HAL
open science

Moderate deviation principle in nonlinear bifurcating autoregressive models

Siméon Valère Bitseki Penda, Adélaïde Olivier

To cite this version:

Siméon Valère Bitseki Penda, Adélaïde Olivier. Moderate deviation principle in nonlinear bifurcating autoregressive models. Statistics and Probability Letters, 2018, 138, pp.20-26. 10.1016/j.spl.2018.02.037 . hal-01710079

HAL Id: hal-01710079

https://hal.science/hal-01710079

Submitted on 15 Feb 2018

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.

MODERATE DEVIATION PRINCIPLE IN NONLINEAR BIFURCATING AUTOREGRESSIVE MODELS

S. VALÈRE BITSEKI PENDA AND ADÉLAÏDE OLIVIER

Abstract. Recently, nonparametric techniques have been proposed to study bifurcating autoregressive processes – an adaptation of autoregressive processes for a binary tree structure. One can build Nadaraya-Watson type estimators of the two autoregressive functions as in [3] and [5]. In the present work, we prove moderate deviation principle for these estimators.

Keywords: Bifurcating Markov chains, binary trees, bifurcating autoregressive processes, nonparametric estimation, Nadaraya-Watson estimator, moderate deviation principle.

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

1. Introduction

1.1. **A generalization of BAR processes.** Bifurcating autoregressive (BAR) processes where introduced by Cowan an Staudte [6] in 1986 to study *E. coli* bacterium. Since then it has been extensively studied. We refer in particular to the recent works of Bercu and Blandin [1], de Saporta, Gégout-Petit and Marsalle [8], see also references therein. Nonlinear bifurcating autoregressive (NBAR) processes, studied in Bitseki Penda *et al.* [4, 5], generalize BAR processes, avoiding an *a priori* linear specification on the two autoregressive functions.

We first need some notation. We introduce the infinite binary tree whose vertices are indexed by the positive integers: the initial individual is indexed by 1 and an individual $k \geq 1$ gives birth to two individuals $2k$ and $2k+1$. For $m \geq 0$, let $\mathbb{G}_m = \{2^m, \ldots, 2^{m+1}-1\}$ be the m-th generation. A given individual $k \geq 1$ lives in the r_k-th generation with $r_k = \lfloor \log_2 k \rfloor$.

Let us now introduce precisely a NBAR process which is specified by **1)** a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m\geq 0}, \mathbb{P})$, together with a measurable state space $(\mathbb{R}, \mathfrak{B})$, **2)** two measurable functions $f_0, f_1 : \mathbb{R} \to \mathbb{R}$ and **3)** a probability density G on $(\mathbb{R}\times\mathbb{R}, \mathfrak{B}\otimes\mathfrak{B})$ with a null first order moment. In this setting we have the following

Definition 1. *A NBAR process is a family $(X_k)_{k\geq 1}$ of random variables with value in $(\mathbb{R}, \mathfrak{B})$ such that, for every $k \geq 1$, X_k is $\mathcal{F}_{r_k}$-measurable and*

$$X_{2k} = f_0(X_k) + \varepsilon_{2k} \quad \text{and} \quad X_{2k+1} = f_1(X_k) + \varepsilon_{2k+1}$$

where $\big((\varepsilon_{2k}, \varepsilon_{2k+1})\big)_{k\geq 1}$ is a sequence of independent bivariate random variables with common density G.

The distribution of $(X_k)_{k\geq 1}$ is thus entirely determined by the autoregressive functions (f_0, f_1), the noise density G and an initial distribution for X_1. Informally, each $k \geq 1$ is viewed as a particle of feature X_k (size, lifetime, growth rate, DNA content and so on) with value in $\mathbb{R}$. Conditional on $X_k = x$, the feature $(X_{2k}, X_{2k+1}) \in \mathbb{R}^2$ of the offspring of k is a perturbed version of $\big(f_0(x), f_1(x)\big)$.

When X_1 is distributed according to a measure $\mu(dx)$ on $(\mathbb{R}, \mathfrak{B})$, we denote by $\mathbb{P}_\mu$ the law of the NBAR process $(X_u)_{u\in\mathbb{T}}$ and by $\mathbb{E}_\mu[\cdot]$ the expectation with respect to the probability $\mathbb{P}_\mu$.

1.2. **Nadaraya-Watson type estimator of the autoregressive functions.** For $n \geq 0$, introduce the genealogical tree up to the $(n+1)$-th generation, $\mathbb{T}_{n+1} = \bigcup_{m=0}^{n+1} \mathbb{G}_m$. Assume we observe $\mathbb{X}^{n+1} = (X_k)_{k\in\mathbb{T}_{n+1}}$, *i.e.* we have $|\mathbb{T}_{n+1}| = 2^{n+2} - 1$ random variables with value in $\mathbb{R}$. Let $\mathcal{D} \subset \mathbb{R}$ 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 $\mathbb{X}^{n+1}$ by

$$\Big(\widehat{f}_{\iota,n}(x) = \frac{|\mathbb{T}_n|^{-1} \sum\limits_{k\in\mathbb{T}_n} K_{h_n}(x - X_k) X_{2k+\iota}}{\big(|\mathbb{T}_n|^{-1} \sum\limits_{k\in\mathbb{T}_n} K_{h_n}(x - X_k)\big) \vee \varpi_n}, \iota \in \{0, 1\} \Big), \tag{1}$$

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 $\int_{\mathbb{R}} K = 1$. Almost sure convergence to $\big(f_0(x), f_1(x)\big)$ and asymptotic normality of these estimators have been studied by Bitseki Penda and Olivier in [5].

1.3. **Objective.** The purpose of this work is to establish a moderate deviation principle for the estimators of the autoregressive functions defined by (1). Roughly speaking, for some range of speed $(b_n, n \in \mathbb{N})$ such that $\sqrt{|\mathbb{T}_n|^{1-\gamma}} \ll b_n \ll |\mathbb{T}_n|^{1-\gamma}$ where $\gamma \in (0, 1)$, for all $x \in \mathbb{R}$ and for all $\delta > 0$, our goal is to establish asymptotic equivalences of the form

$$\frac{|\mathbb{T}_n| h_n}{b_n^2} \log \mathbb{P}\left(\frac{|\mathbb{T}_n| h_n}{b_n} |\widehat{f}_{\iota,n}(x) - f_\iota(x)| > \delta \right) \sim -\frac{\nu(x)\delta^2}{2\sigma_\iota^2 \int_S K^2(y)dy},$$

where σ_ι^2 denotes the variance of $\varepsilon_{1+\iota}$ and the function $\nu(\cdot)$ will be specified later.

Statistical estimators are also studied under the angle of large and moderate deviation 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 Louani and Joutard [12, 13, 11] in the univariate case, see also Mokkadem *et al.* [15] and references therein). We refer to Mokkadem and Pelletier [14] for the study of confidence bands based on the use of moderate deviation principles.

Before we proceed, let us introduce the notion of moderate deviation principle in a general setting. Let $(Z_n)_{n\geq 0}$ be a sequence of random variables with values in $\mathbb{R}$ endowed with its Borel σ-field $\mathfrak{B}$ 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 $I : \mathbb{R} \to \mathbb{R}^+$ be a lower semicontinuous function, that is for all $c > 0$ the sub-level set $\{x \in \mathbb{R}, I(x) \leq c\}$ is a closed set. Such a function I is called *rate function* and it is called *good rate function* if all its sub-level sets are compact sets. Let $(a_n)_{n\geq 0}$ be a positive sequence such that $a_n \to +\infty$ and $a_n/s_n \to 0$ as n goes to $+\infty$.

Definition 2 (Moderate deviation principle, MDP)**.** *We say that $Z_n/\sqrt{a_n s_n}$ satisfies a moderate deviation principle in $\mathbb{R}$ with speed a_n and the rate function I if, for any $A \in \mathfrak{B}$,*

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

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

Our objective is to prove such a MDP for the estimators $\widehat{f}_0(x)$ and $\widehat{f}_1(x)$.

2. Moderate deviation principle

2.1. **Model contraints.** The autoregressive functions f_0 and f_1 will be restricted to belong to the following class. For $\ell > 0$, we introduce the class $\mathcal{F}_\ell$ of bounded functions $f : \mathbb{R} \to \mathbb{R}$ such that $|f|_\infty = \sup_{x\in\mathbb{R}} |f(x)| \leq \ell$.

The two marginals of the noise density $G_0(\cdot) = \int_{\mathbb{R}} G(\cdot, y)dy$ and $G_1(\cdot) = \int_{\mathbb{R}} G(x, \cdot)dx$ 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 $g : \mathbb{R} \to [0, \infty)$ such that

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

for any $x \in \mathbb{R}$.

For any $R > 0$, we set

$$\delta(R) = \min\Big\{ \inf_{|x|\leq R} G_0(x); \inf_{|x|\leq R} G_1(x)\Big\} \tag{2}$$

and

$$\eta(R) = \frac{|G_0|_\infty + |G_1|_\infty}{2} \int_{|y|>R} \int_{x\in\mathbb{R}} \frac{r}{1 + \big|y - \gamma|x| - \ell\big|^\lambda \wedge \big|y + \gamma|x| + \ell\big|^\lambda} dxdy.$$

To guarantee we have geometric ergodicity of the tagged-branch Markov chain (see Guyon [10] or Section 3.1 below), uniformly in the initial value, with an exponential decay rate smaller than 1/2, we will require the following assumption (see Lemma 9).

Assumption 3. *There exists $R_3 > \ell$ such that $2(R_3 - \ell)\delta(R_3) > 1/2$ with $\delta(\cdot)$ defined by (2).*

The following assumption will guarantee that the invariant density ν of the tagged-branch chain is positive on some nonempty interval (see Lemma 17 in [5]).

Assumption 4. *For $R_1 > 0$ such that $\eta(R_1) < 1$, there exists $R_2 > \ell + \gamma R_1$ such that $\delta(R_2) > 0$.*

We also reinforce usual assumptions on the kernel.

Assumption 5. *The kernel $K : \mathbb{R} \to \mathbb{R}$ is bounded with compact support and for some integer $n_0 \geq 1$, we have $\int_{-\infty}^{\infty} x^k K(x)dx = \mathbf{1}_{\{k=0\}}$ for $k = 0, \ldots, n_0$. In addition, $K : \mathbb{R} \to \mathbb{R}$ is such that*

$$\int_{\mathbb{R}} |K^+(x)|dx > \frac{|G_0|_\infty + |G_1|_\infty}{2\delta(R_2)\big(1-\eta(R_1)\big)} \int_{\mathbb{R}} |K^-(x)|dx$$

where R_1, R_2 come from Assumption 4 and $K^+(\cdot) = \max\{K(\cdot); 0\}$, $K^-(\cdot) = \min\{K(\cdot); 0\}$.

Note that the strict inequality required in Assumption 5 is valid for any nonnegative kernel – since $\int_{\mathbb{R}} |K^-(x)|dx = 0$ for such kernels. Thus the triangle kernel, the Epanechnikov kernel *e.g.* satisfie Assumption 5.

To finish, we introduce smooth functions described in the following way: for $\mathcal{X} \subseteq \mathbb{R}$ and $\beta > 0$, with $\beta = \lfloor\beta\rfloor + \{\beta\}$, $0 < \{\beta\} \leq 1$ and $\lfloor\beta\rfloor$ an integer, let $\mathcal{H}^\beta_{\mathcal{X}}$ denote the Hölder space of functions $h : \mathcal{X} \to \mathbb{R}$ possessing a derivative of order $\lfloor\beta\rfloor$ that satisfies $|h^{\lfloor\beta\rfloor}(y) - h^{\lfloor\beta\rfloor}(x)| \leq c(h)|x - y|^{\{\beta\}}$. The minimal constant $c(h)$ such that the previous inequality holds defines a semi-norm $|g|_{\mathcal{H}^\beta_{\mathcal{X}}}$. We equip the space $\mathcal{H}^\beta_{\mathcal{X}}$ with the norm $\|h\|_{\mathcal{H}^\beta_{\mathcal{X}}} = \sup_x |h(x)| + |h|_{\mathcal{H}^\beta_{\mathcal{X}}}$ and the balls $\mathcal{H}^\beta_{\mathcal{X}}(L) = \{h : \mathcal{X} \to \mathbb{R},\ \|h\|_{\mathcal{H}^\beta_{\mathcal{X}}} \leq L\},\ L > 0$.

2.2. **Main result.** We are now ready to state a moderate deviation principle (in the sense of Definition 2) for the estimators $\widehat{f}_0(x)$ and $\widehat{f}_1(x)$ in this framework.

Theorem 6. *Let $(b_n)_{n\geq 0}$ be a positive sequence such that*

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

Let $\ell>0$, $r>0$ and $\lambda>5$. Specify $(\widehat{f}_{0,n},\widehat{f}_{1,n})$ with a kernel K satisfying Assumption 5 for some $n_0>0$, with $h_n\propto|\mathbb{T}_n|^{-\alpha}$ for $\alpha\in(1/(2\beta+1),1)$, and with $\varpi_n>0$ such that $\varpi_n\to 0$ as $n\to+\infty$. For any initial probability measure $\mu(dx)$ on $\mathbb{R}$ for X_1, for every $L,L'>0$ and $0<\beta<n_0$, for every G such that $(G_0,G_1)\in\big(\mathcal{G}_{r,\lambda}\cap\mathcal{H}^{\beta}_{\mathbb{R}}(L')\big)^2$ satisfy Assumptions 3 and 4, there exists $d=d(\ell,G)>0$ such that for every compact interval $\mathcal{D}\subset[-d,d]$ with nonempty interior, for every x in the interior of $\mathcal{D}$ and for every functions $(f_0,f_1)\in\big(\mathcal{F}_\ell\cap\mathcal{H}^{\beta}_{\mathcal{D}}(L)\big)^2$, 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 $\mathbb{R}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and good rate function $J_x:\mathbb{R}^2\to\mathbb{R}$ defined by

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

with Γ the variance-covariance matrix of $(\varepsilon_1,\varepsilon_2)$, where $\boldsymbol{z}^t$ stands for the transpose of vector $\boldsymbol{z}$.

The notation $\propto$ means proportional to up to some positive constant. The contraction principle (see the textbook of Dembo and Zeitouni [7], Chapter 4) enables us to state the following corollary of Theorem 6.

Corollary 7. *In the same setting as Theorem 6, for every $\delta>0$,*

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

with σ_ι^2 denoting the variance of $\varepsilon_{1+\iota}$.

Remark 8. *Let us mention that recently, Bitseki et al. [3] have establish transportation inequality for bifurcating Markov chains. This has allowed them to obtain deviation inequalities for a large class of functions. However, their results do not allow to obtain MDP with optimal speeds as in Theorem 6.*

3. Proofs

3.1. **Preliminaries.** A key object in the study of NBAR processes (particular case of a bifurcating Markov chain, see Guyon [10] for a general definition) is the so-called tagged-branch Markov chain. Consider the Markov chain chain $(Y_m)_{m\geq 0}$ taking values in $\mathbb{R}$, such that $Y_0=X_1$, with transition

$$Q(x,dy)=\frac{1}{2}\Big(G_0\big(y-f_0(x)\big)+G_1\big(y-f_1(x)\big)\Big)dy,\quad x\in\mathbb{R}. \tag{3}$$

It corresponds to a lineage taken randomly (uniformly at each branching event) in the population. In the setting of Theorem 6, we achieve uniform ergodicity for the tagged-branch Markov chain Y as stated in the following

Lemma 9 (Uniform ergodicity)**.** *For every $(f_0, f_1) \in \mathcal{F}_\ell^2$, for every G such that $(G_0, G_1) \in \mathcal{G}_{r,\lambda}^2$ satisfy Assumption 3, the Markov kernel Q admits a unique invariant probability measure ν of the form $\nu(dx) = \nu(x)dx$ on $(\mathbb{R}, \mathfrak{B})$. Moreover, for every G such that $(G_0, G_1) \in \mathcal{G}_{r,\lambda}^2$ satisfy Assumption 3, there exist a constant $R > 0$ and $\rho \in (0, 1/2)$ such that*

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

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

The proof is postponed to the Appendix (Section 4). Uniform ergodicity achieved by Lemma 9 is here crucial in order to make use of deviations inequalities as obtained in Bitseki Penda, Hoffmann and Olivier [4], one key tool for the proof of Theorem 6.

We will intensively use the following two concepts: super-exponential convergence and exponential equivalence. Let $(Z_n)_{n\geq 0}$ be a sequence of random variables and Z a random variable with values in $\mathbb{R}$ endowed with its Borel σ-field $\mathfrak{B}$.

Definition 10 (Super-exponential convergence)**.** *We say that $(Z_n)_{n\geq 0}$ converges $(s_n)_{n\geq 0}$-super-exponentially fast in probability to Z and we note $Z_n \xRightarrow[s_n]{\text{superexp}} Z$ if*

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

for any $\delta > 0$.

Let $(W_n)_{n\geq 0}$ be another sequence of random variables with values in $\mathbb{R}$.

Definition 11 (Exponential equivalence, see [7])**.** *We say that $(Z_n)_{n\geq 0}$ and $(W_n)_{n\geq 0}$ are $(s_n)_{n\geq 0}$-exponentially equivalent and we note $Z_n \overset{\text{superexp}}{\underset{s_n}{\sim}} W_n$ if*

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

for any $\delta > 0$.

3.2. **Proof of Theorem 6.** Set x in the interior of $\mathcal{D}$. In order to prove the moderate deviation 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, $\widehat{\nu}_n(x) = |\mathbb{T}_n|^{-1} \sum_{k\in\mathbb{T}_n} K_{h_n}(x - X_k)$, and for $\iota \in \{0, 1\}$,

$$M_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_n} K_{h_n}(x - X_k)\varepsilon_{2k+\iota}, \tag{4}$$

$$N_{\iota,n}(x) = \frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_n} K_{h_n}(x - X_k)\big(f_\iota(X_k) - f_\iota(x)\big), \tag{5}$$

$$R_{\iota,n}(x) = \Big(\widehat{\nu}_n(x) - \big(\widehat{\nu}_n(x)\vee\varpi_n\big)\Big) f_\iota(x), . \tag{6}$$

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

$$\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}\widehat{f}_{0,n}(x)-f_0(x)\\ \widehat{f}_{1,n}(x)-f_1(x)\end{pmatrix} \underset{b_n^2/(|\mathbb{T}_n|h_n)}{\overset{\text{superexp}}{\sim}} \frac{1}{\widehat{\nu}_n(x)\vee\varpi_n}\frac{|\mathbb{T}_n|h_n}{b_n}\begin{pmatrix}M_{0,n}(x)\\ M_{1,n}(x)\end{pmatrix} \tag{7}$$

in the sense of Definition 11. Consequently, these two quantities satisfy the same MDP (see Dembo and Zeitouni [7], Chapter 4) and we prove a MDP for the right-hand side of (7) in Step 3.

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

We claim that

$$\widehat{\nu}_n(x) \vee \varpi_n \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \tag{8}$$

On the one hand, we know that $K_{h_n} \star \nu(x) \to \nu(x)$ as $n \to +\infty$[1]. Since the previous sequence is deterministic, we conclude that

$$K_{h_n} \star \nu(x) \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \tag{9}$$

On the other hand, from Theorem 4 (ii) of [4],

$$\mathbb{P}_\mu\Big(\Big|\frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_n} K_{h_n}(x-X_k) - K_{h_n}\star\nu(x)\Big| > \delta\Big) \le 2\exp\Big(\frac{-C_1\delta^2|\mathbb{T}_n|h_n}{1+\delta}\Big) \tag{10}$$

where C_1 is a positive constant which depends on K and Q but not on n. Note that Lemma 9 ensures we are in the setting of [4]. Applying the log to the two terms of (10), multiplying by $|\mathbb{T}_n|h_n/b_n^2$ and letting n go to infinity, we are led to

$$\frac{1}{|\mathbb{T}_n|}\sum_{k\in\mathbb{T}_n} K_{h_n}(x-X_k) - K_{h_n}\star\nu(x) \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0 \tag{11}$$

using conditions (i) and (ii) on the sequence b_n. From the foregoing (9) and (11), we obtain

$$\widehat{\nu}_n(x) \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} \nu(x). \tag{12}$$

To reach (8), 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

$$\mathbf{1}_{\{\widehat{\nu}_n(x)<\varpi_n\}} \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{13}$$

We have

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

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, using $\inf_{(f_0,f_1)}\inf_{x\in\mathcal{D}}\nu(x) > 0$ (under Assumption 5, see Lemma 17 of [5]) and $\varpi_n \to 0$. Using the deviations inequality (10) with δ', we obtain (13). We finally have (8), which ends the first step.

[1]By a Taylor expansion of ν up to order $\lfloor\beta\rfloor$ (note that ν has the same regularity as the noise density G and recall that the number n_0 of vanishing moments of K in Assumption 5 satisfies $n_0 > \beta$), we obtain

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

up to some positive constant independent of n, see for instance Proposition 1.2 in Tsybakov [17].

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

We claim that

$$\frac{|\mathbb{T}_n|h_n}{b_n} N_{\iota,n}(x) \xrightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{14}$$

We use decomposition $N_{\iota,n}(x) = N_{\iota,n}^{(1)}(x) + N_{\iota,n}^{(2)}(x)$ with

$$N_{\iota,n}^{(1)}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{k\in\mathbb{T}_n} \mathbb{E}_\nu\big[K_{h_n}(x-X_k)\big(f_\iota(X_k)-f_\iota(x)\big)\big], \tag{15}$$

$$N_{\iota,n}^{(2)}(x) = \frac{1}{|\mathbb{T}_n|} \sum_{k\in\mathbb{T}_n} \Big(K_{h_n}(x-X_k)\big(f_\iota(X_k)-f_\iota(x)\big) - \mathbb{E}_\nu\big[K_{h_n}(x-X_k)\big(f_\iota(X_k)-f_\iota(x)\big]\Big). \tag{16}$$

On the one hand, one can check that $|N_{\iota,n}^{(1)}(x)| \lesssim h_n^\beta$ (see [5], Proof of Theorem 8, Step 1.1). 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 *e.g.* Worms [19] for more details. On the other hand, from Theorem 4 (ii) of [4], slightly refined[2], the following deviations inequality holds

$$\mathbb{P}_\mu\big(|N_{\iota,n}^{(2)}(x)| > \delta\big) \le 2\exp\Big(\frac{-C_2\delta^2|\mathbb{T}_n|h_n}{\big(h_n^\beta + h_n^{2\beta+1}|\mathbb{T}_n| + (|\mathbb{T}_n h_n|)^{-1}\big)+\delta}\Big)$$

where C_2 is a positive constant which depends on K and Q but not on n. Recalling conditions (i) and (ii) on b_n and $h_n \propto |\mathbb{T}_n|^{-\alpha}$ with $\alpha \in (1/(2\beta+1), 1)$, 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 (14) 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 $R_{\iota,n}(x) = \big(\widehat{\nu}_n(x) - \varpi_n\big) f_\iota(x)\mathbf{1}_{\{\widehat{\nu}_n(x)<\varpi_n\}}$ and recalling (12) and (13). 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 (7).

Step 3. Main term $M_{\iota,n}(x)$.

First, we introduce the filtration $\mathcal{G} = (\mathcal{G}_m^{(n)}; n \ge 0, m \le |\mathbb{T}_n|)$, where for all $n \ge 0$,

$$\mathcal{G}_0^{(n)} = \sigma(X_1) \quad \text{and } \forall 1 \le m \le |\mathbb{T}_n|, \quad \mathcal{G}_m^{(n)} = \sigma\big((X_k, X_{2k}, X_{2k+1}), 1 \le k \le m\big).$$

[2]One has to note that the inequality still holds with the refinement $\Sigma_{1,n}(g) = |Qg^2|_\infty + \min_{1\le\ell\le n-1}|Qg|_\infty^2 2^\ell + |g|_\infty^2 2^{-\ell}$, see equation (3) and Theorem 4 in [4]. In our case, with the test function $g : y \rightsquigarrow h_n^{-1}K\big(h_n^{-1}(x-y)\big)\big(f_\iota(y)-f_\iota(x)\big)$ for x fixed, $|Qg^2|_\infty$ is of order $h_n^{\beta-1}$, $|Qg|_\infty^2$ is of order $h_n^{2\beta}$ and $|g|_\infty^2$ is of order h_n^{-2}.

We then consider the triangular array of bivariate random variables $(\overline{E}_k^{(n)}(x))$ defined by

$$\overline{E}_k^{(n)}(x) = \sum_{l=1}^{k} E_l^{(n)}(x) \tag{17}$$

where for $l \leq |\mathbb{T}_n|$

$$E_l^{(n)}(x) = (|\mathbb{T}_n|h_n)^{-1/2} \begin{pmatrix} K\big(h_n^{-1}(x-X_l)\big)\varepsilon_{2l} \\ K\big(h_n^{-1}(x-X_l)\big)\varepsilon_{2l+1} \end{pmatrix}.$$

Note that $\big(\overline{E}_k^{(n)}(x); n \geq 0, 1 \leq k \leq |\mathbb{T}_n|\big)$ is a $\mathcal{G}$-martingale triangular arrays whose bracket is given by

$$\begin{aligned} \langle \overline{E}^{(n)}(x)\rangle_k &= \sum_{l=1}^{k} \mathbb{E}\left[E_l^{(n)}(x)\left(E_l^{(n)}(x)\right)^t \Big| \mathcal{G}_{l-1}^{(n)}\right] \\ &= \left(\frac{1}{|\mathbb{T}_n|h_n}\sum_{l=1}^{k} K^2\left(\frac{X_l - x}{h_n}\right)\right)\Gamma. \end{aligned}$$

In order to make use of the MDP for martingale triangular arrays (see Worms [19, 20] or Puhalskii [16]), one need to check

$$\langle \overline{E}^{(n)}(x)\rangle_{|\mathbb{T}_n|} \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} |K|_2^2\nu(x)\Gamma, \tag{18}$$

and

$$\sum_{k\in\mathbb{T}_n} \mathbb{E}\left[\|\overline{E}_k^{(n)} - \overline{E}_{k-1}^{(n)}\|^4 \Big| \mathcal{G}_{k-1}\right] \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{19}$$

The condition (19) is called exponential Lyapunov condition and it implies exponential Lindeberg condition, we refer to Worms [19, 20] for more details. One can easily check that it suffices to show

$$\frac{1}{|\mathbb{T}_n|h_n}\sum_{k\in\mathbb{T}_n} K^2\big(h_n^{-1}(x-X_k)\big) \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} |K|_2^2\nu(x) \tag{20}$$

and

$$\frac{1}{(|\mathbb{T}_n|h_n)^2}\sum_{k\in\mathbb{T}_n} K^4\big(h_n^{-1}(x-X_k)\big) \xRightarrow[b_n^2/(|\mathbb{T}_n|h_n)]{\text{superexp}} 0. \tag{21}$$

The same argument as in Step 1 (replacing K by $|K|_2^{-2}K^2$ or $|K^2|_2^{-2}K^4$ in (10)) enables us to prove (20) and (21). Gathering (18) and (19) and using the truncation of the martingale $(\overline{E}_k^{(n)}(x))$ as in the proof of Theorem 5.1 of Bitseki Penda and Djellout [2], 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 $\mathbb{R}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and the good rate function defined for all $\boldsymbol{z} \in \mathbb{R}^2$ by

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

Finally, using Lemma 4.1 of Worms [18] (which is a consequence of the contraction principle) and Step 1, we conclude that

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

satisfies a MDP on $\mathbb{R}^2$ with speed $b_n^2/(|\mathbb{T}_n|h_n)$ and rate function J_x defined in Theorem 6. Reminding (7), we get the MDP stated in Theorem 6.

4. Appendix

Proof of Lemma 9. Set $\mathcal{C} = \{y \in \mathbb{R}; |y| \leq R_3 - \ell\} \neq \emptyset$ since $R_3 > \ell$ under Assumption 3. We prove that, for all $y \in \mathcal{C}$,

$$\inf_{x\in\mathbb{R}} Q(x,y) = \frac{1}{2}\Big\{ \inf_{x\in\mathbb{R}} G_0\big(y - f_0(x)\big) + \inf_{x\in\mathbb{R}} G_1\big(y - f_1(x)\big)\Big\} \geq \delta(R_3) > 0$$

recalling that Q is defined by (3) and $\delta(\cdot)$ by (2), since $|y - f_\iota(x)| \leq R_3$ (for $|f_\iota|_\infty \leq \ell$ and $|y| \leq R_3 - \ell$). Thus Q satisfies the Doeblin condition (see Definition 6.10 of Douc *et al.* [9]) with constant $|\mathcal{C}|\delta(R_3) > 0$. By Lemma 6.10 of [9], we are in position to apply Theorem 6.6 of [9]. Thus the Markov kernel Q admits a unique invariant probability measure and one can take $\rho = 1 - |\mathcal{C}|\delta(R_3) = 1 - 2(R_3 - \ell)\delta(R_3)$ for the exponential decay rate. Under Assumption 3, we have $\rho < 1/2$. □

Acknowledgements. We thank G. Fort for suggesting us the reading of the recent textbook [9].

References

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

[2] 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.

[3] S. V. Bitseki Penda, M. Escobar-Bach and A. Guillin. *Transportation cost-information and concentration inequalities for bifurcating Markov chains.* Bernoulli, 23 (2017), 3213–3242

[4] S. V. Bitseki Penda, M. Hoffmann and A. Olivier. *Adaptive estimation for bifurcating Markov chains.* Bernoulli, 23 (2017), 3598–3637

[5] S. V. Bitseki Penda and A. Olivier. *Autoregressive functions estimation in nonlinear bifurcating autoregressive models.* Statistical Inference for Stochastic Processes, 20 (2017), 179–210

[6] R. Cowan and R. G. Staudte. *The bifurcating autoregressive model in cell lineage studies.* Biometrics, 42 (1986), 769–783.

[7] A. Dembo and O. Zeitouni. *Large deviations techniques and applications, second edition.* Second edition, volume 38 of *Applications of Mathematics*, Springer, 1998.

[8] B. de Saporta, A. Gégout-Petit and L. Marsalle. *Random coefficients bifurcating autoregressive processes.* ESAIM: Probability and Statistics, 18 (2014), 365–399.

[9] R. Douc, E. Moulines and D. S. Stoffer. *Nonlinear Time Series.* Chapman & Hall/CRC Texts in Statistical Science, 2014.

[10] J. Guyon. *Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging.* The Annals of Applied Probability, 17 (2007), 1538–1569.

[11] C. Joutard. *Sharp large deviations in nonparametric estimation.* Nonparametric Statistics, 18 (2006), 293–306.

[12] D. Louani. *Large deviations limit theorems for the kernel density estimator.* Scandinavian Journal of Statistics, 25 (1998), 243–253.

[13] D. Louani. *Some large deviations limit theorems in conditional nonparametric statistics.* Statistics: A Journal of Theoretical and Applied Statistics, 33 (1999), 171–196.

[14] A. Mokkadem and M. Pelletier. *Confidence bands for densities, logarithmic point of view.* Alea, 2 (2006), 231–266.

[15] A. Mokkadem, M. Pelletier and B. Thiam. *Large and moderate deviation principles for kernel estimators of the multivariate regression.* Mathematical Methods of Statistics, 17 (2008), 1–27.

[16] A. Puhalskii. *Large deviations of semimartingales via convergence of the predictable characteristics.* Stochastics and Stochastics Reports, 49 (1994), 27–85.

[17] A. Tsybakov. *Introduction to Nonparametric Estimation.* Springer series in statistics, Springer-Verlag, New-York, 2009.

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

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

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

S. Valère Bitseki Penda, IMB, CNRS-UMR 5584, Université Bourgogne Franche-Comté, 9 avenue Alain Savary, 21078 Dijon Cedex, France.

E-mail address: simeon-valere.bitseki-penda@u-bourgogne.fr

Adélaïde Olivier, Département de Mathématiques - Batiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud, F-91405 ORSAY Cedex

E-mail address: adelaide.olivier@u-psud.fr