444017.
- 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.
- 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.
- 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).
- 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.
- 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. doi: 10.1007/978-1-4899-4483-2.
- 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.
- 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. doi: 10.3150/14-BEJ623.
- 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.
- T. Fujii. Nonparametric estimation for a class of piecewise-deterministic Markov processes. *J. Appl. Probab.*, 50(4):931–942, 2013. ISSN 0021-9002. doi: 10.1239/jap/1389370091.
- F. Guillemin, P. Robert, and B. Zwart. AIMD algorithms and exponential functionals. *Ann. Appl. Probab.*, 14(1):90–117, 2004. ISSN 1050-5164. doi: 10.1214/aoap/1075828048. URL <http://dx.doi.org/10.1214/aoap/1075828048>.

- T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005. ISSN 0091-1798. doi: 10.1214/009117905000000044.
- 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. doi: 10.1051/ps/2016013.
- 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.
- Y. Meyer. *Ondelettes et opérateurs. I*. Actualités Mathématiques. [Current Mathematical Topics]. Hermann, Paris, 1990. ISBN 2-7056-6125-0. Ondelettes. [Wavelets].
- 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. doi: 10.1007/978-3-319-12145-1\_15.
- 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.
- G. Viennet. Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 1997.