HAL open science

Autoregressive Functions Estimation in Nonlinear Bifurcating Autoregressive Models

Siméon Valère Bitseki Penda, Adélaïde Olivier

To cite this version:

Siméon Valère Bitseki Penda, Adélaïde Olivier. Autoregressive Functions Estimation in Nonlinear Bifurcating Autoregressive Models. 2015. hal-01159255v1

HAL Id: hal-01159255

https://hal.science/hal-01159255v1

Preprint submitted on 2 Jun 2015 (v1), last revised 11 Feb 2016 (v2)

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

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

AUTOREGRESSIVE FUNCTIONS ESTIMATION IN NONLINEAR BIFURCATING AUTOREGRESSIVE MODELS

S. VALÈRE BITSEKI PENDA AND ADÉLAÏDE OLIVIER

Abstract. Bifurcating autoregressive processes, which can be seen as an adaptation of autoregressive processes for a binary tree structure, have been extensively studied during the last decade in a parametric context. In this work we do not specify any *a priori* form for the two autoregressive functions and we use nonparametric techniques. We investigate both nonasymptotic and asymptotic behavior of the Nadaraya-Watson type estimators of the autoregressive functions. We build our estimators observing the process on a finite subtree denoted by $\mathbb{T}_n$, up to the depth n. Estimators achieve the classical rate $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$ in quadratic loss over Hölder classes of smoothness. We prove almost sure convergence, asymptotic normality giving the bias expression when choosing the optimal bandwidth and a moderate deviations principle. Our proofs rely on specific techniques used to study bifurcating Markov chains. Finally, we address the question of asymmetry and develop an asymptotic test for the equality of the two autoregressive functions.

Keywords: Bifurcating Markov chains, binary trees, bifurcating autoregressive processes, nonparametric estimation, Nadaraya-Watson estimator, minimax rates of convergence, asymptotic normality, moderate deviations principle, asymmetry test.

Mathematics Subject Classification (2010): 62G05, 62G10, 62G20, 60J80, 60F05, 60F10, 60J20, 92D25.

1. Introduction

1.1. A generalization of the bifurcating autoregressive model.

BAR process. Roughly speaking, bifurcating autoregressive processes, BAR processes for short, are an adaptation of autoregressive processes when the index set have a binary tree structure. BAR processes were introduced by Cowan and Staudte [16] in 1986 in order to study cell division in *Escherichia Coli* bacteria. Let $\mathbb{T}$ be the infinite binary tree indexed in the following way

$$\mathbb{T} = \bigcup_{m=0}^{\infty} \{0,1\}^m.$$

We can see the nodes of $\mathbb{T}$ as representing a given population of cells, where the initial individual is denoted by $\emptyset$. We suppose that each individual $u \in \mathbb{T}$ gives birth to two individuals $u0 = (u,0)$ and $u1 = (u,1)$, called respectively type 0 pole and type 1 pole. In [16], the original BAR process is defined as follows. If X_u denotes the observation of some feature of cell u, then the first order BAR process is given, for all $u \in \mathbb{T}$, by

$$\begin{cases} X_{u0} = a + bX_u + \varepsilon_{u0}, \\ X_{u1} = a + bX_u + \varepsilon_{u1}, \end{cases}$$

drawing the initial value $X_\emptyset$ according to a Gaussian law, where a and b are unknown real parameters, with $|b| < 1$, related to inherited effects. The noise sequence $\big((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T}\big)$ forms a

sequence of independent and identically distributed bivariate centered Gaussian random variables and represents environmental effects.

Since then, several extensions of this model have been studied and various estimators for unknown parameters have been proposed. First, one can mention [27]. There Guyon introduces asymmetry to take into account the fact that autoregressive parameters for poles of type 0 or 1 can differ. Introducing the bifurcating Markov chain theory, Guyon studies the asymptotic behavior of the least squares estimators of the unknown parameters. He also introduces some asymptotic tests which allow to decide if the model is symmetric or not. Several extensions of this linear model have been proposed and studied from a parametric point of view, see for instance Basawa and Huggins [2, 3] and Basawa and Zhou [4, 5] where the BAR process is studied for non-Gaussian noise and long memory. Next, around 2010, Bercu, Blandin, Delmas, de Saporta, Gégout-Petit and Marsalle extended in different directions the study of the BAR process. Bercu *et al.* [7] use martingale approach in order to study least squares estimators of unknown parameters for processes with memory greater than 1 and without the normality assumption of the noise sequence. A number of extensions were subsequently surveyed using the approaches of Guyon and Bercu *et al.*, one can cite [17, 20]. Even more recently, [21] takes into account missing data and [6, 14, 22] study the model with random coefficients. However, to our best knowledge, there is no extensive nonparametric study for nonlinear bifurcating autoregressive processes, where there is no *a priori* specification on the two autoregressive functions. We can mention the applications of [12] (section 4) where deviations inequalities are derived for Nadaraya-Watson type estimators of the autoregressive functions.

Nonlinear BAR process. Nonlinear bifurcating autoregressive (NBAR, for short) processes generalize bifurcating autoregressive (BAR, for short) processes. Let us now give a precise definition for NBAR processes. We now work on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For an individual $u \in \mathbb{T}$, we are interested in some feature X_u (it may be its size, its growth rate, its age, ...) with values in the metric space $\mathcal{S}$ endowed with its Borel σ-field $\mathfrak{S}$. We suppose that the features X_{u0} and X_{u1} associated to its offspring, denoted by $u0$ and $u1$, depend on the past only through X_u, via an autoregressive system of equations. In this setting we have the following

Definition 1 (NBAR process)**.** *Let μ be a probability on $(\mathcal{S}, \mathfrak{S})$, g_ε be a probability on $(\mathcal{S} \times \mathcal{S}, \mathfrak{S})$ with a null first order moment and $f_0, f_1 : \mathcal{S} \to \mathcal{S}$ be two functions. We say that $X = (X_u, u \in \mathbb{T})$ is a nonlinear bifurcating autoregressive process (NBAR process for short) with initial distribution μ and autoregressive functions f_0 and f_1 if*

(i) $X_\emptyset \sim \mu$,
(ii) *For any $u \in \mathbb{T}$,*

$$X_{u0} = f_0(X_u) + \varepsilon_{u0} \quad \text{and} \quad X_{u1} = f_1(X_u) + \varepsilon_{u1}$$

where $\big((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T}\big)$ is a sequence of independent and identically distributed centered bivariate random variables with common density g_ε, independent on $X_\emptyset$.

We denote by $\mathbb{P}_\mu$ the law of the process $(X_u, u \in \mathbb{T})$ built in such a way and by $\mathbb{E}_\mu[\cdot]$ the expectation with respect to the probability $\mathbb{P}_\mu$.

1.2. Estimation of the autoregressive functions and objectives.

Autoregressive functions estimators. Our aim is to estimate the unknown autoregressive functions f_0 and f_1 in Definition 1 from the observation of a subpopulation. For that purpose, we propose to make use of a Nadaraya-Watson type estimator. Two seminal articles in nonparametric regression function estimation where written in 1964 independently by Nadaraya [41] and Watson [48].

They introduced a kernel estimator, now called Nadaraya-Watson estimator, for the regression function $\mathbb{E}\big[Y|X=x\big]$ when observing n pairs $(X_1,Y_1),\ldots,(X_n,Y_n)$ of independent random variables distributed as (X,Y). The Nadaraya-Watson estimator was also used in the framework of autoregressive time series. In that case the aim is to reconstruct $\mathbb{E}\big[X_n|X_{n-1}=x\big]$, assuming that $\big(X_n, n\geq 0\big)$ is stationary, see [44, 29]. We generalize here the use of the Nadaraya-Watson estimator to the case of an autoregressive process indexed by a binary tree.

For the sake of simplicity, we assume that $\mathcal{S}\subset\mathbb{R}$. Let $\mathcal{D}\subset\mathcal{S}$ be a compact interval. We propose to estimate $(f_0(x),f_1(x))$ the autoregressive functions at point $x\in\mathcal{D}$ from the observations

$$\big(X_u, u\in\mathbb{T}_{n+1}\big)$$

by

$$\Big(\widehat{f}_{\iota,n}(x)=\frac{|\mathbb{T}_n|^{-1}\sum\limits_{u\in\mathbb{T}_n}K_{h_n}(x-X_u)X_{u\iota}}{\big(|\mathbb{T}_n|^{-1}\sum\limits_{u\in\mathbb{T}_n}K_{h_n}(x-X_u)\big)\vee\varpi_n}\,,\iota\in\{0,1\}\Big), \tag{1}$$

where $\varpi_n>0$ and we set $K_{h_n}(\cdot)=h_n^{-1}K(h_n^{-1}\cdot)$ for $h_n>0$ and a kernel function $K:\mathbb{R}\to\mathbb{R}$ such that $\int_{\mathbb{R}}K=1$.

Objectives. Our first objective in this work is to study the estimators (1) both from nonasymptotic and asymptotic points of view. The nonasymptotic study includes the control of the quadratic loss in a minimax sense (Theorems 8 and 9) and the asymptotic study includes almost sure convergence (Proposition 10), asymptotic normality (Theorems 11 and 12) and a moderate deviations principle (Theorem 16). To this end, we shall make use of nonasymptotic behavior for bifurcating Markov chains (see [27, 11]) and asymptotic behavior of martingales. We are also interested in comparing the two autoregressive functions f_0 and f_1 and to test whether the phenomenon studied is symmetric or not. The test we build to do so relies on the asymptotic study.

The present work is organized as follows. The results are obtained under the assumption of geometric ergodicity of the so-called tagged-branch Markov chain we define in Section 2, together with the nonasymptotic results. In Section 3, we state asymptotic results for our estimators which enable us to address the question of asymmetry and build a test to compare f_0 and f_1. After discussion in Section 4, we also give there some numerical results to illustrate the estimation and test strategies (Section 5). The last part of the article, Section 6 with an appendix, is devoted to the proofs of our results.

2. Nonasymptotic behavior

2.1. **Tagged-branch chain.** Let us first introduce some usual notation for binary trees. We denote by $\mathbb{G}_m$ the set of individuals belonging to the m-th generation ($m\geq 0$), $\mathbb{T}_r$ the set of individuals up to the r-th generation ($r\geq 0$) and $\mathbb{T}$ the infinite binary tree which represent the whole population:

$$\mathbb{G}_m=\{0,1\}^m\,,\quad \mathbb{T}_r=\bigcup_{m=0}^{r}\mathbb{G}_m\,,\quad \mathbb{T}=\bigcup_{m=0}^{\infty}\mathbb{G}_m\,,$$

with the convention $\mathbb{G}_0=\{\emptyset\}$ where $\emptyset$ is called the initial individual. The cardinality of the set $\mathbb{G}_m$, *e.g.*, is denoted by $|\mathbb{G}_m|$. We also use throughout the following standard notation: for $u\in\mathbb{G}_m$, $m\geq 0$, its length (the generation to which it belongs) is $|u|=m$ and we write $u0=(u,0)\in\mathbb{G}_{m+1}$

and $u1 = (u, 1) \in \mathbb{G}_{m+1}$ for its children.

The main tools in the study of the previous estimators come from the following property. A NBAR process $(X_u, u \in \mathbb{T})$ as defined above is a bifurcating Markov chain on $\mathbb{T}$ (see Definition 3) with

$$\boldsymbol{\mathcal{P}}(x, dy, dz) = g_\varepsilon(dy - f_0(x), dz - f_1(x)) \tag{2}$$

as $\mathbb{T}$-transition probability (see Definition 2).

Definition 2 ($\mathbb{T}$-transition probability, see [27])**.** *We call $\mathbb{T}$-transition probability any mapping $\boldsymbol{\mathcal{P}} : \mathcal{S} \times \mathfrak{S}^2 \to [0, 1]$ such that*

(i) *$\boldsymbol{\mathcal{P}}(\cdot, A)$ is measurable for all $A \in \mathfrak{S}^2$,*
(ii) *$\boldsymbol{\mathcal{P}}(x, \cdot)$ is a probability measure on $(\mathcal{S}^2, \mathfrak{S}^2)$ for all $x \in \mathcal{S}$.*

For a $\mathfrak{S}$-measurable function $f : \mathcal{S}^3 \to \mathcal{S}$, we denote by $\boldsymbol{\mathcal{P}} f$ the $\mathfrak{S}$-measurable function

$$x \in \mathcal{S} \rightsquigarrow \boldsymbol{\mathcal{P}} f(x) = \int_{\mathcal{S}^2} f(x, y, z) \boldsymbol{\mathcal{P}}(x, dy, dz),$$

when it is defined.

Definition 3 (Bifurcating Markov chain, see [27])**.** *Let $(X_u, u \in \mathbb{T})$ be a family of $\mathcal{S}$-valued random variables defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_m, m \geq 0), \mathbb{P})$. Let μ be a probability on $(\mathcal{S}, \mathfrak{S})$ and $\boldsymbol{\mathcal{P}}$ be a $\mathbb{T}$-transition probability. We say that $(X_u, u \in \mathbb{T})$ is a $(\mathcal{F}_m)$-bifurcating Markov chain with initial distribution μ and $\mathbb{T}$-transition probability $\boldsymbol{\mathcal{P}}$ if*

(i) *$X_\emptyset \sim \mu$,*
(ii) *X_u is $\mathcal{F}_{|u|}$-measurable for all $u \in \mathbb{T}$,*
(iii) *for all $m \geq 0$ and for all family $(f_u, u \in \mathbb{G}_m)$ of $\mathfrak{S}$-measurable functions from $\mathcal{S}^3$ to $\mathcal{S}$,*

$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_m} f_u(X_u, X_{u0}, X_{u1}) \Big| \mathcal{F}_m\Big] = \prod_{u \in \mathbb{G}_m} \boldsymbol{\mathcal{P}} f_u(X_u).$$

Bifurcating Markov chains are studied by Guyon in [27], it allows the case where the two offspring are not conditionally independent as we can see in the property (iii) of Definition 3.

For a $\mathbb{T}$-transition probability $\boldsymbol{\mathcal{P}}$ on $\mathcal{S} \times \mathfrak{S}^2$, we denote by $\mathcal{P}_0$ and $\mathcal{P}_1$ respectively the first and the second marginal of $\boldsymbol{\mathcal{P}}$, that is

$$\mathcal{P}_0(x, B) = \boldsymbol{\mathcal{P}}(x, B \times \mathcal{S}), \quad \mathcal{P}_1(x, B) = \boldsymbol{\mathcal{P}}(x, \mathcal{S} \times B)$$

for all $x \in \mathcal{S}$ and $B \in \mathfrak{S}$. We denote by Q the mean of $\mathcal{P}_0$ and $\mathcal{P}_1$,

$$Q = (\mathcal{P}_0 + \mathcal{P}_1)/2. \tag{3}$$

Let $Y = (Y_m, m \geq 0)$ be the Markov chain on $\mathcal{S}$ with initial value $Y_0 = X_\emptyset$ and transition probability Q. The chain $(Y_m, m \geq 0)$ corresponds to a random lineage taken in the population. In the specific case of NBAR processes, the tagged-branch chain Y can be constructed as follows. Let $(\iota_m, m \geq 1)$ be a sequence of independent and identically distributed random variables, independent on X, that have the Bernoulli distribution of parameter $1/2$. Let $(\varepsilon'_m, m \geq 1)$ be a sequence of independent and identically distributed random variables, independent on X, that have probability law $(g_0 + g_1)/2$, where

$$g_0(\cdot) = \int_{\mathcal{S}} g_\varepsilon(\cdot, dy) \quad \text{and} \quad g_1(\cdot) = \int_{\mathcal{S}} g_\varepsilon(dx, \cdot) \tag{4}$$

are the marginals of g_ε. Set

$$\left\{ \begin{array}{l} Y_0 = X_\emptyset, \\ Y_m = f_{\iota_m}(Y_{m-1}) + \varepsilon'_m\,, \quad m \geq 1. \end{array} \right. \tag{5}$$

Then $\mathbb{P}(Y_1 \in dy | Y_0 = x) = Q(x, dy)$. More generally, for all $m \geq 1$, we have $\mathbb{P}(Y_m \in dy | Y_0 = x) = Q^m(x, dy)$ where Q^m denotes the m-th iterated of Q recursively defined by the formulae

$$Q^0(x, B) = \delta_x(B) \quad \text{and} \quad Q^{m+1}(x, B) = \int_{\mathcal{S}} Q(x, dy) Q^m(y, B) \quad \forall B \in \mathfrak{S}.$$

It is well known that Q^m is a transition probability on $\mathcal{S} \times \mathfrak{S}$. From (2), the explicit expression of Q is

$$Q(x, dy) = \frac{1}{2}\Big(g_0\big(dy - f_0(x)\big) + g_1\big(dy - f_1(x)\big)\Big). \tag{6}$$

Asymptotic and nonasymptotic studies on bifurcating Markov chains have shown that the limit law of the Markov chain Y plays an important role, we refer to [11] and references therein for more details. In the present work, the estimator of the invariant density of the tagged-branch Markov chain will play a crucial role in the analysis of the estimators defined by (1). More precisely, we will see that the denominator converges almost surely to the invariant density of the Markov chain Y (Proposition 23).

2.2. **Model contraints.** Recall we assume that $\mathcal{S} \subset \mathbb{R}$. The autoregressive functions are devoted to belong to the following class. For $\gamma \in (0, 1)$ and $\ell > 0$, we introduce the class $\mathcal{F}(\gamma, \ell)$ of continuous functions $f : \mathcal{S} \to \mathcal{S}$ such that

$$|f(x)| \leq \gamma |x| + \ell$$

for any $x \in \mathcal{S}$.

The following Assumption will be in force throughout the present work:

Assumption 4 (Noise density)**.** *The probability g_ε is absolutely continuous with respect to the Lebesgue measure with continuous density on $\mathcal{S}$, also denoted by g_ε abusing slightly notation.*

We denote by g_0 and g_1 the densities with respect to the Lebesgue measure of the probability measures defined by (4). The two marginal densities g_0 and g_1 are devoted to belong to the following class. For $r > 0$ and $\lambda > 2$, we introduce the class $\mathcal{G}(r, \lambda)$ of nonnegative continuous functions $g : \mathcal{S} \to [0, \infty)$ such that

$$g(x) \leq \frac{r}{1 + |x|^\lambda}$$

for any $x \in \mathcal{S}$. When $(g_0, g_1) \in \mathcal{G}(r, \lambda)^2$ for some $\lambda > 3$, we denote the covariance matrix of $(\varepsilon_0, \varepsilon_1)$, called noise covariance matrix, by

$$\Gamma = \begin{pmatrix} \sigma_0^2 & \varrho\sigma_0\sigma_1 \\ \varrho\sigma_0\sigma_1 & \sigma_1^2 \end{pmatrix} \tag{7}$$

with $\sigma_0\,, \sigma_1 > 0$ and $\varrho \in (-1, 1)$.

It is crucial for our proofs to study the ergodicity of the tagged-branch Markov chain Y. Geometric ergodicity of nonlinear autoregressive processes has been studied in [8] (see Theorem 1) and also in [1, 15]. The main difference is that we need here a precise control on the ergodicity rate, which should be smaller that $1/2$, due to the binary tree structure. We also see, through (5), that the autoregressive function is random in our case. The following crucial assumption will guarantee

geometric ergodicity of the tagged-branch Markov chain Y with an exponential decay rate smaller than $1/2$ (see Lemma 20). For any $M > 0$, we set

$$\delta(M) = \min\Big\{ \inf_{|x|\le M} g_0(x); \inf_{|x|\le M} g_1(x)\Big\}. \tag{8}$$

Assumption 5. *Set $M_0 = \ell + \mathbb{E}\big[|\varepsilon_1'|\big] < \infty$ with ε_1' drawn according to $(g_0+g_1)/2$. There exists $\eta > 0$ such that*

$$\gamma < \frac{1}{2} - \eta$$

and there exists $M_1 > 2M_0/(1/2-\eta-\gamma)$ such that

$$2M_1\delta\big((1+\gamma)M_1+\ell\big) > \frac{1}{2}. \tag{9}$$

The following assumption will guarantee that the invariant density ν is positive on some nonempty interval (see Lemma 24). For any $M > 0$, we set

$$\eta(M) = \frac{|g_0|_\infty + |g_1|_\infty}{2}\int_{|y|>M}\int_{x\in\mathcal{S}} \frac{r}{1+\big|y-\gamma|x|-\ell\big|^\lambda \wedge \big|y+\gamma|x|+\ell\big|^\lambda} dxdy,$$

where for a function $h:\mathcal{S}\to\mathbb{R}$, $|h|_\infty$ stands for $\sup_{x\in\mathcal{S}}|h(x)|$.

Assumption 6. *For $M_2 > 0$ such that $\eta(M_2) < 1$, there exists $M_3 > \ell+\gamma M_2$ such that $\delta(M_3) > 0$.*

2.3. **Main results.** We need the following property on K:

Assumption 7. *The kernel $K:\mathbb{R}\to\mathbb{R}$ is bounded with compact support and for some integer $n_0 \ge 1$, we have $\int_{-\infty}^{\infty} x^k K(x)dx = \mathbf{1}_{\{k=0\}}$ for $k = 1,\dots,n_0$.*

Assumption 7 will enable us to have nice approximation results over smooth functions f_0 and f_1, described in the following way: for a compact interval $\mathcal{D}\subset(0,\infty)$ and $\beta > 0$, with $\beta = \lfloor\beta\rfloor + \{\beta\}$, $0 < \{\beta\} \le 1$ and $\lfloor\beta\rfloor$ an integer, let $\mathcal{H}^\beta_{\mathcal{D}}$ denote the Hölder space of functions $h:\mathcal{D}\to\mathbb{R}$ possessing a derivative of order $\lfloor\beta\rfloor$ that satisfies

$$|h^{\lfloor\beta\rfloor}(y) - h^{\lfloor\beta\rfloor}(x)| \le c(h)|x-y|^{\{\beta\}}. \tag{10}$$

The minimal constant $c(h)$ such that (10) holds defines a semi-norm $|g|_{\mathcal{H}^\beta_{\mathcal{D}}}$. We equip the space $\mathcal{H}^\beta_{\mathcal{D}}$ with the norm $\|h\|_{\mathcal{H}^\beta_{\mathcal{D}}} = \sup_x|h(x)| + |h|_{\mathcal{H}^\beta_{\mathcal{D}}}$ and the balls

$$\mathcal{H}^\beta_{\mathcal{D}}(L) = \{h:\mathcal{D}\to\mathbb{R},\ \|h\|_{\mathcal{H}^\beta_{\mathcal{D}}} \le L\},\ L > 0.$$

Theorem 8 (Upper rate of convergence)**.** *Work under Assumption 4. Let $\gamma\in(0,1/2)$ and $\ell > 0$, let $r > 0$ and $\lambda > 3$. Specify $(\widehat{f}_{0,n},\widehat{f}_{1,n})$ with a kernel K satisfying Assumption 7 for some $n_0 > 0$, with*

$$h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$$

and $\varpi_n > 0$ such that $\varpi_n \to 0$ as $n\to\infty$. For every $L, L' > 0$ and $0 < \beta < n_0$, for every g_ε such that $(g_0,g_1)\in\big(\mathcal{G}(r,\lambda)\cap\mathcal{H}^\beta_{\mathcal{S}}(L')\big)^2$ satisfy Assumptions 5 and 6, there exists $d = d(\gamma,\ell,g_0,g_1) > 0$ such that for every compact interval $\mathcal{D}\subset[-d,d]$ with non-empty interior and for every x in the interior of $\mathcal{D}$,

$$\sup_{(f_0,f_1)} \mathbb{E}_\mu\Big[\big(\widehat{f}_{0,n}(x)-f_0(x)\big)^2 + \big(\widehat{f}_{1,n}(x)-f_1(x)\big)^2\Big]^{1/2} \lesssim \varpi_n^{-1}|\mathbb{T}_n|^{-\beta/(2\beta+1)}$$

where the supremum is taken among all functions $(f_0, f_1) \in \big(\mathcal{F}(\gamma,\ell) \cap \mathcal{H}^{\beta}_{\mathcal{D}}(L)\big)^2$*, for any initial measure* $\mu(dx)$ *on* $\mathcal{S}$ *for* $X_\emptyset$ *such that* $\mu\big((1+|\cdot|)^2\big) < \infty$*. The notation* $\lesssim$ *means up to some constant independent on* n*.*

Some comments are in order: **1)** The threshold ϖ_n should be chosen such that it inflates the upper-rate of convergence of a slow factor only. Typically, $\varpi_n = (\ln n)^{-1}$ is suitable. **2)** Up to the factor ϖ_n^{-1}, we obtain the classical rate $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$ where $|\mathbb{T}_n|$ is the number of observations. We know it is optimal in a minimax sense in a density estimation framework and we can infer this is optimal in our framework too. To prove it is the purpose of Theorem 9 which follows. **3)** Looking carefully at the proof, see (27), we actually see that $\varpi_n \to 0$ as $n \to \infty$ is not necessary. One could choose, $\varpi_n = \varpi$ with

$$\varpi = \frac{1}{2} \inf_{(f_0,f_1)} \inf_{x\in\mathcal{D}} \nu(x) > 0$$

where the infimum is taken among all $(f_0, f_1) \in \mathcal{F}(\gamma,\ell)^2$ and where $\varpi > 0$ is guaranteed by Lemma 24. However, to calibrate in practice ϖ_n in such a way is not possible since we cannot compute ϖ. **4)** For the sake of simplicity, we have picked a common bandwidth h_n to define the two estimators, but one can immediately generalize our study for two different bandwidths $(h_{\iota,n} = |\mathbb{T}_n|^{-\beta_\iota/(2\beta_\iota+1)}, \iota \in \{0,1\})$ where β_ι is the Hölder smoothness of f_ι. **5)** We do not achieve adaptivity in the smoothness of the autoregressive functions since our choice of bandwidth h_n still depends on β, we refer to the discussion in Section 4.

Theorem 9 (Lower rate of convergence)**.** *Assume the noise density* g_ε *is a bivariate Gaussian density. Let* $\mathcal{D} \subset \mathcal{S}$ *be a compact interval. For every* $\gamma \in (0,1)$ *and every positive* ℓ, β, L*, there exists* $C > 0$ *such that, for every* $x \in \mathcal{D}$*,*

$$\liminf_{n\to\infty} \inf_{(\widehat{f}_{0,n},\widehat{f}_{1,n})} \sup_{(f_0,f_1)} \mathbb{P}\Big(|\mathbb{T}_n|^{\beta/(2\beta+1)}\big(\big|\widehat{f}_{0,n}(x) - f_0(x)\big| + \big|\widehat{f}_{1,n}(x) - f_1(x)\big|\big) \geq C\Big) > 0,$$

where the supremum is taken among all functions $(f_0, f_1) \in \big(\mathcal{F}(\gamma,\ell) \cap \mathcal{H}^{\beta}_{\mathcal{D}}(L)\big)^2$ *and the infimum is taken among all estimators based on* $(X_u, u \in \mathbb{T}_{n+1})$*.*

This result obviously implies a lower rate of convergence for the mean quadratic loss at point x. We see that in a regular case, the Gaussian case, the lower and upper rates match.

3. Asymptotic behavior

3.1. Main results.

Asymptotic normality. The almost-sure convergence of the autoregressive functions estimators is obtained in Proposition 10 for any $h_n = |\mathbb{T}_n|^{-\alpha}$ with $\alpha \in (0,1)$. Choosing $\alpha \geq 1/(2\beta+1)$, the estimator $\big(\widehat{f}_{0,n}(x), \widehat{f}_{1,n}(x)\big)$ recentered by $\big(f_0(x), f_1(x)\big)$ and normalized by $\sqrt{|\mathbb{T}_n| h_n}$ converges in distribution to a Gaussian law. Depending on $\alpha > 1/(2\beta+1)$ or not, the limit Gaussian law is centered or not, as we state in Theorems 11 and 12.

Proposition 10 (Almost sure convergence)**.** *In the same setting as in Theorem 8 with* $h_n = |\mathbb{T}_n|^{-\alpha}$ *for* $\alpha \in (0,1)$*,*

$$\begin{pmatrix} \widehat{f}_{0,n}(x) \\ \widehat{f}_{1,n}(x) \end{pmatrix} \to \begin{pmatrix} f_0(x) \\ f_1(x) \end{pmatrix}, \quad \mathbb{P}_\mu - a.s.$$

as $n \to \infty$*.*

From now on we need to reinforce the assumption on the noise sequence: we require that the noise $(\varepsilon_0, \varepsilon_1)$ has finite moment of order 4, $\mathbb{E}\big[\varepsilon_0^4 + \varepsilon_1^4\big] < \infty$, which is guaranteed by $\lambda > 5$. We denote by $|K|_2^2 = \int_{\mathcal{S}} K(x)^2 dx$.

Theorem 11 (Asymptotic normality)**.** *In the same setting as in Theorem 8 with $\lambda > 5$ and $h_n = |\mathbb{T}_n|^{-\alpha}$ for $\alpha > 1/(2\beta+1)$,*

$$\sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\big(\mathbf{0}_2, \mathbf{\Sigma}_2(x)\big) \quad \textit{with} \quad \mathbf{\Sigma}_2(x) = |K|_2^2 \big(\nu(x)\big)^{-1} \Gamma,$$

Γ being the noise covariance matrix and $\mathbf{0}_2 = (0,0)$. Moreover, for $x_1, \dots, x_k$ distinct points in $\mathcal{D}$, the sequence

$$\left(\begin{pmatrix} \widehat{f}_{0,n}(x_l) - f_0(x_l) \\ \widehat{f}_{1,n}(x_l) - f_1(x_l) \end{pmatrix}, l = 1, \dots, k \right)$$

is asymptotically independent.

The restriction $\alpha > 1/(2\beta+1)$ in Theorem 11 prevents us from choosing $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$, which is the optimal choice to achieve the minimax rate as we have seen in Theorem 8. The following Theorem remedies to this flaw, but at the cost of an unknown bias. We obtain an explicit expression of this bias for β an integer which depends on the β-th derivatives of the autoregressive functions and the invariant measure of the tagged-branch chain.

Theorem 12 (Asymptotic normality with bias expression)**.** *In the same setting as in Theorem 8 with $\lambda > 5$ and β an integer,*

(i) *If $h_n^\beta \sqrt{|\mathbb{T}_n| h_n} \to \kappa$ with $\kappa \in [0, \infty)$ as $n \to +\infty$, then*

$$\sqrt{|\mathbb{T}_n| h_n} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{d} \mathcal{N}_2\left(\kappa \boldsymbol{m}_2(x), \mathbf{\Sigma}_2(x)\right) \quad \textit{with} \quad \mathbf{\Sigma}_2(x) = |K|_2^2 \big(\nu(x)\big)^{-1} \Gamma$$

and

$$\boldsymbol{m}_2(x) = \frac{(-1)^\beta}{\beta!\, \nu(x)} \int_{\mathcal{S}} y^\beta K(y) dy \begin{pmatrix} (\nu f_0)^\beta(x) - \nu^\beta(x) f_0(x) \\ (\nu f_1)^\beta(x) - \nu^\beta(x) f_1(x) \end{pmatrix}.$$

(ii) *If $h_n^\beta \sqrt{|\mathbb{T}_n| h_n} \to \infty$ as $n \to +\infty$, then*

$$h_n^{-\beta} \begin{pmatrix} \widehat{f}_{0,n}(x) - f_0(x) \\ \widehat{f}_{1,n}(x) - f_1(x) \end{pmatrix} \xrightarrow{\mathbb{P}_\mu} \boldsymbol{m}_2(x).$$

If β is not an integer, we could generalize Theorem 12 but at the cost of introducing fractional derivatives. Note that the definition of this notion is not unique (see [45] or [47]). We decided to restrict the parameter β to be an integer in order to avoid here additional technicalities.

Moderate deviations principle. Statistical estimators are also studied under the angle of large and moderate deviations principles. Large and moderate deviations limit theorems are proved in the independent setting for the kernel density estimator and also for the Nadaraya-Watson estimator (see [35, 36, 32] in the univariate case, see also [39] and references therein). We also refer to Mokkadem and Pelletier [38] for the study of confidence bands based on the use of moderate deviations principles. The purpose of this section is to establish a moderate deviations principle for the autoregressive functions estimators defined by (1).

Let us introduce the notion of moderate deviations principle in a general setting. Let $(Z_n, n \geq 0)$ be a sequence of random variables with values in a Polish space $\mathcal{S}$ endowed with its Borel σ-field

$\mathfrak{S}$ and let $(s_n, n \geq 0)$ be a positive sequence that converges to $+\infty$. We assume that Z_n/s_n converges in probability to 0 and that $Z_n/\sqrt{s_n}$ converges in distribution to a centered Gaussian law. Let $I : \mathcal{S} \to \mathbb{R}^+$ be a lower semicontinuous function, that is for all $c > 0$ the sub-level set $\{x \in \mathcal{S}, I(x) \leq c\}$ is a closed set. Such a function I is called a rate function and it is called a good rate function if all the sub-level sets are compact sets. Let $(a_n, n \geq 0)$ be a positive sequence such that a_n/s_n converges to 0.

Definition 13 (Moderate deviations principle)**.** *We say that $Z_n/\sqrt{a_n s_n}$ satisfies a moderate deviations principle on $\mathcal{S}$ with speed a_n and the rate function I if, for any $A \in \mathfrak{S}$,*

$$-\inf_{x\in A^\circ} I(x) \leq \liminf_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}\big(\frac{Z_n}{\sqrt{a_n s_n}} \in A\big) \leq \limsup_{n\to\infty} \frac{1}{a_n} \log \mathbb{P}\big(\frac{Z_n}{\sqrt{a_n s_n}} \in A\big) \leq -\inf_{x\in \bar{A}} I(x),$$

where A° and $\bar{A}$ denote respectively the interior and the closure of A.

We will state such a moderate deviations principle for the estimators $\widehat{f}_0(x)$ and $\widehat{f}_1(x)$ in a restricted framework. Since the moderate deviations principle requires a strong notion of convergence, the super-exponential convergence (see Definition 26), it is not surprising to ask the process to also satisfy strong assumptions. The autoregressive functions f_0 and f_1 will be restricted to belong to the class $\mathcal{F}(\gamma, \ell)$ with $\gamma = 0$, which means we have

$$\max\{|f_0|_\infty; |f_1|_\infty\} \leq \ell.$$

To guarantee we have geometric ergodicity of the tagged-branch Markov chain Y uniformly in the initial value with an exponential decay rate smaller than 1/2, we will require the following assumption to replace Assumption 5 (see Lemma 25).

Assumption 14. *There exists $M_4 > \ell$ such that $2(M_4 - \ell)\delta(M_4) > 1/2$ with $\delta(\cdot)$ defined by (8).*

We also reinforce the assumption on the kernel with

Assumption 15. *The kernel $K : \mathbb{R} \to \mathbb{R}$ is such that*

$$\frac{\int_{\mathbb{R}} |K^+(x)|dx}{\int_{\mathbb{R}} |K^-(x)|dx} > \frac{|g_0|_\infty + |g_1|_\infty}{2\delta(M_3)\big(1 - \eta(M_2)\big)}$$

where M_2, M_3 come from Assumption 6 and $K^+(\cdot) = \max\{K(\cdot); 0\}$, $K^-(\cdot) = \min\{K(\cdot); 0\}$.

Theorem 16 (Moderate deviations principle)**.** *In the same setting as in Theorem 8 with $\gamma = 0$, $\lambda > 5$, K satisfying in addition Assumption 15, $h_n = |\mathbb{T}_n|^{-\alpha}$ for $\alpha > 1/(2\beta+1)$ and Assumption 14 replacing Assumption 5, let $(b_n, n \geq 0)$ be a positive sequence such that*

$$(i)\ \lim_{n\to\infty} \frac{b_n}{(|\mathbb{T}_n|h_n)^{1/2}} = +\infty\,, \quad (ii)\ \lim_{n\to\infty} \frac{b_n}{|\mathbb{T}_n|h_n} = 0\,, \quad (iii)\ \lim_{n\to\infty} \frac{b_n}{|\mathbb{T}_n|h_n^{1+\beta}} = +\infty.$$

Then the sequence

$$\left(\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}\widehat{f}_{0,n}(x) - f_0(x)\\ \widehat{f}_{1,n}(x) - f_1(x)\end{pmatrix}, n \geq 0\right)$$

satisfies a MDP on $\mathcal{S}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and good rate function $J_x : \mathcal{S}^2 \to \mathbb{R}$ defined by

$$J_x(\boldsymbol{z}) = \big(2|K|_2^2\big)^{-1}\nu(x)\boldsymbol{z}^t\Gamma^{-1}\boldsymbol{z}\,, \quad \boldsymbol{z} \in \mathcal{S}^2,$$

where $\boldsymbol{z}^t$ stands for the transpose of vector $\boldsymbol{z}$.

The contraction principle (see [19], Chapter 4) enables us to state the following corollary of Theorem 16.

Corollary 17. *In the same setting as Theorem 16, for every $\delta > 0$,*

$$\lim_{n\to+\infty} \frac{|\mathbb{T}_n|h_n}{b_n^2} \log \mathbb{P}\Big(\frac{|\mathbb{T}_n|h_n}{b_n}\big|\widehat{f}_{\iota,n}(x) - f_\iota(x)\big| > \delta\Big) = -\big(2\sigma_\iota^2|K|_2^2\big)^{-1}\nu(x)\delta^2, \quad \iota \in \{0,1\}.$$

3.2. **Construction of an asymmetry test.** Testing in the context of nonparametric regression is a crucial point, especially in applied contexts. The question of no effect in nonparametric regression is early addressed in [26]. We may also want to compare two regression curves nonparametrically and we refer to [40] and references therein. Specific tools have been developed to compare time series, in particular through autocorrelation functions (see for instance the recent work [31] among many others). However the test we propose to study asymmetry follows from the estimation procedure we propose and is more inspired by the regression case.

Let $x_1, \dots, x_k$ be distinct points in $\mathcal{S}$. We are going to build a statistical test that allows us to segregate between hypothesis

$$\mathcal{H}_0 : \forall l \in \{1,\dots,k\}, f_0(x_l) = f_1(x_l) \quad \text{vs.} \quad \mathcal{H}_1 : \exists l \in \{1,\dots,k\}, f_0(x_l) \neq f_1(x_l).$$

In the parametric studies on *E. Coli* [27, 17], these tests are known as detection of cellular aging and they permit to decide if the cell division is symmetric or asymmetric.

Inspired by [9] we define new estimators in order to both achieve the rate $|\mathbb{T}_n|^{-\beta/(2\beta+1)}$ in the asymptotic normality theorem and remove the bias. Let $\big(\widehat{f}^{(a)}_{\iota,n}(x), \iota \in \{0,1\}\big)$ be the estimators (1) with bandwidth $h_n^{(a)} = |\mathbb{T}_n|^{-1/(2\beta+1)}$ and $\big(\widehat{f}^{(b)}_{\iota,n}(x), \iota \in \{0,1\}\big)$ be the estimators (1) with bandwidth $h_n^{(b)} = |\mathbb{T}_n|^{-\delta/(2\beta+1)}$ for some $\delta \in (0,1)$. We assume that β is an integer so that the results of Theorem 12 are valid. We define

$$\Big(\bar{f}_{\iota,n}(x) = \big(1 - |\mathbb{T}_n|^{\frac{-\beta(1-\delta)}{2\beta+1}}\big)^{-1}\big(\widehat{f}^{(a)}_{\iota,n}(x) - |\mathbb{T}_n|^{\frac{-\beta(1-\delta)}{2\beta+1}}\widehat{f}^{(b)}_{\iota,n}(x)\big), \iota \in \{0,1\}\Big), \tag{11}$$

new estimators such that

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}}\begin{pmatrix}\bar{f}_{0,n}(x) - f_0(x)\\ \bar{f}_{1,n}(x) - f_1(x)\end{pmatrix} \xrightarrow{d} \mathcal{N}_2\big(\mathbf{0}_2, \mathbf{\Sigma}_2(x)\big) \quad \text{with} \quad \mathbf{\Sigma}_2(x) = |K|_2^2\big(\nu(x)\big)^{-1}\Gamma. \tag{12}$$

As announced the trick of [9] enables us to remove the unknown bias while keeping the optimal rate of convergence. We now define a test statistics based on these new estimators by

$$W_n(x_1,\dots,x_k) = \frac{|\mathbb{T}_n|^{\frac{2\beta}{2\beta+1}}}{(\sigma_0^2 + \sigma_1^2 - 2\sigma_0\sigma_1\varrho)|K|_2^2}\sum_{l=1}^k \widehat{\nu}_n(x_l)\big(\bar{f}_{0,n}(x_l) - \bar{f}_{1,n}(x_l)\big)^2 \tag{13}$$

with $\widehat{\nu}_n(\cdot) = |\mathbb{T}_n|^{-1}\sum_{u\in\mathbb{T}_n} K_{h_n}(X_u - \cdot)$ where $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$.

Theorem 18 (Wald test for asymmetry)**.** *In the same setting as in Theorem 12, let $x_1,\dots,x_k$ be distinct points in $\mathcal{D}$. Then the test statistic $W_n(x_1,\dots,x_k)$ converges in distribution to the chi-squared distribution with k degrees of freedom $\chi^2(k)$, under $\mathcal{H}_0$, and $\mathbb{P}_\mu$-almost surely to $+\infty$, under $\mathcal{H}_1$.*

Note that in (13) we could replace σ_0^2, σ_1^2 and ϱ by

$$\widehat{\sigma}^2_{0,n} = |\mathbb{T}_n|^{-1}\sum_{u\in\mathbb{T}_n}(\widehat{\varepsilon}_{u0})^2, \quad \widehat{\sigma}^2_{1,n} = |\mathbb{T}_n|^{-1}\sum_{u\in\mathbb{T}_n}(\widehat{\varepsilon}_{u1})^2 \quad \text{and} \quad \widehat{\varrho}_n = \widehat{\sigma}^{-1}_{0,n}\widehat{\sigma}^{-1}_{1,n}|\mathbb{T}_n|^{-1}\sum_{u\in\mathbb{T}_n}\widehat{\varepsilon}_{u0}\widehat{\varepsilon}_{u1}$$

with the empirical residuals $\widehat{\varepsilon}_{u\iota} = X_{u\iota} - \widehat{f}_{\iota,n}(X_u)$ for $u \in \mathbb{T}_n$. We claim that these estimators are consistent, so that Theorem 18 is still valid with this other test statistics. Proving the convergence

in probability of these three quantities would imply some technical calculations and we do not give here more details.

4. Discussion

Recursive estimators. For the sake of clarity, we have picked a bandwidth that depends on the size of the whole subtree $\mathbb{T}_n$. We could also choose bandwidths that depend on the sizes of the generations. More precisely, we could estimate $\big(f_0(x), f_1(x)\big)$ the autoregressive functions at point $x \in \mathcal{D}$ from the observations $(X_u, u \in \mathbb{T}_{n+1})$ by

$$\Big(\widehat{f}_{\iota,n}(x) = \frac{\sum_{m=0}^{n} \sum_{u \in \mathbb{G}_m} K_{h_m}(x - X_u) X_{u\iota}}{\sum_{m=0}^{n} \sum_{u \in \mathbb{G}_m} K_{h_m}(x - X_u)}, \iota \in \{0,1\}\Big)$$

with the collection of bandwidths $(h_m = |\mathbb{G}_m|^{-\alpha}, 0 \leq m \leq n)$ for $\alpha \in (0,1)$. These estimators can be seen as a version of recursive Nadaraya-Watson estimators when the index set has a binary tree structure. We stress that our results also hold for these alternative estimators.

Heteroscedasticity. More generally, given two functions $\sigma_0, \sigma_1 : \mathcal{S} \to [0, \infty)$, we could consider the autoregressive equations

$$X_{u0} = f_0(X_u) + \sigma_0(X_u)\varepsilon_{u0} \quad \text{and} \quad X_{u1} = f_1(X_u) + \sigma_1(X_u)\varepsilon_{u1}$$

with $\mathbb{E}[\varepsilon_{u0}^2] = \mathbb{E}[\varepsilon_{u1}^2] = 1$ and $\mathbb{E}[\varepsilon_{u0}\varepsilon_{u1}] = \varrho$ where $\varrho \in (-1,1)$. Assuming $0 < \inf_{x \in \mathcal{S}} \sigma_\iota(x) \leq \sup_{x \in \mathcal{S}} \sigma_\iota(x) < \infty$ for $\iota \in \{0,1\}$, Theorems 11 and 12 still hold with

$$\mathbf{\Sigma}_2(x) = |K|_2^2 \big(\nu(x)\big)^{-1} \begin{pmatrix} \sigma_0^2(x) & \varrho\sigma_0(x)\sigma_1(x) \\ \varrho\sigma_0(x)\sigma_1(x) & \sigma_1^2(x) \end{pmatrix}.$$

The estimation of the variance functions σ_0 and σ_1 would be in practice of great interest but the theoretical study of such estimators lies here beyond the scope of the article.

Adaptive estimation. Our estimators of the autoregressive functions are not adaptive in the smoothness β of these two functions since the choice of the bandwidth $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ still depends on β. For classical autoregressive models (*i.e.* non-bifurcating), we refer to the early work by Hoffmann [30]. This has been also investigated in [18] and Delouille and van Sachs propose there a protocol with design-adapted wavelets. It is well-known that the question of adaptivity boils down to proving deviations inequalities. In our framework, we should prove deviations inequalities for the two empirical means

$$|\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) X_{u\iota} \quad \text{and} \quad |\mathbb{T}_n|^{-1} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u).$$

In the restricted class of bounded autoregressive functions $\mathcal{F}(0, \ell)$, with compactly supported noise density g_ε, the results of [13] brings such deviations inequalities. Indeed, on the class $\mathcal{F}(0, \ell)$, we can achieve uniform ergodicity, a key assumption in [13]. However, on the more general class $\mathcal{F}(\gamma, \ell)$ with $\gamma > 0$, the problem of deviations inequalities is still open. We recall that on the class $\mathcal{F}(\gamma, \ell)$ with $\gamma > 0$, only $\mathbb{V}$-geometric ergodicity with $\mathbb{V}$ a Lyapunov function is achieved.

Uniform test. The asymmetry test we have built is based on the choice of a grid of points on $\mathcal{S}$. A theoretical result is needed in order to build a uniform test on a interval $\mathcal{D} \subset \mathcal{S}$. More precisely, to achieve such a uniform test we should study the asymptotic behavior of

$$\sup_{x\in\mathcal{D}} \sqrt{|\mathbb{T}_n|h_n}\big|\widehat{f}_{\iota,n}(x) - f_\iota(x)\big|, \quad \iota \in \{0,1\}.$$

This asymptotic study lies in the scope of the theory of extrema. One can see the study of Liu and Wu [34] for autoregressive processes of order 1: an asymptotic Gumbel behavior is highlighted for the Nadaraya-Watson type estimator of the autoregressive function. One can also see Li [33] where the aim is to compare the two autoregressive functions of two autoregressive processes of order 1: a test statistics based on estimated residuals is studied there.

Weaker assumption on the noise. More realistic assumptions on the noise sequence, which represent environmental effects, should be considered. As in Bercu, de Saporta and Gégout-Petit [7], we could assume that the noise is such that, for any $u \in \mathbb{G}_m$,

$$\mathbb{E}[\varepsilon_u|\mathcal{F}_{m-1}] = 0$$

where $\mathcal{F}_{m-1}$ is the sigma-field generated by $(X_u, u \in \mathbb{T}_{m-1})$. This is much weaker than assuming $\big((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T}\big)$ is an independent and identically distributed sequence but the Markovian structure of $(X_u, u \in \mathbb{T})$ is lost and the techniques we have used in this work are no longer available.

5. Numerical implementation

The goal of this section is to illustrate the theoretical results of the previous sections, in particular the results of Theorem 8 (Upper rate of convergence) and Theorem 18 (Wald test for asymmetry).

Quality of the estimation procedure. We pick trial autoregressive functions defined analytically by

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_1(x) = x(1/8 + \exp(-x^2)/2)$$

for $x \in \mathcal{S} = \mathbb{R}$. We take a Gaussian noise with $\sigma_0^2 = \sigma_1^2 = 1$ and $\varrho = 0.3$. We simulate $M = 500$ times a NBAR process up to generation $n+1 = 15$, with root $X_\emptyset = 1$. We take a Gaussian kernel $K(x) = (2\pi)^{-1/2}\exp(-x^2/2)$ and $h_n = |\mathbb{T}_n|^{-1/5}$ in order to implement estimators given by (1). We did not meet any problem with the denominator in practice and actually set $\varpi_n = 0$. We evaluate $\widehat{f}_{0,n}$ and $\widehat{f}_{1,n}$ on a regular grid of $\mathcal{D} = [-3,3]$ with mesh $\Delta x = |\mathbb{T}_n|^{-1/2}$. For each sample we compute the empirical error

$$e_\iota^{(i)} = \frac{\|\widehat{f}_{\iota,n}^{(i)} - f_\iota\|_{\Delta x}}{\|f_\iota\|_{\Delta x}}, \quad i = 1, \dots, M,$$

where $\|\cdot\|_{\Delta x}$ denotes the discrete norm over the numerical sampling. Table 1 displays the mean-empirical errors together with the empirical standard deviations,

$$\overline{e}_\iota = M^{-1}\sum_{i=1}^M e_\iota^{(i)} \quad \text{and} \quad \big(M^{-1}\sum_{i=1}^M (e_\iota^{(i)} - \overline{e}_\iota)^2\big)^{1/2}$$

for $\iota \in \{0,1\}$. The larger n, the better the reconstruction of f_0 and f_1 as shown in Table 1. This is also true at a visual level, as shown on Figure 1 where 95%-level confidence bands are built so that for each point x, the lower and upper bounds include 95% of the estimators $(\widehat{f}_{0,n}^{(i)}(x), i = 1 \dots M)$. As one can see on Figure 1, the reconstruction is good around 0 and deteriorates for large or small x. The same analysis holds for the reconstruction of f_1, see the thin blue lines. The error is close to $|\mathbb{T}|^{-2/5}$ for both $\widehat{f}_{0,n}$ and $\widehat{f}_{1,n}$ as expected: indeed, for a kernel of order n_0, the bias term

in density estimation is of order $h^{\beta\wedge(n_0+1)}$. For the smooth f_0, f_1 and ν we have here, we rather expect for the rate $|\mathbb{T}_n|^{-(n_0+1)/(2(n_0+1)+1)} = |\mathbb{T}|^{-2/5}$ for the Gaussian kernel with $n_0 = 1$ that we use here, and this is consistent with what we observe in Figure 2.

$\boldsymbol{n}$	8	9	10	11	12	13	14
$\lvert\mathbb{T}_{\boldsymbol{n}}\rvert$	511	1 023	2 047	4 095	8 191	16 383	32 767
$\overline{\boldsymbol{e}}_{\mathbf{0}}$	0.4442	0.3417	0.2633	0.2006	0.1517	0.1285	0.0891
Std. dev.	*0.1509*	*0.1063*	*0.0761*	*0.0558*	*0.0387*	*0.0295*	*0.0209*
$\overline{\boldsymbol{e}}_{\mathbf{1}}$	0.6696	0.5141	0.4006	0.3027	0.2356	0.1776	0.1384
Std. dev.	*0.2482*	*0.1626*	*0.1227*	*0.0831*	*0.0622*	*0.0440*	*0.0326*

TABLE 1. *Mean empirical relative error $\overline{e}_0$ (resp. $\overline{e}_1$) and its standard deviation computed over $M = 500$ Monte-Carlo trees, with respect to $|\mathbb{T}_n|$, for the autoregressive function f_0 (resp. f_1) reconstructed over the interval $\mathcal{D} = [-3, 3]$ by the estimator $\widehat{f}_{0,n}$ (resp. $\widehat{f}_{1,n}$).*

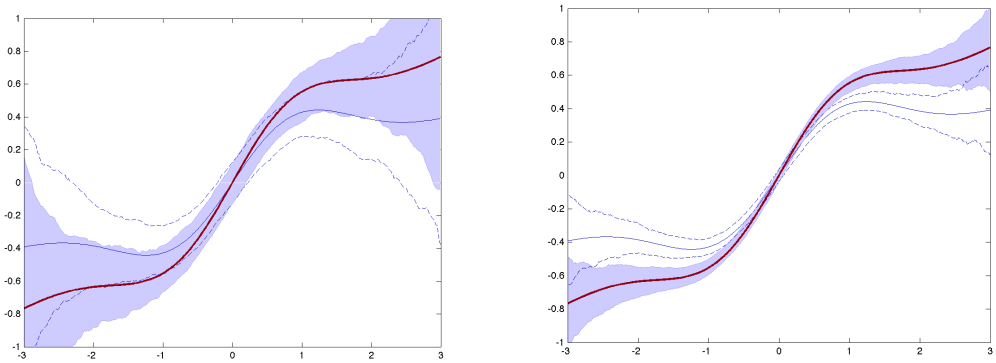

FIGURE 1. *Reconstruction of f_0 over $\mathcal{D} = [-3, 3]$ with 95%-level confidence bands constructed over $M = 500$ Monte-Carlo trees. In bold red line: $x \rightsquigarrow f_0(x)$. In thin blue lines: reconstruction of f_1 with 95%-level confidence bands. Left: $n = 10$ generations. Right: $n = 14$ generations.*

Implementation of the asymmetry test. We implement now the estimators (11) inspired by [9] in order to compute our test statistics (13). We keep a Gaussian kernel and we pick $h_n^{(a)} = |\mathbb{T}_n|^{-1/5}$ and $h_n^{(b)} = |\mathbb{T}_n|^{-1/10}$ (*i.e.* $\delta = 1/2$). The numerical study of $\bar{f}_{0,n}$ and $\bar{f}_{1,n}$ leads to similar results as those of the previous study. For a given grid $\{x_1, \ldots, x_k\}$ of $\mathcal{D} = [-3, 3]$, we reject the null hypothesis $\mathcal{H}_0$ if $W_n(x_1, \ldots, x_k)$ exceeds the 5%-quantile of the chi-squared distribution with k degrees of freedom and thus obtain a test with asymptotic level 5%. We measure the quality of our test procedure computing the proportion of rejections of the null.

We first implement the two following cases:

$$(\textbf{Case} =) \quad f_0(x) = f_1(x) = x(1/4 + \exp(-x^2)/2),$$

$$(\textbf{Case} \neq) \quad f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_1(x) = x(1/8 + \exp(-x^2)/2).$$

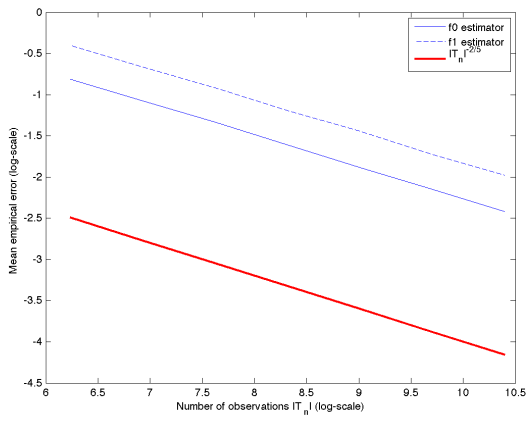

FIGURE 2. *The log-average relative empirical error over $M = 500$ Monte-Carlo trees vs.* $\log(|\mathbb{T}_n|)$ *for f_0 (resp. f_1) reconstructed over $\mathcal{D} = [-3, 3]$ with $\widehat{f}_{0,n}$ (solid blue line) (resp. $\widehat{f}_{1,n}$ (dashed blue line)) compared to the expected log-rate (solid red line).*

The test should reject $\mathcal{H}_0$ in the second case but not in the first one. The larger n, the better the test as one can see in Table 2: $\mathcal{H}_0$ is more and more often rejected for (**Case** $\neq$) and less and less often rejected for (**Case** $=$) as n increases, which is what we expect. We also observe the influence of the number of points of the grid which enables us to build the test statistics. Three grids of $\mathcal{D} = [-3, 3]$ are tested with $k = 13, 25$ and 61 points. The larger the number of points, the larger the proportion of rejections of $\mathcal{H}_0$ in both cases. The test we have built shows good properties only for a reasonable number k of points, as one can see in Table 2.

$\boldsymbol{n}$		8	9	10	11	12	13	14
$\lvert\mathbb{T}_{\boldsymbol{n}}\rvert$		511	1 023	2 047	4 095	8 191	16 383	32 767
	$\Delta x = 0.5$	46.8%	67.2%	87.6%	99.0%	100%	100%	100%
Case $\neq$	$\Delta x = 0.25$	59.6%	77.8%	92.8%	99.8%	100%	100%	100%
	$\Delta x = 0.1$	67.8%	85.4%	95.6%	99.8%	100%	100%	100%
	$\Delta x = 0.5$	19.6%	18.6%	18.2%	16.2%	13.4%	14.8%	12.4%
Case $=$	$\Delta x = 0.25$	30.4%	30.0%	29.0%	24.8%	21.4%	19.4%	19.8%
	$\Delta x = 0.1$	42.6%	42.6%	40.4%	39.8%	35%	31.6%	32.2%

TABLE 2. *Proportions of rejections of the null hypothesis $\mathcal{H}_0 : \{\forall l = 1, \ldots, k, f_0(x_l) = f_1(x_l)\}$ for 5% asymptotic level tests over $M = 500$ Monte-Carlo trees. The test is based on the test statistics $W_n(x_1, \ldots, x_k)$ (13) with the grids $\{x_l = -3 + (l-1)\Delta x \leq 3; l \geq 1\}$ for $\Delta x \in \{0.5; 0.25; 0.1\}$.* (**Case** $\neq$)*: the proportions should be high.* (**Case** $=$)*: the proportions should be low.*

The second experiment aims at studying empirically the power of our test. We keep with the same autoregressive function f_0 for cells of type 0 and parametrize the autoregressive function for

cells of type 1 such that it progressively comes closer to f_0:

$$f_0(x) = x(1/4 + \exp(-x^2)/2) \quad \text{and} \quad f_{1,\tau}(x) = x(\tau + \exp(-x^2)/2)$$

for $\tau \in [1/8, 1/4]$. This choice enables us to interpolate between (**Case** $\neq$) and (**Case** $=$). As τ becomes closer to $1/4$, *i.e.* as $f_{1,\tau}$ becomes closer to f_0, we see the decrease of the proportions of rejections of the null in Figure 3. The steeper the decrease is, the better performs our test. The proportion of rejections of $\mathcal{H}_0$ is higher than 40% only for τ up to 0.1875 for a reasonable number of observations ($|\mathbb{T}_n| = 2\ 047$ on the left in Figure 3). On the right in Figure 3, one can see what become the results for a larger number of observations, $|\mathbb{T}_n| = 32\ 767$: the performance is good for τ up to 0.225, *i.e.* closer to the equality case $\tau = 1/4$.

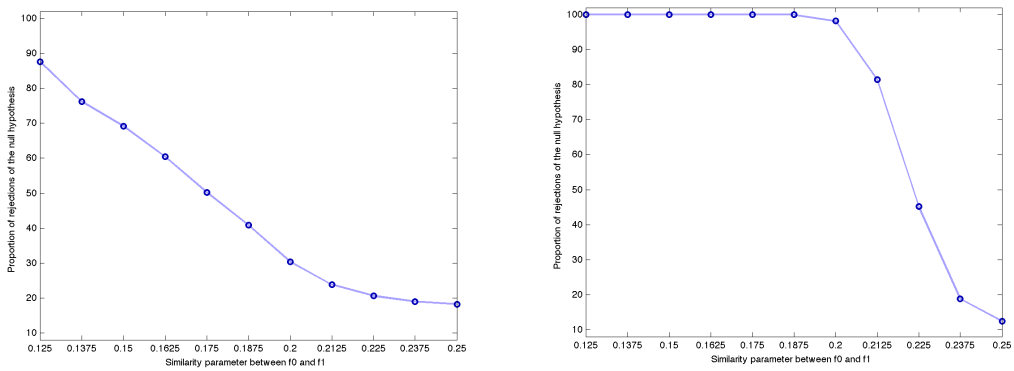

FIGURE 3. *Proportions of rejections of the null hypothesis* $\mathcal{H}_0 : \{\forall l \in \{1, \ldots, k\}, f_0(x_l) = f_{1,\tau}(x_l)\}$ *with respect to* $\tau \in [1/8; 1/4]$ *for 5% asymptotic level tests over* $M = 500$ *Monte-Carlo trees. The test is based on the test statistics* $W_n(x_1, \ldots, x_k)$ (13) *with the grid* $\{x_l = -3 + (l-1)\Delta x \leq 3; l \geq 1\}$ *for* $\Delta x = 0.5$. *Left:* $n = 10$ *generations. Right:* $n = 14$ *generations.*

6. Proofs

The notation $\lesssim$ means up to some constant independent on n and uniform on the class $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$.

For a $\mathfrak{S}$-measurable function $h : \mathcal{S} \to \mathbb{R}$ and a measure μ on $(\mathcal{S}, \mathfrak{S})$ we define $\mu(h) = \int_{\mathcal{S}} h(x)\mu(dx)$. For $K \subset \mathcal{S}$ let

$$|h|_1 = \int_{\mathcal{S}} |h(y)|dy\,, \quad |h|_2^2 = \int_{\mathcal{S}} h(y)^2 dy\,, \quad |h|_K = \sup_{y \in K} |h(y)|$$

and $|h|_\infty = |h|_{\mathcal{S}}$. For a function $h : \mathcal{S}^2 \to \mathbb{R}$ and $K, K' \subset \mathcal{S}$ let

$$|h|_{K,K'} = \sup_{(x,x') \in K \times K'} |h(x, x')|.$$

The following lemma is well-known (see [17], Lemma 2.1 and [27], Equation (7)) and highlights the role of the tagged-branch Markov chain. We prove it in Appendix for the sake of completeness.

Lemma 19 (Many-to-one formulae)**.** *Let $(X_u, u \in \mathbb{T})$ be a bifurcating Markov chain with $\mathbb{T}$-transition $\boldsymbol{\mathcal{P}}$ and let Q be defined by (3). Then for $g : \mathcal{S} \to \mathbb{R}$ such that $|g(x)| \leq 1 + |x|$ for any $x \in \mathcal{S}$, we have*

$$\mathbb{E}_\mu\Big[\sum_{u\in\mathbb{G}_m} g(X_u)\Big] = |\mathbb{G}_m|\mathbb{E}_\mu\big[g(Y_m)\big] = |\mathbb{G}_m|\mu\big(Q^m g\big) \tag{14}$$

and

$$\mathbb{E}_\mu\Big[\sum_{\substack{(u,v)\in\mathbb{G}_m^2\\ u\neq v}} g(X_u)g(X_v)\Big] = |\mathbb{G}_m|\sum_{l=1}^{m} 2^{l-1}\mu\Big(Q^{m-l}\big(\boldsymbol{\mathcal{P}}(Q^{l-1}g\otimes Q^{l-1}g)\big)\Big), \tag{15}$$

for any initial measure $\mu(dx)$ on $\mathcal{S}$ for $X_\emptyset$ such that $\mu\big((1+|\cdot|)^2\big) < \infty$.

6.1. **Preliminary.** Set $\mathbb{V}(x) = 1 + |x|$ for $x \in \mathcal{S}$. It plays the role of the Lyapunov function in the following

Lemma 20 (Ergodicity)**.** *Work under Assumption 4. Let $\gamma \in (0, 1/2)$ and $\ell > 0$, let $r > 0$ and $\lambda > 2$. For every $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$, for every g_ε such that $(g_0, g_1) \in \mathcal{G}(r, \lambda)^2$ satisfy Assumption 5, there exists a unique probability measure ν of the form $\nu(dx) = \nu(x)dx$ on $(\mathcal{S}, \mathfrak{S})$. Moreover, for every g_ε such that $(g_0, g_1) \in \mathcal{G}(r, \lambda)^2$ satisfy Assumption 5, there exists a constant $R > 0$ and $\rho \in (0, 1/2)$ such that*

$$\sup_{(f_0,f_1)}\sup_{|h|\leq\mathbb{V}} |Q^m h(x) - \nu(h)| \leq R\,\mathbb{V}(x)\,\rho^m, \quad x \in \mathcal{S}, \quad m \geq 0,$$

where the supremum is taken among all functions $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ and among all functions $h : \mathcal{S} \to \mathbb{R}$ which satisfy $|h(x)| \leq \mathbb{V}(x)$ for all $x \in \mathcal{S}$.

Proof of Lemma 20. We shall rely on the results of Hairer and Mattingly [28].

Step 1. In order to make use of Theorem 1.2 of [28] we shall verify their Assumptions 1 and 2. Since $Y_1 = f_{\iota_1}(Y_0) + \varepsilon_1'$ with ι_1 drawn according to the Bernoulli distribution with parameter $1/2$ and ε_1' drawn according to $(g_0 + g_1)/2$, we get

$$Q(|\cdot|)(x) = \mathbb{E}_x\big[|Y_1|\big] \leq \mathbb{E}\big[|f_{\iota_1}(x)|\big] + \mathbb{E}[|\varepsilon_1'|] \leq \gamma|x| + M_0$$

using $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$, with $M_0 = \ell + \mathbb{E}[|\varepsilon_1'|]$ as defined previously. We have $\gamma \in (0, 1)$ and $M_0 \geq 0$, so that is Assumption 1 in [28] (with their $V(y) = |y|$).

Set $\mathcal{C} = \{x \in \mathcal{S}; |x| \leq M_1\}$ where M_1 comes from Assumption 5. For any $A \in \mathfrak{S}$ and $x \in \mathcal{C}$, using the expression of Q given by (6),

$$Q(x, A) \geq \frac{1}{2}\int_{A\cap\mathcal{C}} g_0\big(y - f_0(x)\big)dy + \frac{1}{2}\int_{A\cap\mathcal{C}} g_1\big(y - f_1(x)\big)dy.$$

For $(x, y) \in \mathcal{C}^2$, we have $|y - f_\iota(x)| \leq (1 + \gamma)M_1 + \ell$ for $\iota \in \{0, 1\}$. Thus

$$\inf_{x\in\mathcal{C}} Q(x, A) \geq 2M_1\delta\big((1 + \gamma)M_1 + \ell\big)\frac{|A \cap \mathcal{C}|}{|\mathcal{C}|} \quad \forall A \in \mathfrak{S},$$

where $\delta(\cdot)$ is defined by (8) and $|A|$ denotes the Lebesgue measure of $A \in \mathfrak{S}$. That is Assumption 2 in [28] with $\alpha = 2M_1\delta\big((1+\gamma)M_1+\ell\big) > 0$. The existence and uniqueness of an invariant probability measure ν follows from Theorem 1.2 of [28]. Moreover ν is absolutely continuous with respect to the Lebesgue measure, since $Q(x, dy)$ defined by (6) itself is absolutely continuous with respect

to the Lebesgue measure by Assumption 4 on g_ε. By Assumption 5, for M_1 satisfying (9), there exists some $\alpha_0 \in (0, 1/2)$ such that

$$\alpha = 2M_1\delta\big((1+\gamma)M_1 + \ell\big) > 1/2 + \alpha_0. \tag{16}$$

We set $\beta = \alpha_0/M_0$. For all $x \in \mathcal{S}$ we pick $Q^n\delta_x$ and ν for μ_1 and μ_2 in Theorem 1.3 of [28] and apply it recursively. We conclude that for any function h such that $|h(x)| \leq (1+\beta|x|)$ for all $x \in \mathcal{S}$, for some positive constant $C > 0$, we have

$$|Q^m h(x) - \nu(h)| \leq C\rho^m(1+\beta|x|)$$

with $C = 1 + \int_{\mathcal{S}}(1+\beta|x|)\nu(x)dx < \infty$.

Step 2. A precise control of ρ with respect to γ, M_0 and α is established in [28]. Set $\gamma_0 = \gamma + 2M_0/M_1 + \eta \in (\gamma + 2M_0/M_1, 1)$ where η comes from Assumption 5. Theorem 1.3 of [28] states one can take

$$\rho = \big(1 - (\alpha - \alpha_0)\big) \vee \big(\frac{2 + M_1\beta\gamma_0}{2 + M_1\beta}\big).$$

Condition (16) gives immediately $1 - (\alpha - \alpha_0) < 1/2$. Note that

$$\frac{2 + M_1\beta\gamma_0}{2 + M_1\beta} < \frac{1}{2} \quad \Longleftrightarrow \quad M_1 > \frac{(1+2\alpha_0)M_0}{1/2 - \gamma - \eta}$$

which is satisfied for $\alpha_0 \in (0, 1/2)$ with our choice of M_1 (which satisfies (9)). Thus Assumption 5 guarantees we can take $\rho < 1/2$.

It immediately follows from Step 1 and Step 2 that, for any function h such that $|h| \leq \mathbb{V}$, for some positive constant $R > 0$, we have

$$|Q^m h(x) - \nu(h)| \leq R\rho^m \mathbb{V}(x)$$

with $\mathbb{V}(x) = 1 + |x|$, as asserted. This bound holds uniformly over $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ by construction (the uniform choice of ρ is guaranteed by Step 2 and for a uniform choice of C in Step 1 recall that $(g_0, g_1) \in \mathcal{G}(r, \lambda)^2$). □

Step 2 of this proof highlights that Assumption 5 is written to readily obtain $\rho < 1/2$. Note that to prove the existence of some $\rho \in (0, 1)$ in Step 1, one only need the existence of some $M_1 > 2M_0/(1-\gamma)$ such that $2M_1\delta\big((1+\gamma)M_1 + \ell\big) > 0$ with $\delta(\cdot)$ defined by (8).

6.2. **Estimation of the density of the invariant measure.** For $x \in \mathcal{D}$, set

$$\widehat{\nu}_n(x) = \frac{1}{|\mathbb{T}_n|}\sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u), \tag{17}$$

a kernel estimator of the density ν of the invariant measure of the tagged-branch chain Y of transition Q.

Proposition 21. *Work under Assumption 4. Let $\gamma \in (0, 1/2)$ and $\ell > 0$, let $r > 0$ and $\lambda > 3$. Specify $\widehat{\nu}_n$ with a kernel K satisfying Assumption 7 for some $n_0 > 0$ and*

$$h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}.$$

For every $L' > 0$ and $0 < \beta < n_0$, for every g_ε such that $(g_0, g_1) \in \big(\mathcal{G}(r, \lambda) \cap \mathcal{H}^\beta_{\mathcal{S}}(L')\big)^2$ satisfy Assumptions 5 and 6, for every compact interval $\mathcal{D} \subset \mathcal{S}$ with non-empty interior and every x in the interior of $\mathcal{D}$,

$$\sup_{(f_0, f_1)} \mathbb{E}_\mu\Big[\big(\widehat{\nu}_n(x) - \nu(x)\big)^2\Big] \lesssim |\mathbb{T}_n|^{\frac{-2\beta}{2\beta+1}}$$

where the supremum is taken among all functions $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$*, for any initial measure* $\mu(dx)$ *on* $\mathcal{S}$ *for* $X_\emptyset$ *such that* $\mu\big((1+|\cdot|)^2\big) < \infty$*.*

Proof of Proposition 21. The usual bias-variance decomposition can be written here as

$$\begin{aligned}\mathbb{E}\Big[\big(\widehat{\nu}_n(x) - \nu(x)\big)^2\Big] &= \mathbb{E}\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - \nu(x)\big)^2\Big] \\ &= \mathbb{E}\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n}\star\nu(x)\big)^2\Big] + \big(K_{h_n}\star\nu(x) - \nu(x)\big)^2\end{aligned}$$

where $\star$ stands for the convolution. For $(g_0, g_1) \in \mathcal{H}^\beta_{\mathcal{S}}(L')$, we have $\nu \in \mathcal{H}^\beta_{\mathcal{D}}(L'')$ for some $L'' > 0$: since ν is invariant for Q, using (6) which defines Q and Assumption 4, we can write

$$\nu(y) = \int_{\mathcal{S}} \nu(x)Q(x,y)dx = \frac{1}{2}\int_{\mathcal{S}} \nu(x)\Big(g_0\big(y - f_0(x)\big) + g_1\big(y - f_1(x)\big)\Big)dx,$$

where we immediately see that the regularity of ν is inherited from the regularity of g_0 and g_1 the marginals of the noise density g_ε. By a Taylor expansion up to order $\lfloor\beta\rfloor$ (recall that the number n_0 of vanishing moments of K in Assumption 7 satisfies $n_0 > \beta$), we obtain

$$\big(K_{h_n}\star\nu(x) - \nu(x)\big)^2 \lesssim h_n^{2\beta}, \tag{18}$$

see for instance Proposition 1.2 in Tsybakov [46]. In addition, we claim that

$$\mathbb{E}\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n}\star\nu(x)\big)^2\Big] \lesssim (|\mathbb{T}_n|h_n)^{-1}. \tag{19}$$

Choosing $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ brings the announced result. Let us now prove (19) in two steps.

Step 1. Result over one generation. We heavily rely on the following controls:

Lemma 22. *Let* F *be a bounded function with compact support and* G *be a locally bounded function. For* $h_n > 0$ *and* x *in the interior of* $\mathcal{D}$*, we define the function* $H_n : \mathcal{S} \to \mathbb{R}$ *by*

$$H_n(\cdot) = F\big(h_n^{-1}(x - \cdot)\big)G(\cdot).$$

For h_n *such that* $x - h_n\mathrm{supp}(F) \subset \mathcal{D}$*, we have*

(i) $|QH_n|_\infty \le h_n|F|_1|G|_{\mathcal{D}}|Q|_{S,\mathcal{D}}$ *and* $|\nu(H_n)| \le h_n|F|_1|G|_{\mathcal{D}}|\nu|_{\mathcal{D}}$*.*

(ii) *Under Assumption 5, for* $m \ge 0$*,* $y \in \mathcal{S}$*,*

$$|Q^m H_n(y) - \nu(H_n)| \lesssim h_n \wedge \big(\mathbb{V}(y)\rho^m\big)$$

up to the constant $\max\{|F|_1|G|_{\mathcal{D}}\big(|Q|_{\mathcal{S},\mathcal{D}} + |\nu|_{\mathcal{D}}\big), R|F|_\infty|G|_{\mathcal{D}}\}$*.*

The proof of this lemma is postponed to the Appendix. Note that we have $|\nu|_{\mathcal{D}} \le |Q|_{\mathcal{S},\mathcal{D}} \le |Q|_{\mathcal{S},\mathcal{S}} \le (|g_0|_\infty + |g_1|_\infty)/2 < \infty$ (recall that ν is invariant for Q defined by (6), we work under Assumption 4 and $(g_0, g_1) \in \mathcal{G}(r,\lambda)^2$). Set

$$H_n(\cdot) = K\big(h_n^{-1}(x - \cdot)\big) \quad \text{and} \quad \widetilde{H}_n(\cdot) = H_n(\cdot) - \nu(H_n),$$

with h_n sufficiently small such that $x - h_n\mathrm{supp}(K) \subset \mathcal{D}$. On the one hand,

$$\mathbb{E}_\mu\Big[\sum_{u\in\mathbb{G}_m} \widetilde{H}_n^2(X_u)\Big] = |\mathbb{G}_m|\mu\big(Q^m\widetilde{H}_n^2\big) \lesssim |\mathbb{G}_m|h_n$$

relying on the many-to-one formula (14) and Lemma 22(i). Inspired by [24] (proof of Proposition 8), set $l^* = \lfloor|\log h_n|/|\log\rho|\rfloor$. Since $\mu\big(Q^{m-l}\big(\boldsymbol{\mathcal{P}}(\mathbb{V}\otimes\mathbb{V})\big)\big) < \infty$ (use $\mathbb{V}(x) = 1 + |x|$ for $x \in \mathcal{S}$,

$(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$ and finally $\mu(\mathbb{V}^2) < \infty$, one can look at Lemmae 25 and 26 of [27]), by the many-to-one formula (15),

$$\mathbb{E}_\mu\Big[\sum_{u\neq v\in\mathbb{G}_m} \widetilde{H}_n(X_u)\widetilde{H}_n(X_v)\Big] \lesssim |\mathbb{G}_m|\Big(\sum_{l=1}^{l^*} 2^{l-1}h_n^2 + \sum_{l=l^*+1}^{m} 2^{l-1}\rho^{2(l-1)}\Big) \lesssim 2^m h_n\,,$$

using the first upper-bound given by Lemma 22(ii) before l^*, the second one after l^*, and using $\rho \in (0, 1/2)$. Finally, we conclude that

$$\mathbb{E}_\mu\Big[\big(\sum_{u\in\mathbb{G}_m} \widetilde{H}_n(X_u)\big)^2\Big] \lesssim |\mathbb{G}_m|h_n.$$

Thus, we have uniformly over $(f_0, f_1) \in \mathcal{F}(\gamma, \ell)^2$,

$$\mathbb{E}_\mu\Big[\big(\frac{1}{|\mathbb{G}_m|}\sum_{u\in\mathbb{G}_m} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\big)^2\Big] \lesssim (|\mathbb{G}_m h_n|)^{-1}. \tag{20}$$

Step 2. Result over a subtree. We rely on the previous inequality (20). Decomposing by generation and by the triangular inequality, we obtain

$$\begin{aligned}
&\mathbb{E}_\mu\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\big)^2\Big] \\
&\qquad \le \Big(\sum_{m=0}^{n} \frac{|\mathbb{G}_m|}{|\mathbb{T}_n|}\Big(\mathbb{E}_\mu\Big[\big(\frac{1}{|\mathbb{G}_m}\sum_{u\in\mathbb{G}_m} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\big)^2\Big]\Big)^{1/2}\Big)^2 \\
&\qquad \lesssim \Big(\sum_{m=0}^{n} \frac{|\mathbb{G}_m|}{|\mathbb{T}_n|}(|\mathbb{G}_m|h_n)^{-1/2}\Big)^2 \lesssim (|\mathbb{T}_n|h_n)^{-1}.
\end{aligned}$$

This proves inequality (19) we claimed and the proof is now complete. Note that we have removed the log-term which appears in Proposition 8 of [24]. □

Proposition 23. *In the same setting as in Proposition 21,*

$$\widehat{\nu}_n(x) \to \nu(x), \quad \mathbb{P}_\mu - a.s.$$

as $n \to \infty$.

Proof of Proposition 23. Write

$$\widehat{\nu}_n(x) - \nu(x) = \big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\big) + \big(K_{h_n} \star \nu(x) - \nu(x)\big).$$

From (18), we deduce that

$$|K_{h_n} \star \nu(x) - \nu(x)| \to 0 \text{ as } n \to \infty. \tag{21}$$

Note that we could obtain (21) invoking the result stated by Theorem 2.1.1 of [42], result also known as the Bochner Lemma (see section 7.1.2 of [25]). Using (19) for $h_n = |\mathbb{T}_n|^{-\alpha}$ with $\alpha \in (0, 1)$,

$$\sum_{n\ge 0} \mathbb{E}_\mu\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\big)^2\Big] < \infty,$$

and by the Borel-Cantelli lemma, we deduce that

$$\Big|\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u) - K_{h_n} \star \nu(x)\Big| \to 0, \quad \mathbb{P}_\mu - a.s.$$

as $n \to \infty$. Thus $|\widehat{\nu}_n(x) - \nu(x)| \to 0$, $\mathbb{P}_\mu - a.s.$ as $n \to \infty$. □

6.3. **Proof of Theorem 8.** For x in the interior of $\mathcal{D}$, for $\iota \in \{0,1\}$, we plan to use the decomposition

$$\begin{aligned}\widehat{f}_{\iota,n}(x) - f_\iota(x) &= \frac{M_{\iota,n}(x) + L_{\iota,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n} - \frac{(\nu f_\iota)(x)}{\nu(x)} \\ &= \frac{M_{\iota,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n} + \frac{L_{\iota,n}(x) - (\nu f_\iota)(x)}{\widehat{\nu}_n(x) \vee \varpi_n} - \frac{\widehat{\nu}_n(x) \vee \varpi_n - \nu(x)}{\widehat{\nu}_n(x) \vee \varpi_n} f_\iota(x)\end{aligned}$$

with

$$M_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u)\varepsilon_{u\iota}, \tag{22}$$

$$L_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u). \tag{23}$$

Thus

$$\begin{aligned}\mathbb{E}_\mu\Big[\big(\widehat{f}_{\iota,n}(x) - f_\iota(x)\big)^2\Big] \lesssim \varpi_n^{-2}\Big(&\mathbb{E}_\mu\Big[\big(M_{\iota,n}(x)\big)^2\Big] + \mathbb{E}_\mu\Big[\big(L_{\iota,n}(x) - (\nu f_\iota)(x)\big)^2\Big] \\ &+ \mathbb{E}_\mu\Big[\big(\widehat{\nu}_n(x) \vee \varpi_n - \nu(x)\big)^2\Big]\Big)\end{aligned} \tag{24}$$

using $|f_\iota|_{\mathcal{D}} < \infty$, uniformly over the class $\mathcal{F}(\gamma, \ell)$ for $\mathcal{D}$ compact interval. We successively treat the three terms in Steps from 1 to 3.

Step 1. Term $M_{\iota,n}(x)$. For all $m \geq 0$ and $\iota \in \{0,1\}$ fixed, the sequence $(\varepsilon_{u\iota}, u \in \mathbb{G}_m)$ is a family independent random variables such that $\mathbb{E}[\varepsilon_{u\iota}^2] = \sigma_\iota^2$. Thus

$$\begin{aligned}\mathbb{E}_\mu\Big[\Big(\frac{1}{|\mathbb{G}_m|}\sum_{u \in \mathbb{G}_m} K_{h_n}(x - X_u)\varepsilon_{u\iota}\Big)^2\Big] &= \frac{\sigma_\iota^2}{|\mathbb{G}_m| h_n^2}\mathbb{E}_\mu\Big[\frac{1}{|\mathbb{G}_m|}\sum_{u \in \mathbb{G}_m} K^2\big(h_n^{-1}(x - X_u)\big)\Big] \\ &= \frac{\sigma_\iota^2}{|\mathbb{G}_m| h_n^2}\mu\Big(K^2\big(h_n^{-1}(x - \cdot)\big)\Big) \lesssim (|\mathbb{G}_m| h_n)^{-1}\end{aligned}$$

by the many-to-one formula (14) and using Lemma 22(i). The result over a subtree follows by the triangular inequality (as in Step 2 of the proof of Proposition 21),

$$\mathbb{E}_\mu\Big[\big(M_{\iota,n}(x)\big)^2\Big] \lesssim (|\mathbb{T}_n| h_n)^{-1}. \tag{25}$$

Step 2. Term $L_{\iota,n}(x)$. By usual the bias-variance decomposition,

$$\begin{aligned}\mathbb{E}_\mu\Big[\big(L_{\iota,n}(x) - (\nu f_\iota)(x)\big)^2\Big] = \mathbb{E}_\mu\Big[\Big(\frac{1}{|\mathbb{T}_n|}\sum_{u \in \mathbb{T}_n} K_{h_n}(x - X_u) f_\iota(X_u) - K_{h_n} \star (\nu f_\iota)(x)\Big)^2\Big] \\ + \big(K_{h_n} \star (\nu f_\iota)(x) - (\nu f_\iota)(x)\big)^2.\end{aligned}$$

First, since $(\nu f_\iota) \in \mathcal{H}_{\mathcal{D}}^\beta(L'')$ for some constant $L'' > 0$ and since Assumption 7 is valid with $n_0 > \beta$,

$$\big(K_{h_n} \star (\nu f_\iota)(x) - (\nu f_\iota)(x)\big)^2 \lesssim h_n^{2\beta}.$$

Secondly, we do the same study as in the proof of Proposition 21 for

$$H_n(\cdot) = K\big(h_n^{-1}(x - \cdot)\big) f_\iota(\cdot),$$

relying on Lemma 22 (using $|f_\iota|_{\mathcal{D}} < \infty$ uniformly over the class $\mathcal{F}(\gamma,\ell)$, with $\mathcal{D}$ compact interval), with h_n sufficiently small such that $x - h_n\mathrm{supp}(K) \subset \mathcal{D}$. We obtain

$$\mathbb{E}_\mu\Big[\big(\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x-X_u)f_\iota(X_u) - K_{h_n}\star(\nu f_\iota)(x)\big)^2\Big] \lesssim (|\mathbb{T}_n|h_n)^{-1}.$$

Thus

$$\mathbb{E}_\mu\Big[\big(L_{\iota,n}(x) - (\nu f_\iota)(x)\big)^2\Big] \lesssim h_n^{2\beta} + (|\mathbb{T}_n|h_n)^{-1}. \tag{26}$$

Step 3. Denominator $\widehat{\nu}_n(x)\vee\varpi_n$. We prove the following lemma in Appendix.

Lemma 24. *Work under Assumption 4. Let $\gamma\in(0,1/2)$ and $\ell>0$, let $r>0$ and $\lambda>3$. For every g_ε such that $(g_0,g_1)\in\mathcal{G}(r,\lambda)^2$ satisfy Assumptions 5 and 6, there exists $d=d(\gamma,\ell,g_0,g_1)>0$ such that for every $\mathcal{D}\subset[-d,d]$,*

$$\inf_{(f_0,f_1)}\inf_{x\in\mathcal{D}}\nu(x) > 0$$

where the infimum is taken among all functions $(f_0,f_1)\in\mathcal{F}(\gamma,\ell)^2$.

Relying first on Lemma 24, we choose n large enough such that

$$0 < \varpi_n \le \frac{1}{2}\inf_{(f_0,f_1)}\inf_{x\in\mathcal{D}}\nu(x), \tag{27}$$

and we get

$$\mathbb{E}_\mu\Big[\big(\widehat{\nu}_n(x)\vee\varpi_n - \nu(x)\big)^2\Big] \lesssim \mathbb{E}_\mu\Big[\big(\widehat{\nu}_n(x)-\nu(x)\big)^2\Big] \lesssim h_n^{2\beta} + (|\mathbb{T}_n|h_n)^{-1}, \tag{28}$$

uniformly over $(f_0,f_1)\in\mathcal{F}(\gamma,\ell)^2$, using the upper-bound obtained in the proof of Proposition 21 for the second inequality.

Finally, gathering (25), (26) and (28) in (24) and choosing $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$ we obtain the asserted result.

6.4. **Proof of Theorem 9.** In the following, $\mathbb{P}^n_{(f_0,f_1)}$ will denote the law on $\mathcal{S}^{|\mathbb{T}_{n+1}|}$ of the vector $(X_u, u\in\mathbb{T}_{n+1})$, NBAR process, in the sense of Definition 1, driven by the autoregressive functions f_0 and f_1 with initial probability measure $\mu(dx)$ on $\mathcal{S}$ for $X_\emptyset$ and with a Gaussian noise *i.e.*

$$g_\varepsilon(x,y) = \big(2\pi(\sigma_0^2\sigma_1^2(1-\rho^2)\big)^{-1/2}\exp\Big(-\frac{\sigma_1^2x^2 - 2\sigma_0\sigma_1\rho xy + \sigma_0^2y^2}{2\big(\sigma_0^2\sigma_1^2(1-\rho^2)\big)}\Big), \quad (x,y)\in\mathcal{S}^2,$$

with $\sigma_0,\sigma_1>0$ and $\rho\in(-1,1)$. When $f_0=f_1=f$, we shorten $\mathbb{P}^n_{(f_0,f_1)}$ into $\mathbb{P}^n_f$. We denote by $\mathbb{E}^n_f[\cdot]$ the expectation with respect to $\mathbb{P}^n_f$.

Step 1. Let $\delta>0$. Fix $f_0=f_1=f^*$ with $f^*\in\mathcal{F}(\gamma,\ell)\cap\mathcal{H}^\beta_{\mathcal{D}}(L-\delta)$ and $x\in\mathcal{D}$. Then, for large enough n, setting $h_n = |\mathbb{T}_n|^{-1/(2\beta+1)}$, we construct a perturbation $(f_{0,n},f_{1,n})$ of (f_0,f_1) defined by

$$f_{0,n}(y) = f_{1,n}(y) = f^*_n(y) = f^*(y) + ah_n^\beta K\big(h_n^{-1}(x-y)\big), \quad y\in\mathcal{S},$$

for some smooth kernel K with compact support such that $K(0)=1$, and for some $a=a_{\delta,K}>0$ chosen in such a way that $f_n\in\mathcal{F}(\gamma,\ell)\cap\mathcal{H}^\beta_{\mathcal{D}}(L)$. Note that at point $y=x$, $|f_{0,n}(x)-f_0(x)| = |f_{1,n}(x)-f_1(x)| = a_{\delta,K}h_n^\beta = a_{\delta,K}|\mathbb{T}_n|^{-\beta/(2\beta+1)}$.

Step 2. In the sequel, to shorten expressions, we set

$$\big|(\widehat{f}_{0,n},\widehat{f}_{1,n}) - f\big| = \big|\widehat{f}_{0,n}(x)-f(x)\big| + \big|\widehat{f}_{1,n}(x)-f(x)\big|.$$

For arbitrary estimators $(\widehat{f}_{0,n}(x), \widehat{f}_{1,n}(x))$ and a constant $C > 0$, the maximal risk is bounded below by

$$\begin{aligned}
&\max_{f\in\{f^*;f_n^*\}} \mathbb{P}_f^n\Big(|\mathbb{T}_n|^{\beta/(2\beta+1)}\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f\big|\geq C\Big)\\
\geq&\frac{1}{2}\Big(\mathbb{P}_{f^*}^n\big(|\mathbb{T}_n|^{\beta/(2\beta+1)}\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f^*\big|\geq C\big)+\mathbb{P}_{f_n^*}^n\big(|\mathbb{T}_n|^{\beta/(2\beta+1)}\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f_n^*\big|\geq C\big)\Big)\\
\geq&\frac{1}{2}\mathbb{E}_{f^*}^n\Big[\mathbf{1}_{\{|\mathbb{T}_n|^{\beta/(2\beta+1)}|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f^*|\geq C\}}+\mathbf{1}_{\{|\mathbb{T}_n|^{\beta/(2\beta+1)}|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f_n^*|\geq C\}}\Big]-\frac{1}{2}\|\mathbb{P}_{f^*}^n-\mathbb{P}_{f_n^*}^n\|_{TV}.
\end{aligned}$$

By the triangle inequality, we have

$$|\mathbb{T}_n|^{\beta/(2\beta+1)}\Big(\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f^*\big|+\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f_n^*\big|\Big)\geq 2|\mathbb{T}_n|^{\beta/(2\beta+1)}\big|f_n^*(x)-f^*(x)\big|=2a_{\delta,K}$$

by Step 1, so if we now take $C < a_{\delta,K}/4$, one of the two indicators within the expectation above must be equal to one with full $\mathbb{P}_{f^*}^n$-probability. In that case,

$$\max_{f\in\{f^*;f_n^*\}} \mathbb{P}_f^n\Big(|\mathbb{T}_n|^{\beta/(2\beta+1)}\big|(\widehat{f}_{0,n},\widehat{f}_{1,n})-f\big|\geq C\Big)\geq\frac{1}{2}\big(1-\|\mathbb{P}_{f^*}^n-\mathbb{P}_{f_n^*}^n\|_{TV}\big)$$

and Theorem 9 is thus proved if $\limsup_{n\to\infty}\|\mathbb{P}_{f^*}^n-\mathbb{P}_{f_n^*}^n\|_{TV}<1$.

Step 3. By the Pinsker inequality, we have $\|\mathbb{P}_{f^*}^n-\mathbb{P}_{f_n^*}^n\|_{TV}\leq\frac{\sqrt{2}}{2}\Big(\mathbb{E}_{f^*}^n\Big[\ln\frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n}\Big]\Big)^{1/2}$ and the log-likelihood ratio can be written

$$\begin{aligned}
\mathbb{E}_{f^*}^n\Big[\ln\Big(\frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n}\Big)\Big]&=\mathbb{E}_{f^*}^n\Big[\sum_{u\in\mathbb{T}_n}\ln\Big(\frac{g_\varepsilon\big(X_{u0}-f^*(X_u),X_{u1}-f^*(X_u)\big)}{g_\varepsilon\big(X_{u0}-f_n^*(X_u),X_{u1}-f_n^*(X_u)\big)}\Big)\Big]\\
&=-\sum_{u\in\mathbb{T}_n}\mathbb{E}_{f^*}^n\Big[\frac{(\sigma_1^2-\sigma_0\sigma_1\rho)\varepsilon_{u0}+(\sigma_0^2-\sigma_0\sigma_1\rho)\varepsilon_{u1}}{\sigma_0^2\sigma_1^2(1-\rho^2)}(f_n^*-f^*)(X_u)\Big]\\
&\qquad+\frac{\sigma_0^2+\sigma_1^2-2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)}\sum_{u\in\mathbb{T}_n}\mathbb{E}_{f^*}^n\Big[(f_n^*-f^*)^2(X_u)\Big],
\end{aligned}$$

since g_ε is chosen to be the bivariate Gaussian density and, under $\mathbb{P}_{f^*}^n$, we know $X_{u0}=f^*(X_u)+\varepsilon_{u0}$ and $X_{u1}=f^*(X_u)+\varepsilon_{u1}$. Recall now that X_u is independent on $(\varepsilon_{u0},\varepsilon_{u1})$ which is centered. Thus

$$\begin{aligned}
\mathbb{E}_{f^*}^n\Big[\ln\Big(\frac{d\mathbb{P}_{f^*}^n}{d\mathbb{P}_{f_n^*}^n}\Big)\Big]&=\frac{\sigma_0^2+\sigma_1^2-2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)}\sum_{u\in\mathbb{T}_n}\mathbb{E}_{f^*}^n\Big[(f^*-f_n^*)^2(X_u)\Big]\\
&=\frac{\sigma_0^2+\sigma_1^2-2\sigma_0\sigma_1\rho}{2\sigma_0^2\sigma_1^2(1-\rho^2)}\sum_{m=0}^{n}|\mathbb{G}_m|\mu\Big(Q_{f^*}^m\big((f^*-f_n^*)^2\big)\Big)
\end{aligned}\tag{29}$$

using the many-to-one formula (14), with

$$Q_{f^*}(x,y)=\frac{1}{2}\Big(g_0\big(y-f^*(x)\big)+g_1\big(y-f^*(x)\big)\Big),$$

where g_0 and g_1 are the marginals of g_ε. Since

$$Q_{f^*}\big((f^*-f_n^*)^2\big)(y)=a_{\delta,K}^2h_n^{2\beta}\int_{\mathcal{S}}K^2\big(h_n^{-1}(x-z)\big)Q_{f^*}(y,z)dz\leq a_{\delta,K}^2|K|_2^2|Q_{f^*}|_{\mathcal{S},\mathcal{S}}h_n^{2\beta+1}$$

where $|Q_{f^*}|_{\mathcal{S},\mathcal{S}}$ only depends on σ_0, σ_1 and ρ. Gathering this last upper-bound together with (29) and the Pinsker inequality, we finally get, with our choice of h_n,

$$\|\mathbb{P}^n_{f^*} - \mathbb{P}^n_{f^*_n}\|_{TV} \lesssim a^2_{\delta,K}$$

and this term can be made arbitrarily small by picking $a_{K,\delta}$ small enough.

6.5. **Proof of Proposition 10.** For the choice $h_n = |\mathbb{T}_n|^{-\alpha}$ with $\alpha \in (0,1)$, relying successively on (25) and (26), we deduce

$$\sum_{n\geq 0} \mathbb{E}_\mu\Big[\big(M_{\iota,n}(x)\big)^2\Big] < \infty \quad \text{and} \quad \sum_{n\geq 0} \mathbb{E}_\mu\Big[\big(L_{\iota,n}(x) - (\nu f_\iota)(x)\big)^2\Big] < \infty.$$

Thus, as $n \to \infty$,

$$M_{\iota,n}(x) \to 0, \quad \mathbb{P}_\mu - a.s. \quad \text{and} \quad L_{\iota,n}(x) \to (\nu f_\iota)(x), \quad \mathbb{P}_\mu - a.s.$$

From Proposition 23 and since $\varpi_n \to 0$,

$$\widehat{\nu}_n(x) \vee \varpi_n \to \nu(x), \quad \mathbb{P}_\mu - a.s.$$

We conclude reminding that

$$\widehat{f}_{\iota,n}(x) = \frac{M_{\iota,n}(x) + L_{\iota,n}(x)}{\widehat{\nu}_n(x) \vee \varpi_n}.$$

6.6. **Proofs of Theorems 11 and 12.**

Proof of Theorem 11. Set x in the interior of $\mathcal{D}$. The strategy is to use the following decomposition, which is slightly different from the one used in the proof of Theorem 8,

$$\sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}\widehat{f}_{0,n}(x) - f_0(x)\\ \widehat{f}_{1,n}(x) - f_1(x)\end{pmatrix} = \tfrac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\Bigg\{\sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}M_{0,n}(x)\\ M_{1,n}(x)\end{pmatrix} + \sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}N_{0,n}(x)\\ N_{1,n}(x)\end{pmatrix} + \sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}R_{0,n}(x)\\ R_{1,n}(x)\end{pmatrix}\Bigg\}$$

where, for $\iota \in \{0,1\}$, $M_{\iota,n}(x)$ is defined by (22),

$$N_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} K_{h_n}(x - X_u)\big(f_\iota(X_u) - f_\iota(x)\big), \tag{30}$$

$$R_{\iota,n}(x) = \Big(\widehat{\nu}_n(x) - \big(\widehat{\nu}_n(x) \vee \varpi_n\big)\Big) f_\iota(x), \tag{31}$$

and $\widehat{\nu}_n(x)$ is defined by (17). The first part of the decomposition is called main term, the second part negligible term and the third part is a remainder term due to the truncation of the denominator of the estimators. The strategy is the following: prove first that the last two terms goes to zero almost surely and prove a central limit theorem for the main term in a second step.

Step 1. Negligible and remainder terms, $N_{\iota,n}(x)$ and $R_{\iota,n}(x)$. We use the decomposition $N_{\iota,n}(x) = N^{(1)}_{\iota,n}(x) + N^{(2)}_{\iota,n}(x)$ where

$$N^{(1)}_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} \mathbb{E}_\nu\big[K_{h_n}(x - X_u)\big(f_\iota(X_u) - f_\iota(x)\big)\big] \tag{32}$$

$$N^{(2)}_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n} \Big(K_{h_n}(x - X_u)\big(f_\iota(X_u) - f_\iota(x)\big) - \mathbb{E}_\nu\big[K_{h_n}(x - X_u)\big(f_\iota(X_u) - f_\iota(x)\big]\Big), \tag{33}$$

We claim that

$$\sqrt{|\mathbb{T}_n|h_n}\, N^{(1)}_{\iota,n}(x) \longrightarrow 0 \quad \text{and} \quad \sqrt{|\mathbb{T}_n|h_n}\, N^{(2)}_{\iota,n}(x) \to 0, \quad \mathbb{P}_\mu - a.s.$$

as $n \to \infty$.

Step 1.1. Set

$$H_n(\cdot) = K\big(h_n^{-1}(x-\cdot)\big)\big(f_\iota(\cdot) - f_\iota(x)\big) \tag{34}$$

with h_n sufficiently small such that $x - h_n \mathrm{supp}(K) \subset \mathcal{D}$. By the many-to-one formula (14), after a decomposition of the subtree $\mathbb{T}_n$ in $\cup_{m=0}^n \mathbb{G}_m$,

$$N^{(1)}_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{m=0}^{n} |\mathbb{G}_m|\, \mathbb{E}_\nu\big[K_{h_n}(x-Y_m)\big(f_\iota(Y_m) - f_\iota(x)\big)\big] = h_n^{-1}\nu(H_n) \tag{35}$$

since ν is the invariant measure of the tagged-branch chain $(Y_m, m \geq 0)$ and

$$\begin{aligned}
\nu(H_n) &= \int_{\mathcal{S}} K\big(h_n^{-1}(x-y)\big)\big(f_\iota(y) - f_\iota(x)\big)\nu(y)dy \\
&= h_n \int_{\mathcal{S}} K(y)\big(f_\iota(x-h_n y) - f_\iota(x)\big)\nu(x-h_n y)dy \\
&= h_n \int_{\mathcal{S}} K(y)\Big(\big((\nu f_\iota)(x-h_n y) - (\nu f_\iota)(x)\big) - \big(\nu(x-h_n y) - \nu(x)\big) f_\iota(x)\Big)dy.
\end{aligned}$$

We now use that both (νf_ι) and f_ι have derivatives up to order $\lfloor\beta\rfloor$. Also remind that K is of order $n_0 > \beta$. By a Taylor expansion, for some ϑ and $\vartheta' \in (0,1)$,

$$\begin{aligned}
\nu(H_n) &= h_n \int_{\mathcal{S}} K(y)\Big(\frac{(-h_n y)^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}(\nu f_\iota)^{\lfloor\beta\rfloor}(x-\vartheta h_n y) - \frac{(-h_n y)^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}\nu^{\lfloor\beta\rfloor}(x-\vartheta' h_n y) f_\iota(x)\Big)dy \\
&= h_n \int_{\mathcal{S}} K(y)\frac{(-h_n y)^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}\Big(\big((\nu f_\iota)^{\lfloor\beta\rfloor}(x-\vartheta h_n y) - (\nu f_\iota)^{\lfloor\beta\rfloor}(x)\big) \\
&\qquad\qquad - \big(\nu^{\lfloor\beta\rfloor}(x-\vartheta' h_n y) - \nu^{\lfloor\beta\rfloor}(x)\big) f_\iota(x)\Big)dy.
\end{aligned} \tag{36}$$

Thus, using $(\nu f_\iota) \in \mathcal{H}^\beta_{\mathcal{D}}(L'')$ and $\nu \in \mathcal{H}^\beta_{\mathcal{D}}(L'')$ for some $L'' > 0$,

$$|\nu(H_n)| \leq h_n \int_{\mathcal{S}} |K(y)|\frac{|h_n y|^{\lfloor\beta\rfloor}}{\lfloor\beta\rfloor!}\Big(\big(L''|\vartheta h_n y|^{\{\beta\}}\big) + \big(L''|\vartheta' h_n y|^{\{\beta\}}\big) f_\iota(x)\Big)dy \lesssim h_n^{1+\beta}.$$

Hence, recalling (35), $N^{(1)}_{\iota,n}(x) \lesssim h_n^\beta$ and $\sqrt{|\mathbb{T}_n|h_n}\, N^{(1)}_{\iota,n}(x)$ goes to zero when n goes to infinity choosing $h_n = |\mathbb{T}_n|^{-\alpha}$ with $\alpha > 1/(1+2\beta)$.

Step 1.2. In the same way we have proved $|\nu(H_n)| \lesssim h_n^{1+\beta}$, we prove $|QH_n(y)| \lesssim h_n^{1+\beta}$ using the fact that $z \rightsquigarrow f_\iota(z)Q(y,z)$ and $z \rightsquigarrow Q(y,z)$ belong to $\mathcal{H}^\beta_{\mathcal{D}}(L'')$ for some other $L'' > 0$ for any fixed $y \in \mathcal{S}$. This enables us to reinforce the inequality of Lemma 22(i) and using Lemma 22(ii) we obtain

$$|Q^l H_n(y) - \nu(H_n)| \lesssim h_n^{1+\beta} \wedge \big(\mathbb{V}(y)\rho^l\big), \quad l \geq 1.$$

It brings the following upper-bound using the same technique as in Step 1 and Step 2 of Proposition 21:

$$\mathbb{E}_\mu\big[\big(N^{(2)}_{\iota,n}(x)\big)^2\big] \lesssim h_n^\beta\big(|\mathbb{T}_n|h_n\big)^{-1},$$

which, by the Borel-Cantelli lemma, leads to the $\mathbb{P}_\mu$- almost sure convergence of $\sqrt{|\mathbb{T}_n|h_n}N^{(2)}_{\iota,n}(x)$ to zero, choosing $h_n = |\mathbb{T}_n|^{-\alpha}$ with $\alpha > 0$.

Step 1.3. To end the first step of the proof, we prove that the remainder term is such that

$$\sqrt{|\mathbb{T}_n|h_n}R_{\iota,n}(x) \to 0, \quad \mathbb{P}_\mu - a.s. \tag{37}$$

as $n \to \infty$. Write

$$R_{\iota,n}(x) = \big(\widehat{\nu}_n(x) - \varpi_n\big) f_\iota(x) \mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}} \tag{38}$$

where $\big(\widehat{\nu}_n(x) - \varpi_n\big)$ converges $\mathbb{P}_\mu$-almost surely to $\nu(x)$. We can easily prove that $\mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}}$ converges $\mathbb{P}_\mu$-almost surely to 0, since $\widehat{\nu}_n(x)$ converges $\mathbb{P}_\mu$-almost surely to $\nu(x)$ and $\varpi_n \to 0$, which means that $\mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}}$ is null $\mathbb{P}_\mu$-almost surely beyond some integer. So

$$\sqrt{|\mathbb{T}_n|h_n}\mathbf{1}_{\{\widehat{\nu}_n(x) < \varpi_n\}} = 0, \quad \mathbb{P}_\mu - a.s.$$

beyond some integer and (37) is thus proved.

Step 2. Main term $M_{\iota,n}(x)$. We will make use of the central limit theorem for martingale triangular arrays (see for instance [25], Theorem 2.1.9, page 46). We follow [17] (section 4) in order to define the notion of the n first individuals of $\mathbb{T}$. Let $\big(\Pi^*_m, m \geq 1\big)$ be independent random variables, where for each m, Π^*_m is uniformly distributed over the set of permutations of $\mathbb{G}_m$. The collection $\big(\Pi^*_m(1), \ldots, \Pi^*_m(|\mathbb{G}_m|)\big)$ is a random drawing without replacement of all the elements of $\mathbb{G}_m$. For $k \geq 1$, set $\rho_k = \inf\{k' \geq 0, k \leq |\mathbb{T}_{k'}|\}$ (it can be seen as the number of generation to which belongs the k-th element of $\mathbb{T}$). We finally define a random order on $\mathbb{T}$ through $\widetilde{\Pi}$ the function from $\{1, 2, \ldots\}$ to $\mathbb{T}$ such that $\widetilde{\Pi}(1) = \emptyset$ and for $k \geq 2$, $\widetilde{\Pi}(k) = \Pi^*_{\rho_k}(k - |\mathbb{T}_{\rho_k - 1}|)$. We introduce the filtration $\mathcal{G} = (\mathcal{G}_n, n \geq 0)$ defined by $\mathcal{G}_0 = \sigma(X_\emptyset)$ and for each $n \geq 1$,

$$\mathcal{G}_n = \sigma\Big(\big((X_{\widetilde{\Pi}(k)}, X_{(\widetilde{\Pi}(k),0)}, X_{(\widetilde{\Pi}(k),1)}), 1 \leq k \leq n\big), (\widetilde{\Pi}(k), 1 \leq k \leq n+1)\Big).$$

For $n \geq 1$, we consider the vector of bivariate random variables $\overline{E}^{(n)}(x) = (\overline{E}^{(n)}_k(x), 1 \leq k \leq |\mathbb{T}_n|)$ defined by

$$\overline{E}^{(n)}_k = \sum_{l=1}^{k} E^{(n)}_l(x) \quad \text{with} \quad E^{(n)}_l(x) = (|\mathbb{T}_n|h_n)^{-1/2} \begin{pmatrix} K\big(h_n^{-1}(x - X_{\widetilde{\Pi}(l)})\big)\varepsilon_{(\widetilde{\Pi}(l),0)} \\ K\big(h_n^{-1}(x - X_{\widetilde{\Pi}(l)})\big)\varepsilon_{(\widetilde{\Pi}(l),1)} \end{pmatrix}. \tag{39}$$

Notice that $\overline{E}^{(n)}(x)$ is a square-integrable martingale adapted to $\mathcal{G} = (\mathcal{G}_n, n \geq 0)$. Then, $(\overline{E}^{(n)}(x), n \geq 1)$ is a square-integrable $\mathcal{G}$-martingale triangular array whose bracket is given by

$$\langle \overline{E}^{(n)}(x) \rangle_k = \sum_{l=1}^{k} \mathbb{E}\Big[E^{(n)}_l(x)\big(E^{(n)}_l(x)\big)^t \Big| \mathcal{G}_{l-1}\Big] = \Big(\frac{1}{|\mathbb{T}_n|h_n} \sum_{l=1}^{k} K^2\big(h_n^{-1}(x - X_{\widetilde{\Pi}(l)})\big)\Big)\Gamma$$

where Γ is the noise covariance matrix. We apply Proposition 23 (with $|K|_2^{-2}K^2$ replacing K as kernel function) and we obtain

$$\langle \overline{E}^{(n)}(x) \rangle_{|\mathbb{T}_n|} = \Big(\frac{1}{|\mathbb{T}_n|h_n} \sum_{u \in \mathbb{T}_n} K^2\big(h_n^{-1}(x - X_u)\big)\Big)\Gamma \to |K|_2^2 \nu(x)\,\Gamma, \quad \mathbb{P}_\mu - a.s. \tag{40}$$

as $n \to \infty$. Condition (A1) of Theorem 2.1.9 of [25] is satisfied, this is exactly (40). Since the bivariate random variables $\big((\varepsilon_{u0}, \varepsilon_{u1}), u \in \mathbb{T}\big)$ are independent and identically distributed and since

ε_0 and ε_1 have finite moment of order four,

$$\sum_{k=1}^{|\mathbb{T}_n|}\mathbb{E}\Big[\|\overline{E}_k^{(n)}(x)-\overline{E}_{k-1}^{(n)}(x)\|^4\Big|\mathcal{G}_{k-1}\Big]=\frac{\mathbb{E}\big[(\varepsilon_0^2+\varepsilon_1^2)^2\big]}{(|\mathbb{T}_n|h_n)^2}\sum_{k=1}^{|\mathbb{T}_n|}K^4\big(h_n^{-1}(x-X_{\widetilde{\Pi}(k)})\big)$$
$$=\frac{\mathbb{E}\big[(\varepsilon_0^2+\varepsilon_1^2)^2\big]}{|\mathbb{T}_n|h_n}\Big(\frac{1}{|\mathbb{T}_n|h_n}\sum_{u\in\mathbb{T}_n}K^4\big(h_n^{-1}(x-X_u)\big)\Big)$$

where $\|\cdot\|$ denotes the Euclidian norm for vectors and setting $\overline{E}_0^{(n)}(x)=0$. We apply Proposition 23 (with $|K^2|_2^{-2}K^4$ replacing K as kernel function) and we conclude that

$$\sum_{k=1}^{\mathbb{T}_n}\mathbb{E}\Big[\|\overline{E}_k^{(n)}(x)-\overline{E}_{k-1}^{(n)}(x)\|^4\Big|\mathcal{G}_{k-1}\Big]\to 0,\quad \mathbb{P}_\mu-a.s. \tag{41}$$

The Lyapunov condition (41) implies the Lindeberg condition (A2) of Theorem 2.1.9 of [25] (see section 2.1.4, page 47 of [25]). Therefore, by the central limit theorem for martingale triangular arrays,

$$\overline{E}_{|\mathbb{T}_n|}^{(n)}(x)=\sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}M_{0,n}(x)\\M_{1,n}(x)\end{pmatrix}\xrightarrow{d}\mathcal{N}_2\big(\mathbf{0}_2,|K|_2^2\nu(x)\,\Gamma\big).$$

We conclude gathering Step 1 and Step 2, together with

$$\widehat{\nu}_n(x)\vee\varpi_n\to\nu(x),\quad \mathbb{P}_\mu-a.s.$$

and the Slutsky lemma.

Step 3. Independence. Let $x_1\neq x_2\in\mathcal{D}$. We repete Step 2 for

$$\overline{E}_k^{(n)}(x_1,x_2)=\sum_{l=1}^{k}E_l^{(n)}(x_1,x_2)\quad\text{with}\quad E_l^{(n)}(x_1,x_2)=\begin{pmatrix}E_l^{(n)}(x_1)\\E_l^{(n)}(x_2)\end{pmatrix}$$

and we are led to

$$\overline{E}_{|\mathbb{T}_n|}^{(n)}(x_1,x_2)=\sqrt{|\mathbb{T}_n|h_n}\begin{pmatrix}M_{0,n}(x_1)\\M_{1,n}(x_1)\\M_{0,n}(x_2)\\M_{1,n}(x_2)\end{pmatrix}\xrightarrow{d}\mathcal{N}_4\left(\mathbf{0}_4,|K|_2^2\begin{pmatrix}\nu(x_1)\,\Gamma & \mathbf{0}_{2,2}\\ \mathbf{0}_{2,2} & \nu(x_2)\,\Gamma\end{pmatrix}\right)$$

with $\mathbf{0}_4$ and $\mathbf{0}_{2,2}$ respectively the null vector of size 4 and the null matrix of size 2×2, which shows asymptotic independence between $\big(M_{0,n}(x_1),M_{1,n}(x_1)\big)$ and $\big(M_{0,n}(x_2),M_{1,n}(x_2)\big)$ and thus between $\big(\widehat{f}_{0,n}(x_1),\widehat{f}_{1,n}(x_1)\big)$ and $\big(\widehat{f}_{0,n}(x_2),\widehat{f}_{1,n}(x_2)\big)$ as asserted. □

Proof of Theorem 12. Step 1. Case $\kappa<\infty$. We look carefully at the proof of Theorem 11 and see that only Step 1.1 has to be reconsidered. We prove that

$$\sqrt{|\mathbb{T}_n|h_n}\,N_{\iota,n}^{(1)}(x)\longrightarrow\kappa\nu(x)\boldsymbol{m}(x).$$

Indeed for β an integer, using (35) and (36),

$$\sqrt{|\mathbb{T}_n|h_n}\,N_{\iota,n}^{(1)}(x)=\frac{(-1)^\beta h_n^\beta\sqrt{|\mathbb{T}_n|h_n}}{\beta!}\int_{\mathcal{S}}y^\beta K(y)\big((\nu f_\iota)^\beta(x-\vartheta h_n y)-\nu^\beta(x-\vartheta' h_n y)f_\iota(x)\big)dy$$

and we conclude letting n go to infinity since $(\nu f_\iota)^\beta$ and ν^β are continuous.

Step 2. Case $\kappa=\infty$. With the same argument we prove that in that case,

$$h_n^{-\beta}\,N_{\iota,n}^{(1)}(x)\longrightarrow\nu(x)\boldsymbol{m}(x)$$

as $n \to \infty$. Looking at Steps 1.2 and 1.3 in the proof of Theorem 11, we obtain

$$h_n^{-\beta}\, N_{\iota,n}^{(2)}(x) \longrightarrow 0 \quad \text{and} \quad h_n^{-\beta}\, R_{\iota,n}(x) \longrightarrow 0, \quad \mathbb{P}_\mu - a.s.$$

In addition,

$$h_n^{-\beta}\, M_{\iota,n}(x) = \left(h_n^{\beta}\sqrt{|\mathbb{T}_n|h_n}\right)^{-1} \sqrt{|\mathbb{T}_n|h_n}\, M_{\iota,n}(x) \xrightarrow{\mathbb{P}_\mu} 0,$$

since we work in the case $h_n^{\beta}\sqrt{|\mathbb{T}_n|h_n} \to \kappa = \infty$ and $\sqrt{|\mathbb{T}_n|h_n}\, M_{\iota,n}(x)$ is asymptotically Gaussian as proved previously (Step 3 in the proof of Theorem 11). □

6.7. **Proof of Theorem 16.**

Preliminaries. In the setting of Theorem 16, we achieve uniform ergodicity as stated in the following

Lemma 25 (Uniform ergodicity)**.** *Work under Assumption 4. For every $(f_0, f_1) \in \mathcal{F}(0,\ell)^2$, for every g_ε such that $(g_0, g_1) \in \mathcal{G}(r,\lambda)^2$ satisfy Assumption 14, there exists a unique probability measure ν of the form $\nu(dx) = \nu(x)dx$ on $(\mathcal{S}, \mathfrak{S})$. Moreover, for every g_ε such that $(g_0, g_1) \in \mathcal{G}(r,\lambda)^2$ satisfy Assumption 14, there exists a constant $R > 0$ and $\rho \in (0, 1/2)$ such that*

$$\sup_{(f_0,f_1)} |Q^m h(x) - \nu(h)| \leq R\, |h|_\infty\, \rho^m,$$

where the supremum is taken among all functions $(f_0, f_1) \in \mathcal{F}(0,\ell)^2$, for all $x \in \mathcal{S}$ and $m \geq 0$.

The proof is postponed to the Appendix. Uniform ergodicity achieved by Lemma 25 is here crucial in order to make use of deviations inequalities as obtained in [13]. We recall that the ergodicity proved in Lemma 20 was not uniform in the initial point.

We will intensively use the following two concepts: super-exponential convergence and super-exponential equivalence. Let $(Z_n, n \geq 0)$ be a sequence of random variables and Z a random variable with values in a Polish space $\mathcal{S}$ endowed with its Borel σ-field $\mathfrak{S}$ and endowed with d a metric which makes $(\mathcal{S}, d)$ complete.

Definition 26 (Super-exponential convergence)**.** *We say that (Z_n) converges (s_n)-super-exponentially fast in probability to Z and we note $Z_n \xRightarrow[s_n]{\text{superexp}} Z$ if*

$$\limsup_{n\to+\infty} \frac{1}{s_n} \log \mathbb{P}\left(d(Z_n, Z) > \delta\right) = -\infty$$

for any $\delta > 0$.

Let $(Y_n, n \geq 0)$ be another sequence of random variables with values in $\mathcal{S}$.

Definition 27 (Exponential equivalence, see [19])**.** *We say that (Z_n) and (Y_n) are (s_n)-exponentially equivalent and we note $Z_n \overset{\text{superexp}}{\underset{s_n}{\sim}} Y_n$ if*

$$\limsup_{n\to+\infty} \frac{1}{s_n} \log \mathbb{P}\left(d(Z_n, Y_n) > \delta\right) = -\infty.$$

for any $\delta > 0$.

Proof of Theorem 16. Set x in the interior of $\mathcal{D}$. In order to prove the moderate deviations principle, MDP for short, we use the decomposition

$$\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}\widehat{f}_{0,n}(x)-f_0(x)\\ \widehat{f}_{1,n}(x)-f_1(x)\end{pmatrix}=\tfrac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\Bigg\{\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}M_{0,n}(x)\\ M_{1,n}(x)\end{pmatrix}+\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}N_{0,n}(x)\\ N_{1,n}(x)\end{pmatrix}+\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}R_{0,n}(x)\\ R_{1,n}(x)\end{pmatrix}\Bigg\}$$

where, for $\iota\in\{0,1\}$, $M_{\iota,n}(x)$, $N_{\iota,n}(x)$ and $R_{\iota,n}(x)$ are defined respectively by (22), (30) and (31). The strategy of the proof is the following. After studying the denominator term $\widehat{\nu}_n(x)\vee\varpi_n$ in Step 1, we will prove in Step 2 that the last two terms of the decomposition are negligible in the sense of moderate deviations which leads us to

$$\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}\widehat{f}_{0,n}(x)-f_0(x)\\ \widehat{f}_{1,n}(x)-f_1(x)\end{pmatrix}\underset{b_n^2/(|\mathbb{T}_n|h_n)}{\overset{\text{superexp}}{\sim}}\tfrac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}M_{0,n}(x)\\ M_{1,n}(x)\end{pmatrix}\tag{42}$$

in the sense of Definition 27. Consequently, this two quantities satisfy the same MDP (see [19], Chapter 4) and we prove a MDP for the right-hand side of (42) in Step 3.

Step 1. Denominator $\widehat{\nu}_n(x)\vee\varpi_n$. We claim that

$$\widehat{\nu}_n(x)\vee\varpi_n\xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}\nu(x).\tag{43}$$

On the one hand, we have seen that

$$K_{h_n}\star\nu(x)\to\nu(x)$$

as $n\to\infty$ (see (18)). Since the previous sequence is deterministic, we conclude that

$$K_{h_n}\star\nu(x)\xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}\nu(x).\tag{44}$$

On the other hand, from [13],

$$\mathbb{P}_\mu\Big(\Big|\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}K_{h_n}(x-X_u)-K_{h_n}\star\nu(x)\Big|>\delta\Big)\le 2\exp\Big(\frac{-\delta^2|\mathbb{T}_n|h_n}{c_1+c_2\delta}\Big)\tag{45}$$

where c_1 and c_1 are positive constants which depend on K, R, ρ, Q but not on n. Note that Lemma 25 ensures we are in the setting of [13]. Applying the log to the two terms of (45), multiplying by $|\mathbb{T}_n|h_n/b_n^2$ and letting n go to infinity, we are led to

$$\frac{1}{|\mathbb{T}_n|}\sum_{u\in\mathbb{T}_n}K_{h_n}(x-X_u)-K_{h_n}\star\nu(x)\xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0\tag{46}$$

using conditions (i) and (ii) on the sequence b_n. From the foregoing (44) and (46), we conclude that

$$\widehat{\nu}_n(x)\xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}\nu(x).\tag{47}$$

To reach (43), write the decomposition

$$|\widehat{\nu}_n(x)\vee\varpi_n-\nu(x)|=|\widehat{\nu}_n(x)-\nu(x)|\mathbf{1}_{\{\widehat{\nu}_n(x)\ge\varpi_n\}}+|\varpi_n-\nu(x)|\mathbf{1}_{\{\widehat{\nu}_n(x)<\varpi_n\}},$$

and note that it just remains to prove

$$\mathbf{1}_{\{\widehat{\nu}_n(x)<\varpi_n\}}\xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0.\tag{48}$$

We have

$$\mathbb{P}_\mu\big(\mathbf{1}_{\{\widehat{\nu}_n(x)<\varpi_n\}}>\delta\big)\leq\delta^{-1}\mathbb{P}_\mu\big(\widehat{\nu}_n(x)<\varpi_n\big)\leq\mathbb{P}\big(|\widehat{\nu}_n(x)-K_{h_n}\star\nu(x)\mid>\delta'\big)$$

with $\delta'=\inf_{(f_0,f_1)}\inf_{x\in\mathcal{D}}K_{h_n}\star\nu(x)-\varpi_n>0$ for n large enough under Assumption 15. Using the deviations inequality (45) with δ', we obtain (43), which ends the first step.

Step 2. Negligible and remainder terms, $N_{\iota,n}(x)$ and $R_{\iota,n}(x)$. We claim that

$$\frac{|\mathbb{T}_n|h_n}{b_n}N_{\iota,n}(x)\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0. \tag{49}$$

Remind the decomposition $N_{\iota,n}(x)=N_{\iota,n}^{(1)}(x)+N_{\iota,n}^{(2)}(x)$ with $N_{\iota,n}^{(1)}(x)$ and $N_{\iota,n}^{(2)}(x)$ defined by (32) and (33). On the one hand, we have proved that $|N_{\iota,n}^{(1)}(x)|\lesssim h_n^\beta$ (Step 1.1 in the proof of Theorem 11). Thus

$$\frac{|\mathbb{T}_n|h_n}{b_n}|N_{\iota,n}^{(1)}(x)|\lesssim\frac{|\mathbb{T}_n|h_n^{1+\beta}}{b_n}$$

and since the right-hand side of the previous inequality is deterministic and tends to 0 (condition (iii) on the sequence b_n), we conclude that

$$\frac{|\mathbb{T}_n|h_n}{b_n}N_{\iota,n}^{(1)}(x)\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0,$$

see for e.g. [50] for more details. On the other hand, from [13], the following deviations inequality holds

$$\mathbb{P}_\mu\big(|N_{\iota,n}^{(2)}(x)|>\delta\big)\leq 2\exp\Big(\frac{-\delta^2|\mathbb{T}_n|h_n}{c_1'(h_n^\beta+h_n^{2\beta+1}|\mathbb{T}_n|)+c_2'\delta}\Big)$$

where c_1' and c_1' are positive constants which depend on K, R, ρ, Q but not on n. More precisely, we apply the results of [13] replacing the terms $|Q|_{\mathcal{S},\mathcal{D}}\|g\|_{\mathbb{L}^1(\mu)}$ and $|Q|_{\mathcal{S},\mathcal{D}}\|g\|^2_{\mathbb{L}^2(\mu)}$ by respectively $|Qg|_\infty$ and $|Qg^2|_\infty$, controlled by h_n^β and $h_n^{\beta-1}$ respectively for the test function $g(\cdot)=h_n^{-1}H_n(\cdot)$ with H_n defined by (34). Recalling conditions (i) and (ii) on b_n and $h_n=|\mathbb{T}_n|^{-\alpha}$ with $\alpha>1/(2\beta+1)$, it brings

$$\frac{|\mathbb{T}_n|h_n}{b_n}N_{\iota,n}^{(2)}(x)\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0.$$

Hence (49) is proved. We also have

$$\frac{|\mathbb{T}_n|h_n}{b_n}R_{\iota,n}(x)\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0,$$

using (38) and recalling (47) and (48). Together with Step 1, it leads us to

$$\tfrac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}N_{0,n}(x)\\N_{1,n}(x)\end{pmatrix}+\tfrac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}R_{0,n}(x)\\R_{1,n}(x)\end{pmatrix}\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}0$$

and finally to (42).

Step 3. Main term $M_{\iota,n}(x)$. Consider the $\mathcal{G}$-martingale triangular array $(\overline{E}_k^{(n)}(x),1\leq k\leq n,n\geq 1)$ defined by (39). In order to make use of the MDP for martingale triangular arrays (see [50, 51, 43]), one need to check

$$\langle\overline{E}^{(n)}(x)\rangle_{|\mathbb{T}_n|}\xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}}|K|_2^2\nu(x)\Gamma, \tag{50}$$

and

$$\sum_{k=1}^{|\mathbb{T}_n|} \mathbb{E}\left[\|\overline{E}_k^{(n)} - \overline{E}_{k-1}^{(n)}\|^4 \Big| \mathcal{G}_{k-1}\right] \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{51}$$

The condition (51) is called exponential Lyapunov condition and it implies exponential Lindeberg condition, we refer to [50, 51] for more details. Using (40) and (41), one see that to prove (50) and (51), it suffices to show

$$\frac{1}{|\mathbb{T}_n|h_n} \sum_{u\in\mathbb{T}_n} K^2\big(h_n^{-1}(x - X_u)\big) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} |K|_2^2 \nu(x) \tag{52}$$

and

$$\frac{1}{(|\mathbb{T}_n|h_n)^2} \sum_{u\in\mathbb{T}_n} K^4\big(h_n^{-1}(x - X_u)\big) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{53}$$

The same argument as in Step 1 (replacing K by $|K|_2^{-2}K^2$ or $|K^2|_2^{-2}K^4$ in (45)) enables us to prove (52) and (53). Gathering (50) and (51) and using the truncation of the martingale $(\overline{E}_k^{(n)}(x), 1 \le k \le n, n \ge 1)$ as in the proof of Theorem 5.1 of [10], we conclude that

$$\frac{\sqrt{|\mathbb{T}_n|h_n}}{b_n} \overline{E}_{|\mathbb{T}_n|}^{(n)}(x) = \frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}$$

satisfies a MDP on $\mathcal{S}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and the good rate function defined for all $\boldsymbol{z} \in \mathcal{S}^2$ by

$$I_x(\boldsymbol{z}) = \sup_{\boldsymbol{\lambda}\in\mathbb{R}^2} \left\{\boldsymbol{\lambda}^t \boldsymbol{z} - \frac{1}{2}\nu(x)|K|_2^2 \boldsymbol{z}^t \Gamma \boldsymbol{z}\right\} = \big(2|K|_2^2\big)^{-1} \nu(x)^{-1} \boldsymbol{z}^t \Gamma^{-1} \boldsymbol{z}.$$

Finally, using Lemma 4.1 of [49] (which is a consequence of the contraction principle) and Step 1, we conclude that

$$\frac{1}{\widehat{\nu}_n(x)\vee\varpi_n} \frac{|\mathbb{T}_n|h_n}{b_n} \begin{pmatrix} M_{0,n}(x) \\ M_{1,n}(x) \end{pmatrix}$$

satisfies a MDP on $\mathcal{S}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and rate function J_x defined in Theorem 16. Reminding (42), we get the MDP stated in Theorem 16. □

6.8. **Proof of Theorem 18.** On the one hand, applying Theorem 12 to the estimator $\widehat{f}_{\iota,n}^{(a)}$ built with the bandwidth $h_n^{(a)} = |\mathbb{T}_n|^{-1/(2\beta+1)}$,

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} A_n(x) \xrightarrow{d} \mathcal{N}_2\big(\boldsymbol{m}_2(x), \boldsymbol{\Sigma}_2(x)\big) \quad \text{with} \quad A_n(x) = \begin{pmatrix} \widehat{f}_{0,n}^{(a)}(x) - f_0(x) \\ \widehat{f}_{1,n}^{(a)}(x) - f_1(x) \end{pmatrix}.$$

On the other hand, applying Theorem 12 to the estimator $\widehat{f}_{\iota,n}^{(b)}$ built with the bandwidth $h_n^{(b)} = |\mathbb{T}_n|^{-\delta/(2\beta+1)}$ for $\delta \in (0,1)$,

$$|\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}} B_n(x) \xrightarrow{\mathbb{P}_\mu} \boldsymbol{m}_2(x) \quad \text{with} \quad B_n(x) = \begin{pmatrix} \widehat{f}_{0,n}^{(b)}(x) - f_0(x) \\ \widehat{f}_{1,n}^{(b)}(x) - f_1(x) \end{pmatrix}.$$

Combining these two results, we obtain

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} \begin{pmatrix} \bar{f}_{0,n}(x) - f_0(x) \\ \bar{f}_{1,n}(x) - f_1(x) \end{pmatrix} = \frac{|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}} A_n(x) - |\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}} B_n(x)}{1 - |\mathbb{T}_n|^{-\frac{(1-\delta)\beta}{2\beta+1}}} \xrightarrow{d} \mathcal{N}_2\big(\mathbf{0}_2, \boldsymbol{\Sigma}_2(x)\big),$$

as announced in (12). The study of the test statistics $W_n(x_1,\dots,x_k)$ under $\mathcal{H}_0$ and $\mathcal{H}_1$ then follows classical lines. We give here the main argument for $k=2$. By (12) and using in addition the asymptotical independence stated in Theorem 11, we obtain

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}}\begin{pmatrix}\bar f_{0,n}(x_1)-f_0(x_1)\\ \bar f_{1,n}(x_1)-f_1(x_1)\\ \bar f_{0,n}(x_2)-f_0(x_2)\\ \bar f_{1,n}(x_2)-f_1(x_2)\end{pmatrix}\xrightarrow{d}\mathcal{N}_4\big(\mathbf{0}_4,\mathbf{\Sigma}_4(x_1,x_2)\big)\ \text{ with }\ \mathbf{\Sigma}_4(x_1,x_2)=\begin{pmatrix}\mathbf{\Sigma}_2(x_1) & \mathbf{0}_{2,2}\\ \mathbf{0}_{2,2} & \mathbf{\Sigma}_2(x_2)\end{pmatrix}.$$

Then, using the Delta-method,

$$|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}}\begin{pmatrix}\big(\bar f_{0,n}(x_1)-\bar f_{1,n}(x_1)\big)-\big(f_0(x_1)-f_1(x_1)\big)\\ \big(\bar f_{0,n}(x_2)-\bar f_{1,n}(x_2)\big)-\big(f_0(x_2)-f_1(x_2)\big)\end{pmatrix}\xrightarrow{d}\mathcal{N}_2\big(\mathbf{0}_2,\mathbf{\Sigma}_2(x_1,x_2)\big)$$

with

$$\mathbf{\Sigma}_2(x_1,x_2)=|K|_2^2(\sigma_0^2+\sigma_1^2-2\sigma_0\sigma_1\varrho)\begin{pmatrix}\big(\nu(x_1)\big)^{-1} & 0\\ 0 & \big(\nu(x_2)\big)^{-1}\end{pmatrix},$$

which boils down to, under $\mathcal{H}_0$,

$$\frac{|\mathbb{T}_n|^{-\frac{\beta}{2\beta+1}}}{\sqrt{|K|_2^2(\sigma_0^2+\sigma_1^2-2\sigma_0\sigma_1\varrho)}}\begin{pmatrix}\nu(x_1)^{1/2}\big(\bar f_{0,n}(x_1)-\bar f_{1,n}(x_1)\big)\\ \nu(x_2)^{1/2}\big(\bar f_{0,n}(x_2)-\bar f_{1,n}(x_2)\big)\end{pmatrix}\xrightarrow{d}\mathcal{N}_2\big(\mathbf{0}_2,\boldsymbol{I}_2\big),$$

with $\boldsymbol{I}_2$ the identity matrix of size 2×2. The replacement of $\nu(\cdot)$ by its estimator $\widehat{\nu}_n(\cdot)$ is licit by the Stlutsky theorem. Thus, under $\mathcal{H}_0$, $W_n(x_1,x_2)\xrightarrow{d}\chi^2(2)$, the chi-squared distribution with 2 degrees of freedom. Under $\mathcal{H}_1$, we prove that $W_n(x_1,x_2)$ converges $\mathbb{P}_\mu$-almost surely to $+\infty$ following the same lines and using

$$\sum_{l=1}^{2}\widehat{\nu}_n(x_l)\big(\bar f_{0,n}(x_l)-\bar f_{1,n}(x_l)\big)^2\longrightarrow\sum_{l=1}^{2}\nu(x_l)\big(f_0(x_l)-f_1(x_l)\big)^2\neq 0,\quad \mathbb{P}_\mu-a.s.$$

when $\mathcal{H}_1$ is valid.

7. Appendix

7.1. **Proof of Lemma 19.** Let us first prove (i), see [17] (Lemma 2.1). For $u=(u_1,u_2,\dots,u_m)\in\mathbb{G}_m$,

$$\mathbb{E}_\mu\big[g(X_u)\big]=\mu\big(\mathcal{P}_{u_1}\mathcal{P}_{u2}\dots\mathcal{P}_{u_m}(g)\big).$$

Then

$$\mathbb{E}_\mu\big[\sum_{u\in\mathbb{G}_m}g(X_u)\big]=\sum_{\substack{(u_1,\dots,u_m)\\ \in\{0,1\}^m}}\mu\big(\mathcal{P}_{u_1}\mathcal{P}_{u2}\dots\mathcal{P}_{u_m}(g)\big)=\mu\Big(\sum_{\substack{(u_1,\dots,u_m)\\ \in\{0,1\}^m}}\mathcal{P}_{u_1}\mathcal{P}_{u2}\dots\mathcal{P}_{u_m}(g)\Big)$$
$$=\mu\big(2^m(\mathcal{P}_0+\mathcal{P}_1)^m(g)\big)=|\mathbb{G}_m|\mu(Q^m g)$$

since $Q=(\mathcal{P}_0+\mathcal{P}_1)/2$ and $|\mathbb{G}_m|=2^m$. We also know that $\mu(Q^m g)=\mathbb{E}_\mu[g(Y_m)]$.

We now turn to (ii). We refer to [27] for another strategy of proof (proof of Equation (7)). Some notation first: for $u=(u_1,\dots,u_m)$ and $v=(v_1,\dots,v_n)$ in $\mathbb{T}$, we write $uv=(u_1,\dots,u_m,v_1,\dots,v_n)$ for the concatenation. For $m\geq 0$, we denote by $\mathcal{F}_m$ the sigma-field generated by $(X_u,|u|\leq m)$.

For $m\geq 1$, whenever $u\neq v\in\mathbb{G}_m$, there exists $w\in\mathbb{T}_{m-1}$ together with $i\neq j\in\{0,1\}$ and $\tilde u,\tilde v$ such that $u=wi\tilde u$ and $v=wj\tilde v$, where we call w the most recent common ancestor of u and v. The main argument uses consecutively a first conditioning by $\mathcal{F}_{|w|+1}$ which lets X_u and X_v

conditionally independent and a conditional many-to-one formula of kind (i), a second conditioning by $\mathcal{F}_{|w|}$ and the definition of the $\mathbb{T}$-transition $\boldsymbol{\mathcal{P}}$ and finally the many-to-one formula (i),

$$\begin{aligned}
\mathbb{E}_\mu\Big[\sum_{u\neq v\in\mathbb{G}_m} g(X_u)g(X_v)\Big] &= 2\sum_{l=1}^m \sum_{w\in\mathbb{G}_{m-l}} \mathbb{E}_\mu\big[\sum_{\substack{\tilde u\in\mathbb{G}_{l-1}\\ \tilde v\in\mathbb{G}_{l-1}}} g(X_{w0\tilde u})g(X_{w1\tilde v})\big]\\
&= 2\sum_{l=1}^m \sum_{w\in\mathbb{G}_{m-l}} |\mathbb{G}_{l-1}|^2\mathbb{E}_\mu\Big[Q^{l-1}g(X_{w0})Q^{l-1}g(X_{w1})\Big]\\
&= 2\sum_{l=1}^m (2^{l-1})^2 \sum_{w\in\mathbb{G}_{m-l}} \mathbb{E}_\mu\Big[\boldsymbol{\mathcal{P}}(Q^{l-1}g\otimes Q^{l-1}g)(X_w)\Big]\\
&= 2\sum_{l=1}^m (2^{l-1})^2 |\mathbb{G}_{m-l}|\mathbb{E}_\mu\Big[\boldsymbol{\mathcal{P}}(Q^{l-1}g\otimes Q^{l-1}g)(Y_{m-l})\Big]\\
&= 2^m\sum_{l=1}^m 2^{l-1}\mu\Big(Q^{m-l}\big(\boldsymbol{\mathcal{P}}(Q^{l-1}g\otimes Q^{l-1}g)\big)\Big),
\end{aligned}$$

as asserted.

7.2. **Proof of Lemma 22.** First,

$$\begin{aligned}
|QH_n(y)| &\le \int_{\mathcal{S}} |F\big(h_n^{-1}(x-z)\big)||G(z)|Q(y,z)dz\\
&= h_n\int_{\mathrm{supp}(F)} |F(z)||G(x-h_nz)|Q(y,x-h_nz)dz \le h_n\,|G|_{\mathcal{D}}|Q|_{\mathcal{S},\mathcal{D}}|F|_1
\end{aligned}$$

for h_n such that $x-h_n\mathrm{supp}(F)\subset\mathcal{D}$ (remind that x belongs to the interior of $\mathcal{D}$). We prove in the same way the bound on $\nu(H_n)$. Hence we have proved (i) and we now turn to (ii). The first bound h_n obviously comes from (i) and it remains to prove the second bound $\mathbb{V}(y)\rho^m$. Under Assumption 5, we apply Proposition 20 to $g=H_n/|H_n|_\infty$ and it brings

$$|Q^mH_n(y)-\nu(H_n)|\le R|H_n|_\infty\rho^m\mathbb{V}(y).$$

Since $|H_n|_\infty=|F|_\infty|G|_{\mathcal{D}}$ for h_n such that $x-h_n\mathrm{supp}(F)\subset\mathcal{D}$, we obtain the announced upper-bound.

7.3. **Proof of Lemma 24.** For every $|z|\le d$,

$$\nu(z)=\int_{\mathcal{S}}\nu(y)Q(y,z)dy \ge \inf_{\substack{|y|\le M_2,\\ |z|\le d}} Q(y,z)\int_{|y|\le M_2}\nu(y)dy.$$

On the one hand,

$$\inf_{\substack{|x|\le M_2,\\ |y|\le d}} Q(x,y)\ge\frac12\Big\{\inf_{\substack{|x|\le M_2,\\ |y|\le d}} g_0\big(y-f_0(x)\big)+\inf_{\substack{|x|\le M_2,\\ |y|\le d}} g_1\big(y-f_1(x)\big)\Big\}\ge\delta\big(d+(\ell+\gamma M_2)\big)\ge\delta(M_3)>0$$

if $d>0$ is such that $d+(\ell+\gamma M_2)\le M_3$, which is possible from Assumption 6. On the other hand,

$$\begin{aligned}
\int_{|y|>M_2}\nu(y)dy &= \int_{|y|>M_2}\int_{x\in\mathcal{S}}\nu(x)Q(x,y)dxdy\\
&\le |\nu|_\infty\int_{|y|>M_2}\int_{x\in\mathcal{S}}\frac12\Big(g_0\big(y-f_0(x)\big)+g_1\big(y-f_1(x)\big)\Big)dxdy\le\eta(M_2)
\end{aligned}$$

since (f_0, f_1) belongs to $\mathcal{F}(\gamma,\ell)^2$ and (g_0, g_1) to $\mathcal{G}(r,\lambda)^2$. We know in addition that $|\nu|_\infty \le |Q|_{\mathcal{S},\mathcal{S}} \le (|g_0|_\infty + |g_1|_\infty)/2 < \infty$. Using $\eta(M_2) < 1$ brings

$$\int_{|y|\le M_2} \nu(y)dy > 0.$$

Hence $\nu(y) \ge \delta(M_3)\big(1-\eta(M_2)\big) > 0$ for any $|y| \le d$. We have uniformity in f_0 and f_1 since $\delta(M_3)$ and $\eta(M_2)$ are uniform bounds on the class $\mathcal{F}(\gamma,\ell)^2$.

7.4. **Proof of Lemma 25.** Set $\mathcal{C} = \{y \in \mathcal{S}; |y| \le M_4 - \ell\} \neq \emptyset$ since $M_4 > \ell$ under Assumption 14. We prove that, for all $y \in \mathcal{C}$,

$$\inf_{x\in\mathcal{S}} Q(x,y) = \frac{1}{2}\Big\{ \inf_{x\in\mathcal{S}} g_0\big(y - f_0(x)\big) + \inf_{x\in\mathcal{S}} g_1\big(y - f_1(x)\big)\Big\} \ge \delta(M_4) > 0$$

with $\delta(\cdot)$ defined by (8), since $|y - f_\iota(x)| \le M_4$ (for $|f_\iota|_\infty \le \ell$ and $|y| \le M_4 - \ell$). From Douc *et al.* [23], Theorem 6.6 and Lemma 6.10, one can take $\rho = 1 - |\mathcal{C}|\delta(M_4) = 1 - 2(M_4 - \ell)\delta(M_4)$ and we obtain $\rho < 1/2$ under Assumption 14.

Acknowledgements. We are grateful to A. Guillin and M. Hoffmann for helpful comments. We thank G. Fort for suggesting us the enlightening reading of the recent textbook [23]. S.V. B.P. thanks the Hadamard Mathematics Labex of the *Fondation Mathématique Jacques Hadamard* for financial support.

References

[1] H. Z. An and F.C. Huang. *The geometric ergodicity of nonlinear autoregressive model.* Statistica Sinica, 6 (1996), 943–956.

[2] I. V. Basawa and R. M. Huggins. *Extensions of the bifurcating autoregressive model for cell lineage studies.* Journal of Applied Probability, 36 (1999), 1225–1233.

[3] I. V. Basawa and R. M. Huggins. *Inference for the extended bifurcating autoregressive model for cell lineage studies.* Australian & New Zealand Journal of Statistics, 42 (2000), 423–432.

[4] I. V. Basawa and J. Zhou. *Non-Gaussian bifurcating models and quasi-likelihood estimation.* Journal of Applied Probability, 41 (2004), 55-64.

[5] I. V. Basawa and J. Zhou. *Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors.* Journal of Time Series Analysis, 26 (2005), 825–842.

[6] B. Bercu, V. Blandin. *A Rademacher-Menchov approach for random coefficient bifurcating autoregressive processes.* Stochastic Processes and their Applications, 125 (2015), 1218-1243.

[7] B. Bercu, B. de Saporta and A. Gégout-Petit. *Asymptotic analysis for bifurcating autoregressive processes via a martingale approach.* Electronic Journal of Probability, 87 (2009), 2492–2526.

[8] R. Bhattacharya and C. Lee. *On geometric ergodicity of nonlinear autoregressive models.* Statistics & Probability Letters, 22 (1995), 311–315.

[9] H. J. Bierens. *Kernel estimator of regressions functions.* Advances in econometrics: Fifth world congress, Vol. 1 (1987).

[10] S. V. Bitseki Penda and H. Djellout. *Deviation inequalities and moderate deviations for estimators of parameters in bifurcating autoregressive models.* Annales de l'IHP-PS, 50 (2014), 806–844.

[11] S. V. Bitseki Penda, H. Djellout and A. Guillin. *Deviation inequalities, moderate deviations and some limit theorems for bifurcating Markov chains with application.* The Annals of Applied Probability, 24 (2014), 235–291.

[12] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains.* Submitted (2015)

[13] S. V. Bitseki Penda, M. Hoffmann and A. Olivier. *Adaptive estimation for bifurcating Markov chains.* In preparation.

[14] V. Blandin. *Asymptotic results for random coefficient bifurcating autoregressive processes.* Statistics (2013), to appear.

[15] D. B. H. Cline. *Stability of nonlinear stochastic recursions with application to nonlinear AR-GARCH models.* Advances in Applied Probability, 39 (2007), 462–491.

[16] R. Cowan and R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies*. Biometrics, 42 (1986), 769–783.
[17] J. F. Delmas, L. Marsalle. *Detection of cellular aging in Galton-Watson process*. Stochastic Processes and their Applications, 12 (2010), 2495–2519.
[18] V. Delouille and R. von Sachs. *Estimation of nonlinear autoregressive models using design-adapted wavelets.* Annals of the Institute of Statistical Mathematics, 57 (2005), 235–253.
[19] A. Dembo and O. Zeitouni. *Large deviations techniques and applications, second edition*. Second edition, volume 38 of *Applications of Mathematics*, Springer, 1998.
[20] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Parameters estimation for asymmetric bifurcating autoregressive processes with missing data*. Electronic Journal of Statistics, 5 (2011), 1313–1353.
[21] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Asymmetry tests for bifurcating autoregressive processes with missing data*. Statistics & Probability Letters, 82 (2012), 1439–1444.
[22] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Random coefficients bifurcating autoregressive processes*. ESAIM PS, to appear.
[23] R. Douc, E. Moulines and D. S. Stoffer. *Nonlinear Time Series*. Chapman & Hall/CRC Texts in Statistical Science, 2014.
[24] M. Doumic, M. Hoffmann, N. Krell and L. Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree*. Bernoulli, to appear.
[25] M. Duflo. *Random Iterative Models*. Springer, Berlin (1998)
[26] R. L. Eubank and V. N. LaRiccia. *Testing for no effect in nonparametric regression.* Journal of statistical planning and inference, 36 (1993), 1–14.
[27] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging*. The Annals of Applied Probability, 17 (2007), 1538–1569.
[28] M. Hairer and J. Mattingly. *Yet another look at Harris' ergodic theorem for Markov chains*. Seminar on Stochastic Analysis, Random Fields and Applications VI, Progress in Probability, 63 (2011), 109–117.
[29] W. Härdle and P. Vieu. *Kernel regression smoothing of time series.* Journal of Time Series Analysis, 13 (1992), 209–232.
[30] M. Hoffmann. *On nonparametric estimation in nonlinear AR(1)-models*. Statistics & Probability Letters, 44 (1999), 29–45.
[31] L. Jin. *A data-driven test to compare two or multiple time series.* Computational Statistics & Data Analysis, 55 (2011), 2183–2196.
[32] C. Joutard. *Sharp large deviations in nonparametric estimation*. Nonparametric Statistics, 18 (2006), 293–306.
[33] F. Li. *Comparison of autoregressive curves through partial sums of quasi-residuals*. Contemporary Developments in Statistical Theory, Springer Proceedings in Mathematics & Statistics, 68 (2014), 239-256.
[34] W. Liu and W. B. Wu. *Simultaneous nonparametric inference for times series.* The Annals of Statistics, 38 (2010), 2388–2421.
[35] D. Louani. *Large deviations limit theorems for the kernel density estimator*. Scandinavian Journal of Statistics, 25 (1998), 243–253.
[36] D. Louani. *Some large deviations limit theorems in conditional nonparametric statistics*. Statistics: A Journal of Theoretical and Applied Statistics, 33 (1999), 171–196.
[37] S. Meyn and R. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, 2009.
[38] A. Mokkadem and M. Pelletier. *Confidence bands for densities, logarithmic point of view.* Alea, 2 (2006), 231–266.
[39] A. Mokkadem, M. Pelletier and B. Thiam. *Large and moderate deviations principles for kernel estimators of the multivariate regression.* Mathematical Methods of Statistics, 17 (2008), 1–27.
[40] A. Munk and H. Dette. *Nonparametric comparison of several regression functions: exact and asymptotic theory.* Annals of Statistics, 26 (1998), 2339–2368.
[41] E. A. Nadaraya. *On estimating regression*. Theory of Probability & Its Applications, 9 (1964), 141–142.
[42] B. L. S. Prakasa Rao. *Nonparametric Functional Estimation*. Academic Press, New York (1983)
[43] A. Puhalskii. *Large deviations of semimartingales via convergence of the predictable characteristics*. Stochastics and Stochastics Reports, 49 (1994), 27–85.
[44] P. M. Robinson. *Nonparametric estimators for time series.* Journal of Time Series Analysis, 4 (1983), 185–207.
[45] J. J. Trujillo, M. Rivero and B. Bonilla. *On a Riemann-Liouville generalized Taylor's formula.* Journal of Mathematical Analysis and Applications, 231 (1999), 255–265.
[46] A. Tsybakov. *Introduction to Nonparametric Estimation*. Springer series in statistics, Springer-Verlag, New-York, 2009.

[47] D. Usero. *Fractional Taylor series for Caputo fractional derivatives. Construction of numerical schemes.* Preprint (2008)

[48] G. S. Watson. *Smooth regression analysis.* Sankhyā: The Indian Journal of Statistics, Series A, 26 (1964), 359–372.

[49] J. Worms. *Principes de déviations modérées pour des martingales et applications statistiques.* Thèse de Doctorat à l'Université Marne-la-Vallée (2000).

[50] J. Worms. *Moderate deviations of some dependent variables. I. Martingales.* Mathematical Methods of Statistics, 10 (2001), 38–72.

[51] J. Worms. *Moderate deviations of some dependent variables. II. Some kernel estimators* Mathematical Methods of Statistics, 10 (2001), 161–193.

S. Valère Bitseki Penda, CMAP, École Polytechnique, Route de Saclay, 91128 Palaiseau, France.
E-mail address: bitseki@cmap.polytechnique.fr

Adélaïde Olivier, CEREMADE, CNRS-UMR 7534 and CREST-ENSAE, 3, avenue Pierre Larousse, 92245 Malakoff Cedex, France.
E-mail address: adelaide.olivier@ensae.fr