Statistics and Probability Letters 138 (2018) 20–26

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Moderate deviation principle in nonlinear bifurcating autoregressive models

S. Valère Bitseki Penda[a,*], Adélaïde Olivier[b]

[a] IMB, CNRS-UMR 5584, Université Bourgogne Franche-Comté, 9 avenue Alain Savary, 21078 Dijon Cedex, France
[b] Laboratoire de Mathématiques d'Orsay, Univ. Paris-Sud, CNRS, Université Paris-Saclay, 91405 Orsay, France

ARTICLE INFO

Article history:
Received 10 July 2017
Received in revised form 5 February 2018
Accepted 13 February 2018
Available online 27 February 2018

MSC:
62G05
62G20
60J80
60F05
60F10

Keywords:
Bifurcating Markov chain
Binary tree
Bifurcating autoregressive process
Nonparametric estimation
Nadaraya–Watson estimator
Moderate deviation principle

ABSTRACT

Recently, nonparametric techniques have been proposed to study bifurcating autoregressive processes. One can build Nadaraya–Watson type estimators of the two autoregressive functions as in Bitseki Penda et al. (2017) and Bitseki Penda and Olivier (2017). In the present work, we prove moderate deviation principle for these estimators.

1. Introduction

1.1. A generalization of BAR processes

Bifurcating autoregressive (BAR) processes were introduced by Cowan and Staudte (1986) in 1986 to study *E. coli* bacteria. Since then it has been extensively studied. We refer in particular to the recent works of Bercu and Blandin (2015), de Saporta et al. (2014), see also references therein. Nonlinear bifurcating autoregressive (NBAR) processes, studied in Bitseki Penda et al. (2017b); Bitseki Penda and Olivier (2017), 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 $\mathbb{G}_m = \{2^m, \dots, 2^{m+1}-1\}$ be the mth generation. A given individual $k \geq 1$ lives in the r_kth 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

* Corresponding author.
E-mail addresses: simeon-valere.bitseki-penda@u-bourgogne.fr (S.V. Bitseki Penda), adelaide.olivier@u-psud.fr (A. Olivier).

https://doi.org/10.1016/j.spl.2018.02.037
0167-7152/

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

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

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 in Bitseki Penda and Olivier (2017).

1.3. Objective

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 (1998) and Joutard (2006) in the univariate case). We refer to Mokkadem and Pelletier (2006) 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, for any $x \in \mathbb{R}$, $g(x) \leq \frac{r}{1+|x|^\lambda}$. 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

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