



**HAL**  
open science

## Moderate deviations of functional of Markov Processes

Valère Bitseki Penda, Hacène Djellout, Laure Dumaz, Florence Merlevède,  
Frédéric Proïa

► **To cite this version:**

Valère Bitseki Penda, Hacène Djellout, Laure Dumaz, Florence Merlevède, Frédéric Proïa. Moderate deviations of functional of Markov Processes. ESAIM: Proceedings, 2014, 44, p. 214. hal-00872322

**HAL Id: hal-00872322**

**<https://hal.science/hal-00872322>**

Submitted on 14 Oct 2013

**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 DEVIATIONS OF FUNCTIONAL OF MARKOV PROCESSES

S.VALÈRE BITSEKI PENDA<sup>1</sup>, HACÈNE DJELLOUT<sup>1</sup>, LAURE DUMAZ<sup>2</sup>, FLORENCE  
MERLEVÈDE<sup>3</sup> AND FRÉDÉRIC PROÏA<sup>4</sup>

**Abstract.** This paper presents recent developments on the principle of moderate deviations for some classes of dependent random variables.

**Résumé.** Ce papier présente les récents développements sur le principe des déviations modérées d'une certaine classe de variables aléatoires dépendantes.

### INTRODUCTION

This paper groups the contributions of the speakers of the session dedicated to moderate deviations of functional of Markov processes organized during the Journées MAS, which took place in Clermont-Ferrand in August 2012.

The principle of moderate deviations (PDM, in short) is a subject of classic study of the probability theory. Indeed, in the study of the limit theorems of a probability or statistical model, the PDM is one of main questions that we look, after the laws of large numbers, the central limit theorem (CLT, in short) and the law of the iterated logarithm.

The MDP can be seen as an intermediate behavior between the CLT and large deviations principle (LDP, in short). Usually, the MDP exhibit a simpler rate function (quadratic) inherited from the approximated Gaussian process, and holds for a larger class of dependent random variables than the LDP.

The LDP and MDP of sums of random variables is now a wide and fastly growing branch of probability theory. It was created initially in the framework of the theory of sums of independent identically distributed random variables and then extended to a wide class of random processes, i.e., random functions in one variable, with some general conditions of weak dependence traditional for the theory of random processes.

We refer to Dembo and Zeitouni [13], for an exposition of the general theory of large deviations and limit ourself below to the statement of some important facts and definitions which are useful for our needs.

It is best to think of a specific example to clarify the idea

---

<sup>1</sup> Laboratoire de Mathématiques, CNRS UMR 6620, Université Blaise Pascal, 24 avenue des Landais, BP 80026, 63177 Aubière ; e-mail: Valere.Bitsekipenda@math.univ-bpclermont.fr & Hacene.Djellout@math.univ-bpclermont.fr

<sup>2</sup> École Normale Supérieure, Département de Mathématiques et Applications, 75230 Paris cedex 05, & Université Paris XI & BME (Budapest) ; e-mail: laure.dumaz@ens.fr

<sup>3</sup> Université Paris-Est, LAMA, CNRS UMR 8050, Bâtiment Copernic, 5 Boulevard Descartes, 77435 Champs-Sur-Marne ; e-mail: florence.merlevede@univ-mlv.fr

<sup>4</sup> Université de Bordeaux 1, Institut de Mathématiques de Bordeaux, UMR 5251, and INRIA Bordeaux, team ALEA, 351 Cours de la Libération, 33405 Talence cedex ; e-mail: frederic.proia@inria.fr

**Example 0.1.** (Cramér-Chernoff's Theorem) Let  $X_1, X_2, \dots$  a sequence of i.i.d. centered real valued random variables. Define  $S_n = \sum_{i=1}^n X_i$ . Notice that

$$L(t) := n^{-1} \log \mathbb{E}(e^{tS_n}) = \log \mathbb{E}(e^{tX_1}).$$

Whence, by the Gärtner-Ellis theorem, for all Borelians  $A$ ,

$$- \inf_{t \in A^\circ} I(t) \leq \liminf_n \frac{1}{n} \log \mathbb{P} \left( \frac{S_n}{n} \in A \right) \leq \limsup_n \frac{1}{n} \log \mathbb{P} \left( \frac{S_n}{n} \in A \right) \leq - \inf_{t \in \bar{A}} I(t),$$

where  $I$  is the Fenchel-Legendre dual of  $L$  given by  $I(x) = \sup_{t>0} (tx - \log \mathbb{E}(e^{tX_1}))$ .

**Remark 0.2.** The LDP is not distribution free, it holds for a small class of dependent sequences (see Bryc-Dembo [6]), it requires the existence of moment generating function and, it is restricted to stationarity.

**Example 0.3.** (Heuristic for the MDP). Consider again  $(X_i)_{i \geq 1}$  a sequence of i.i.d. centered real valued random variables. Take  $a_n \rightarrow 0$  and  $na_n \rightarrow \infty$  (instead of  $a_n = n^{-1}$ ). Make blocks  $Y_{k,n}$  of size  $[na_n]$ . Hence  $S_n \sim Y_{1,n} + \dots + Y_{[a_n^{-1}],n}$ , and

$$a_n \log \mathbb{E}(e^{tS_n/\sqrt{na_n}}) \sim a_n \log \left( \mathbb{E}(e^{tY_{1,n}/\sqrt{na_n}}) \right)^{1/a_n}.$$

Under conditions, via the convergence of moments in the CLT,

$$\log \mathbb{E}(e^{tS_{[na_n]}/\sqrt{na_n}}) \rightarrow \frac{\sigma^2 t^2}{2},$$

where  $\sigma^2 = \mathbb{E}(X_1^2)$ . Therefore, by the Gärtner-Ellis theorem, for all Borelians  $A$ ,

$$- \inf_{t \in A^\circ} I(t) \leq \liminf_n a_n \log \mathbb{P} \left( \sqrt{a_n} \frac{S_n}{\sqrt{n}} \in A \right) \leq \limsup_n a_n \log \mathbb{P} \left( \sqrt{a_n} \frac{S_n}{\sqrt{n}} \in A \right) \leq - \inf_{t \in \bar{A}} I(t),$$

where  $I(x) = x^2/(2\sigma^2)$ .

In the i.i.d. setting, we refer to Arcones [1,2], and to Eichelsbacher-Löwe [16](see Theorem 2.14 in section 2) for necessary and sufficient conditions for the MDP to hold.

**Remark 0.4.** MDP is distribution free, it holds for a larger class of dependent sequences, it does not require the existence of the moment generating function, it is non restricted to stationarity.

Before presenting the plan of our paper, let us now give the precise definition of a MDP: let  $(a_n)_{n \geq 0}$  be a positive sequence such that

$$a_n \xrightarrow[n \rightarrow \infty]{} 0, \quad \text{and} \quad na_n \xrightarrow[n \rightarrow \infty]{} \infty. \quad (1)$$

**Definition 0.5.** A family of random variables  $\{Z_n, n > 0\}$  with values in a topological space  $\mathcal{X}$  equipped with  $\sigma$ -field  $\mathcal{B}$  satisfies the MDP with speed  $a_n$  satisfying (1) and good rate function  $I(\cdot)$  if the level sets  $\{x, I(x) \leq \alpha\}$  are compact for all  $\alpha < \infty$ , and for all  $\Gamma \in \mathcal{B}$

$$- \inf_{t \in \Gamma^\circ} I(t) \leq \liminf_{n \rightarrow \infty} a_n \log \mathbb{P}(\sqrt{a_n} Z_n \in \Gamma) \leq \limsup_{n \rightarrow \infty} a_n \log \mathbb{P}(\sqrt{a_n} Z_n \in \Gamma) \leq - \inf_{t \in \bar{\Gamma}} I(t),$$

where  $\Gamma^\circ$  (resp.  $\bar{\Gamma}$ ) denotes the interior (resp. closure) of  $\Gamma$  in  $\mathcal{X}$ .

The paper is organized as follows. In Section 1, we state the MDP for weakly dependent random variables with applications to functions of mixing sequences, Markov chains, and functions of linear process. Section 2 is

devoted to the MDP for the Durbin-Watson statistic related to the first-order autoregressive process. Section 3 is dedicated to the MDP for bifurcating Markov chains and application to the bifurcating autoregressive process. Section 4 is very different from subjects approached on the previous sections. It concerns essentially the large deviations of the true self-repelling motion.

## 1. MODERATE DEVIATIONS FOR WEAKLY DEPENDENT SEQUENCES

In this section, we are interested in the MDP and its functional form for a class of weakly dependent sequences. Examples that can be treated this way include some classes of Markov chains, iterated Lipschitz models and functions of linear processes with absolutely regular innovations.

Concerning the traditional LDP, it is known from the paper by Bryc and Dembo [6] that it is not satisfied by many classes of weakly dependent random variables. This is the reason why it is convenient to look at MDP.

We shall use the new developed Bernstein-type inequalities to obtain sharp moderate deviation asymptotic results for some classes of dependent random variables.

Given  $X_1, X_2, \dots$  a sequence of centered real valued random variables, our aim is to give dependence conditions to get the MDP for the partial sum and the normalized partial sum processes

$$S_n = \sum_{i=1}^n X_i \quad \text{or} \quad W_n(\cdot) = \left\{ \sum_{i=1}^{\lfloor nt \rfloor} X_i, t \in [0, 1] \right\}$$

suitably normalized ( $W_n$  is an element of  $\mathbb{D}[0, 1]$ , the space of functions on  $[0, 1]$  with left-hand limits and continuous from the right, equipped with the Skorohod topology).

### 1.1. MDP under projective conditions

What projective conditions can we expect? Let us recall two results about the functional form of the CLT for stationary sequences. With this aim, it is convenient to define a stationary sequence  $(X_i)_{i \in \mathbb{Z}}$  as follows. Let  $\theta : \Omega \mapsto \Omega$  be a bijective bimeasurable transformation preserving  $\mathbb{P}$  on  $(\Omega, \mathcal{A})$ . For any  $i \in \mathbb{Z}$ , let  $X_i = X_0 \circ \theta^i$  where  $X_0$  is a real-valued random variable defined on  $(\Omega, \mathcal{A})$ . For a subfield  $\mathcal{F}_0$  satisfying  $\mathcal{F}_0 \subseteq \theta^{-1}(\mathcal{F}_0)$ , let  $\mathcal{F}_i = \theta^{-i}(\mathcal{F}_0)$ . Denote also by  $\mathcal{I}$  the  $\theta$ -invariant sigma field.

**Theorem 1.1.** [Maxwell-Woodroffe [21]. Peligrad-Utev [27]]. *Assume that  $X_0$  is  $\mathcal{F}_0$ -measurable, in  $\mathbb{L}^2$  and that*

$$\sum_{n>0} n^{-3/2} \|\mathbb{E}(S_n | \mathcal{F}_0)\|_2 < \infty.$$

*Then  $\{n^{-1/2}W_n(t), t \in [0, 1]\}$  converges in distribution in  $\mathbb{D}[0, 1]$  to  $\sqrt{\eta}W$  where  $W$  is a standard Brownian motion independent of  $\mathcal{I}$  and  $\eta$  is a  $\mathcal{I}$ -measurable nonnegative r.v. such that  $\mathbb{E}(\eta) = \sigma^2$  and*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{E}(S_n^2 | \mathcal{I})}{n} = \eta \quad \text{in } \mathbb{L}^1.$$

**Theorem 1.2.** [Heyde [20]. Dedecker-Merlevède-Volný [7]]. *Assume that  $X_0$  is  $\mathcal{F}_0$ -measurable, in  $\mathbb{L}^2$  and such that  $\mathbb{E}(X_0 | \mathcal{F}_{-\infty}) = 0$  a.s. Assume that*

$$\sum_{n \geq 0} \|P_0(X_n)\|_2 < \infty \quad \text{where} \quad P_0(X_n) = \mathbb{E}(X_n | \mathcal{F}_0) - \mathbb{E}(X_n | \mathcal{F}_{-1}).$$

*Then  $\{n^{-1/2}W_n(t), t \in [0, 1]\}$  converges in distribution in  $\mathbb{D}[0, 1]$  to  $\sqrt{\eta}W$  where  $W$  is a standard Brownian motion independent of  $\mathcal{I}$  and*

$$\eta = \sum_{k \in \mathbb{Z}} \mathbb{E}(X_0 X_k | \mathcal{I}).$$

**Remark 1.3.** The theorems have different ranges of applicability. Both of them are satisfied under

$$\sum_{k>0} k^{-1/2} \|\mathbb{E}(X_k | \mathcal{F}_0)\|_2 < \infty. \quad (2)$$

Let  $\phi(k) = \left\| \sup_{t \in \mathbb{R}} |\mathbb{E}(\mathbf{1}_{X_k \leq t} | \mathcal{F}_0) - \mathbb{E}(\mathbf{1}_{X_k \leq t})| \right\|_\infty$ . If  $X_0 \in \mathbb{L}^p$  for  $p \geq 2$ , then (2) holds under

$$\sum_{k>0} k^{-1/2} (\phi(k))^{(p-1)/p} < \infty.$$

*Proof.* The proofs are based on approximation by a stationary martingale in  $\mathbb{L}^2$ . □

Starting from the so called coboundary decomposition ( $X_k = d_k + Z_k - Z_{k+1}$ ) and using the MDP for martingale (see Puhalskii [29]), Gao [17] and Djellout [11] obtained the MDP for  $\varphi$ -mixing sequences with summable mixing rate.

In the context of Markov process, starting from the Poisson equation, Delyon-Juditsky-Lipster [10] proved the MDP for  $n^{-1/2} \sum_{k=1}^n H(Y_k)$  where  $H$  is a Lipschitz function and  $Y_k = F(Y_{k-1}, \varepsilon_k)$  where  $|F(x, z) - F(y, t)| \leq \kappa|x - y| + L|z - t|$  with  $\kappa < 1$  and  $(\varepsilon_n)$  an i.i.d. sequence of r.v. independent of  $Y_0$  such that  $\mathbb{E}(e^{\delta|\varepsilon_0|}) < \infty$ .

Usually, in dealing with dependent random variables, to brake the dependence, a standard procedure is to divide first the variables in blocks. This technique introduces a new parameter. The second step is then to approximate these blocks either by martingale differences, either by independent blocks using coupling results.

**Proposition 1.4.** [A modification of Puhalskii's result [29]]. *Let  $\{d_{j,n}^{(m)}\}_{1 \leq j \leq k_{n,m}}$  be a martingale difference sequence adapted to  $\mathcal{F}_{j,n}^{(m)}$ . Define  $Z_n^{(m)}(t) = n^{-1/2} \sum_{i=1}^{\lfloor k_{n,m} t \rfloor} d_{i,n}^{(m)}$ . Let  $(a_n)$  be a sequence of positive numbers satisfying (1). Assume that for all  $m \geq 1$*

$$\sup_{1 \leq j \leq k_{n,m}} \|d_{j,n}^{(m)}\|_\infty = o(\sqrt{na_n}) \quad \text{as } n \rightarrow \infty$$

and that for all  $\delta > 0$ , there exists  $\sigma^2 \geq 0$  such that

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} a_n \log \mathbb{P} \left( \left| \frac{1}{n} \sum_{j=1}^{k_{n,m}} \mathbb{E}((d_{j,n}^{(m)})^2 | \mathcal{F}_{(j-1),n}^{(m)}) - \sigma^2 \right| \geq \delta \right) = -\infty.$$

Let  $\{\zeta_n(t), t \in [0, 1]\}$  be a  $\mathbb{D}[0, 1]$ -valued process such that for all  $\delta > 0$ ,

$$\lim_{m \rightarrow \infty} \limsup_{n \rightarrow \infty} a_n \log \mathbb{P} \left( \sqrt{a_n} \sup_{t \in [0,1]} |\zeta_n(t) - Z_n^{(m)}(t)| \geq \delta \right) = -\infty.$$

Then, the processes  $\zeta_n(\cdot)$  satisfy the MDP with rate function  $I_\sigma(\cdot)$  given by

$$I_\sigma(h) = \frac{1}{2\sigma^2} \int_0^1 (h'(u))^2 du \quad (3)$$

if simultaneously  $\sigma > 0$ ,  $h(0) = 0$  and  $h$  is absolutely continuous, and  $I_\sigma(h) = \infty$  otherwise.

**Theorem 1.5.** [Dedecker-Merlevède-Peligrad-Utev [8]]. *Assume that  $\|X_0\|_\infty < \infty$  and that  $X_0$  is  $\mathcal{F}_0$ -measurable. In addition, assume that*

$$\sum_{n=1}^{\infty} n^{-3/2} \|\mathbb{E}(S_n | \mathcal{F}_0)\|_\infty < \infty,$$

and that there exists  $\sigma^2 \geq 0$  with

$$\lim_{n \rightarrow \infty} \|n^{-1} \mathbb{E}(S_n^2 | \mathcal{F}_0) - \sigma^2\|_\infty = 0.$$

Then, for all positive sequences  $(a_n)$  satisfying (1), the normalized partial sums processes  $n^{-1/2} W_n(\cdot)$  satisfy the MDP with the good rate function  $I_\sigma(\cdot)$  given in (3).

**Remark 1.6.** The conditions hold under

$$\sum_{n=1}^{\infty} n^{-1/2} \|\mathbb{E}(X_n | \mathcal{F}_0)\|_\infty < \infty,$$

and

$$\lim_{n \rightarrow \infty} \|\mathbb{E}(X_i X_j | \mathcal{F}_{-n}) - \mathbb{E}(X_i X_j)\|_\infty = 0 \quad \text{for all } i, j \geq 1.$$

In this case  $\sigma^2 = \sum_{k \in \mathbb{Z}} \mathbb{E}(X_0 X_k)$ .

**Remark 1.7.** Let

$$\phi_2(n) = \sup_{i > j \geq n} \left\| \sup_{(s,t) \in \mathbb{R}^2} \left| \mathbb{E}(\mathbf{1}_{X_i \leq s} \mathbf{1}_{X_j \leq t} | \mathcal{F}_0) - \mathbb{E}(\mathbf{1}_{X_i \leq s} \mathbf{1}_{X_j \leq t}) \right| \right\|_\infty.$$

The above conditions are satisfied if

$$\sum_{k > 0} k^{-1/2} \phi_2(k) < \infty.$$

This improves the condition imposed by Gao [17].

*Proof.* The proof of Theorem 1.5 is based on an approximation by a stationary martingale plus an Hoeffding-type inequality.

Let  $m$  be an integer and  $k = k_{n,m} = [n/m]$ . Let  $X_{i,m} = \sum_{j=(i-1)m+1}^{im} X_j$ ,

$$M_k^{(m)} = \sum_{i=1}^k (X_{i,m} - \mathbb{E}(X_{i,m} | \mathcal{F}_{(i-1)m})) := \sum_{i=1}^{[n/m]} D_{i,m} \quad \text{and} \quad M_k^{(m)}(t) := M_{[kt]}^{(m)}.$$

We first notice that

$$\limsup_{n \rightarrow \infty} \left\| \frac{1}{n} \sum_{j=1}^{[n/m]} (\mathbb{E}(D_{j,m}^2 | \mathcal{F}_{(j-1)m}) - \sigma^2) \right\|_\infty \leq \frac{\|\mathbb{E}(S_m | \mathcal{F}_0)\|_\infty^2}{m} + \|m^{-1} \mathbb{E}(S_m^2 | \mathcal{F}_0) - \sigma^2\|_\infty,$$

and

$$\sup_{t \in [0,1]} |S_{[nt]} - M_k^{(m)}(t)| \leq o(\sqrt{na_n}) + \max_{1 \leq j \leq [n/m]} \left| \sum_{i=1}^j \mathbb{E}(X_{i,m} | \mathcal{F}_{(i-1)m}) \right|.$$

To handle the last term in the right-hand side of the above inequality, we apply the Hoeffding-type inequality of Peligrad-Utev-Wu [28]. This gives

$$\begin{aligned} & a_n \log \mathbb{P} \left( \sqrt{\frac{a_n}{n}} \max_{1 \leq j \leq [n/m]} \left| \sum_{i=1}^j \mathbb{E}(X_{i,m} | \mathcal{F}_{(i-1)m}) \right| \geq \delta \right) \\ & \leq a_n \log(4\sqrt{e}) - \frac{\delta^2 m}{2(\|\mathbb{E}(S_m | \mathcal{F}_0)\|_\infty + 80 \sum_{j=1}^{\infty} j^{-3/2} \|\mathbb{E}(S_{jm} | \mathcal{F}_0)\|_\infty)^2}. \end{aligned}$$

□

## 1.2. Applications

### 1.2.1. Contracting Markov chain

Let  $(Y_n)_{n \geq 0}$  be a stationary Markov chain of bounded random variables with invariant measure  $\mu$  and transition kernel  $K$ . Denote by  $\|\cdot\|_{\infty, \mu}$  the essential supremum norm with respect to  $\mu$ . Let  $\Lambda_1$  be the set of 1-Lipschitz functions. Assume that

$$\text{there exist } C > 0 \text{ and } \rho \in ]0, 1[ \text{ such that } \sup_{g \in \Lambda_1} \|K^n(g) - \mu(g)\|_{\infty, \mu} \leq C\rho^n, \quad (4)$$

$$\text{for any } f, g \in \Lambda_1 \text{ and any } m \geq 0 \quad \lim_{n \rightarrow \infty} \|K^n(fK^m(g)) - \mu(fK^m(g))\|_{\infty, \mu} = 0. \quad (5)$$

Let  $\mathcal{L}$  be the class of functions  $f$  from  $\mathbb{R}$  to  $\mathbb{R}$  such that  $|f(x) - f(y)| \leq c(|x - y|)$ , for some concave and non decreasing function  $c$  satisfying

$$\int_0^1 \frac{c(t)}{t\sqrt{|\log t|}} dt < \infty.$$

Assume that the stationary Markov chain  $(Y_n)_{n \geq 0}$  satisfies (4) and (5). If  $f$  belongs to  $\mathcal{L}$ , then the MDP holds for  $\{n^{-1/2} \sum_{k=1}^{\lfloor nt \rfloor} (f(Y_k) - \mu(f)), t \in [0, 1]\}$  with

$$\sigma^2 = \sigma^2(f) = \mu((f - \mu(f))^2) + 2 \sum_{n > 0} \mu(K^n(f) \cdot (f - \mu(f))).$$

### 1.2.2. Linear process

When we deal with functions of a linear process for instance, the next theorem can be more adapted.

**Theorem 1.8.** [Dedecker-Merlevède-Peligrad-Utev [8]]. *Assume that  $\|X_0\|_{\infty} < \infty$ , that  $X_0$  is  $\mathcal{F}_0$ -measurable and that  $\mathbb{E}(X_0 | \mathcal{F}_{-\infty}) = 0$  a.s. In addition, assume that*

$$\sum_{n \geq 0} \|P_0(X_n)\|_{\infty} < \infty,$$

and that for all  $j \geq 0$ ,

$$\lim_{n \rightarrow \infty} \|n^{-1} \sum_{i=1}^n \mathbb{E}(X_i X_{i+j} | \mathcal{F}_0) - \mathbb{E}(X_0 X_j)\|_{\infty} = 0.$$

Then, for all positive sequences  $(a_n)$  satisfying (1), the normalized partial sums processes  $n^{-1/2} W_n(\cdot)$  satisfy the MDP with the good rate function  $I_{\sigma}(\cdot)$  given in (3) where  $\sigma^2 = \sum_{k \in \mathbb{Z}} \mathbb{E}(X_0 X_k)$ .

**Remark 1.9.** This can be extended to linear processes  $Y_k = \sum_{j \geq 0} a_j X_{k-j}$  with  $(X_i)$  satisfying the conditions of the above theorem and  $(a_k)$  in  $\ell^2$ . The rate function is then eventually inherited from a fractional Brownian motion (see Merlevède-Peligrad [23]) and the normalizing sequence is  $\sqrt{\text{Var}(\sum_{k=1}^n Y_k)}$ .

### 1.2.3. Functions of linear processes

Let  $(c_i)_{i \geq 0}$  in  $\ell^1$ ,  $\{\varepsilon_i\}_{i \in \mathbb{Z}}$  a sequence of i.i.d. bounded r.v.'s and

$$X_k = f\left(\sum_{i \geq 0} c_i \varepsilon_{k-i}\right) - \mathbb{E}\left(f\left(\sum_{i \geq 0} c_i \varepsilon_{k-i}\right)\right)$$

Let  $\delta(\varepsilon_0) = 2 \inf\{\|\varepsilon_0 - x\|_{\infty}, x \in \mathbb{R}\}$ , and

$$w_f(h) = \sup_{|t| \leq h, x \in \mathbb{R}} |f(x+t) - f(x)|.$$

- If  $\sum_{n \geq 1} n^{-1/2} w_f(\delta(\varepsilon_0) \sum_{k \geq n} |c_k|) < \infty$  the conditions of Theorem 1.5 hold.
- If  $\sum_{n \geq 0} w_f(\delta(\varepsilon_0) |c_n|) < \infty$ , the conditions of Theorem 1.8 hold.
- Bernoulli Shifts:  $c_i = 2^{-i-1}$  and  $\varepsilon_0$  is such that  $P(\varepsilon_0 = 1) = P(\varepsilon_0 = 0) = 1/2$ . Then the MDP holds as soon as

$$\int_0^1 \frac{w_f(t)}{t \sqrt{|\log t|}} dt < \infty.$$

### 1.3. MDP for strong mixing sequences

The previous theorems do not allow to consider strong mixing sequences and unbounded random variables. Let us consider  $\alpha$ -mixing sequences; i.e.

$$\alpha(n) = \sup_{p \in \mathbb{Z}} \alpha(\mathcal{F}_p, \mathcal{G}_{n+p}) \rightarrow 0 \quad n \rightarrow \infty,$$

where  $\mathcal{F}_p = \sigma(X_j, j \leq p)$ ,  $\mathcal{G}_{n+p} = \sigma(X_j, j \geq n+p)$  and

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|.$$

Assume that for all  $n \geq 1$ ,

$$\alpha(n) \leq \exp(-cn^{\gamma_1}) \quad \text{where } \gamma_1 > 0 \text{ and } c > 0,$$

and the following tail condition: there exist  $b \in ]0, \infty[$ ,  $\gamma_2 \in ]0, +\infty[$  such that

$$\sup_{i > 0} \mathbb{P}(|X_i| > t) \leq \exp(1 - (t/b)^{\gamma_2}) \quad \text{for all } t > 0.$$

In the next theorem, we present a Bernstein-type inequality

**Theorem 1.10.** [Merlevède-Peligrad-Rio [22]]. *Assume that*

$$\gamma < 1 \quad \text{where} \quad \frac{1}{\gamma} = \frac{1}{\gamma_1} + \frac{1}{\gamma_2}.$$

*Then there exists  $\eta > 0$  such that for  $n \geq 4$  and  $\lambda \geq C(\log n)^\eta$*

$$\mathbb{P} \left( \sup_{k \in [1, n]} |S_k| \geq \lambda \right) \leq \exp(-\lambda^2 / (C_1 + C_1 n V)) + (n + 1) \exp(-\lambda^\gamma / C_2),$$

where for  $\varphi_M(x) = (x \wedge M) \vee (-M)$ ,

$$V = \sup_{M \geq 1} \sup_{i > 0} \left( \text{Var}(\varphi_M(X_i)) + 2 \sum_{j > i} |\text{Cov}(\varphi_M(X_i), \varphi_M(X_j))| \right).$$

**Remark 1.11.** Heuristically the inequality can be useful to study  $a_n \log \mathbb{P}(\frac{\sqrt{a_n}}{\sqrt{n}} S_n > t)$  provided  $a_n n^{\gamma/(2-\gamma)} \rightarrow \infty$ .

**Theorem 1.12.** [Merlevède, Peligrad-Rio [22]]. *Assume that*

$$\alpha(n) \leq \exp(-cn^{\gamma_1}) \quad \text{and} \quad \sup_{i > 0} \mathbb{P}(|X_i| > t) \leq \exp(1 - (t/b)^{\gamma_2}),$$



and  $\gamma < 1$  where  $\frac{1}{\gamma} = \frac{1}{\gamma_1} + \frac{1}{\gamma_2}$ . Let  $\sigma_n^2 = \text{Var}S_n$  and assume that

$$\liminf_{n \rightarrow \infty} \frac{\sigma_n^2}{n} > 0.$$

Then, for all positive sequence  $(a_n)$  such that

$$a_n \rightarrow 0 \quad \text{and} \quad a_n n^{\gamma/(2-\gamma)} \rightarrow \infty$$

$\{\sigma_n^{-1}S_n\}$  satisfies the MDP with  $I(t) = t^2/2$ .

**Remark 1.13.** If  $(X_i)$  is a second order stationary sequence, under the mixing and tail conditions  $\lim_{n \rightarrow \infty} \frac{\sigma_n^2}{n} = \sigma^2 > 0$  as soon as  $\sigma_n^2 \rightarrow \infty$ .

*Proof.* We give some hints for the proof of Theorem 1.12. The Bernstein type inequality allows to consider the r.v.'s truncated at a level  $T_n$ . Let

$$I(n, j) = \{(j-1)(p_n + q_n) + 1, \dots, (j-1)(p_n + q_n) + p_n\}$$

$$J(n, j) = \{(j-1)(p_n + q_n) + p_n + 1, \dots, j(p_n + q_n)\}$$

let  $S'(K) = \sum_{i \in K} X'_i$  and  $m_n = \lfloor n/(p_n + q_n) \rfloor$

$$S'_n = \sum_{j=1}^{m_n} S'(I(n, j)) + \sum_{j=1}^{m_n} S'(J(n, j)) + R_n.$$

The idea is to consider discrete Cantor type sets. We construct a set

$$K_{I(n, j)}^{(\ell_n)} = \bigcup_{i=1}^{2^{\ell_n}} I_{\ell_n, i}(p_n, j),$$

where the  $I_{\ell_n, i}(p_n, j)$  are disjoint sets of consecutive integers, each of same cardinal such that

$$\frac{p_n}{2^{\ell_n}(1 + \varepsilon_n)} \leq \text{Card}I_{\ell_n, i}(p_n, j) \leq \frac{p_n}{2^{\ell_n}}.$$

We have the following

$$\sum_{j=1}^{m_n} S'(I(n, j)) = \sum_{j=1}^{m_n} S'(K_{I(n, j)}^{(\ell_n)}) + \sum_{j=1}^{m_n} S'((K_{I(n, j)}^{(\ell_n)})^c).$$

Using coupling arguments, we reduce the problem of studying the MDP for  $\sum_{j=1}^{m_n} S^*(K_{I(n, j)}^{(\ell_n)})$  where the random variables  $(S^*(K_{I(n, j)}^{(\ell_n)}))_{1 \leq j \leq m_n}$  are independent with the same distribution as the random variables  $S'(K_{I(n, j)}^{(\ell_n)})$ .

We need to show that

$$a_n \sum_{j=1}^{m_n} \log \mathbb{E} \exp \left( t S'(K_{I(n, j)}^{(\ell_n)}) / \sqrt{a_n \sigma_n^2} \right) \rightarrow \frac{t^2}{2} \quad \text{as } n \rightarrow \infty.$$

This can be done by decorrelation step by step on the Cantor set that implies

$$a_n \left| \sum_{j=1}^{m_n} \log \mathbb{E} \exp \left( t S' (K_{I(n,j)}^{(\ell_n)}) / \sqrt{a_n \sigma_n^2} \right) - \sum_{j=1}^{m_n} \sum_{i=1}^{2^{\ell_n}} \log \mathbb{E} \exp \left( t S' (I_{\ell_n,i}(p_n, j)) / \sqrt{a_n \sigma_n^2} \right) \right| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

□

**Remark 1.14.** *On the mixing-coefficients.* The  $\alpha$ -mixing coefficients are not needed in their full generality. We need mixing-coefficients allowing coupling and decorrelation of blocks of random variables. We can use of the  $\tau$ -mixing coefficients as introduced by Dedecker and Prieur [9]. These coefficients are easily computable in a lot of situations as for instance iterated Lipschitz models or functions of linear processes generated by absolutely regular innovations.

### 1.4. Applications

#### 1.4.1. Application to Markov chains

Let  $(Y_j)_{j \geq 0}$  be an  $E$ -valued, irreducible ergodic and stationary Markov chain with a transition probability  $P$  having a unique invariant probability measure  $\pi$ . Assume that the chain has an atom: there exists  $A \in E$  with  $\pi(A) > 0$  and  $\nu$  a probability measure such that  $P(x, \cdot) = \nu(\cdot)$  for all  $x \in A$ . Assume that there exists  $\delta > 0$  and  $\gamma_1 \in ]0, 1]$  such that for  $\tau = \inf\{n \geq 0; Y_n \in A\}$ :

$$\mathbb{E}_\nu(\exp(\delta \tau^{\gamma_1})) < \infty.$$

Djellout and Guillin [12]: for each bounded function  $f$  from  $E$  to  $\mathbb{R}$  with  $\pi(f) = 0$ , the MDP holds for  $n^{-1/2} \sum_{i=1}^n f(Y_i)$  with  $a_n$  such that  $a_n \rightarrow 0$  and  $a_n n^{\gamma_1/(2-\gamma_1)} \rightarrow \infty$ .

Applying our MDP result we obtain: Suppose that  $\pi(f) = 0$  and there exist  $b \in ]0, \infty[$  and  $\gamma_2 \in ]0, \infty]$  such that

$$\pi(|f| > t) \leq \exp(1 - (t/b)^{\gamma_2}) \text{ for any } t > 0.$$

If  $1/\gamma_1 + 1/\gamma_2 > 1$  then the MDP holds with speed  $(a_n)$  satisfying  $a_n \rightarrow 0$  and  $a_n n^{\gamma_1/(2-\gamma_1)} \rightarrow \infty$ .

#### 1.4.2. Autoregressive Lipschitz model

For  $\delta$  in  $[0, 1[$  and  $C$  in  $]0, 1]$ , let  $\mathcal{L}(C, \delta)$  be the class of 1-Lipschitz functions  $f$  which satisfy

$$f(0) = 0 \text{ and } |f'(t)| \leq 1 - C(1 + |t|)^{-\delta} \text{ almost everywhere.}$$

Let  $(\varepsilon_i)_{i \in \mathbb{Z}}$  be a sequence of i.i.d. real-valued random variables. For  $\eta \in ]0, 1]$ , let  $ARL(C, \delta, \eta)$  be the class of Markov chains on  $\mathbb{R}$  defined by

$$Y_n = f(Y_{n-1}) + \varepsilon_n \text{ with } f \in \mathcal{L}(C, \delta) \text{ and } \mathbb{E}(\exp(\lambda |\varepsilon_0|^\eta)) < \infty \text{ for a } \lambda > 0.$$

Let  $g$  be a 1-Lipschitz function such that  $|g(x)| \leq c(1 + |x|^\zeta)$  for  $\zeta$  in  $[0, 1]$ . If  $\delta + \zeta > 0$ , then  $(g(Y_i) - \mathbb{E}(g(Y_i)))_{i \in \mathbb{Z}}$  satisfies the MDP with

$$\gamma_2 = \eta(1 - \delta)/\zeta \text{ and } \gamma_1 = \eta(1 - \delta)(\eta(1 - \delta) + \delta)^{-1}.$$

## 2. MODERATE DEVIATIONS FOR THE DURBIN-WATSON STATISTIC RELATED TO THE FIRST-ORDER AUTOREGRESSIVE PROCESS

The purpose of this section is to investigate MDP for the Durbin-Watson statistic associated with the stable first-order autoregressive process where the driven noise is also given by a first-order autoregressive process. We first establish a MDP for both the least squares estimator of the unknown parameter of the autoregressive process as well as for the serial correlation estimator associated with the driven noise. It enables us to provide a MDP for the Durbin-Watson statistic in the easy case where the driven noise is normally distributed and in the more general case where the driven noise satisfies a less restrictive Chen-Ledoux type condition.

We start by introducing some notations and definitions, see [13]. In the whole paper, for any matrix  $M$ ,  $M'$  and  $\|M\|$  stand for the transpose and the euclidean norm of  $M$ , respectively. Let  $(b_n)$  be a sequence of increasing positive numbers such that

$$b_n \longrightarrow \infty, \quad \frac{b_n}{\sqrt{n}} \longrightarrow 0. \quad (6)$$

**Remark 2.1.** The condition (6) corresponds to (1) with  $a_n = 1/b_n^2$ .

**Definition 2.2.** We say that  $Z_n$  converges  $(b_n^2)$ -superexponentially fast in probability to some random variable  $Z$  if, for all  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\|Z_n - Z\| > \delta) = -\infty, \quad \text{and we note } Z_n \xrightarrow[b_n^2]{\text{superexp}} Z.$$

**Definition 2.3.** We say that  $Y_n$  and  $Z_n$  are  $(b_n^2)$ -exponentially equivalent if  $\|Y_n - Z_n\|$  is negligible with respect to the large deviations, that is, for all  $\delta > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P}(\|Y_n - Z_n\| > \delta) = -\infty, \quad \text{and we note } Y_n \xrightarrow[b_n^2]{\text{superexp}} Z_n.$$

Let us introduce the autoregressive process of order 1 with autocorrelated driven noise that we shall consider. For all  $n \geq 1$ ,

$$\begin{cases} X_n &= \theta X_{n-1} + \varepsilon_n \\ \varepsilon_n &= \rho \varepsilon_{n-1} + V_n \end{cases} \quad (7)$$

where  $X_0$  and  $\varepsilon_0$  are square-integrable,  $(V_n)$  is i.i.d. with  $\mathbb{E}[V_1^2] = \sigma^2$  and  $\mathbb{E}[V_1^4] = \tau^4$ . The stability of the process is insured by  $|\theta| < 1$  and  $|\rho| < 1$ . We introduce the least squares estimates

$$\hat{\theta}_n = \frac{\sum_{k=1}^n X_k X_{k-1}}{\sum_{k=1}^n X_{k-1}^2}, \quad \hat{\rho}_n = \frac{\sum_{k=1}^n \hat{\varepsilon}_k \hat{\varepsilon}_{k-1}}{\sum_{k=1}^n \hat{\varepsilon}_{k-1}^2}, \quad \hat{D}_n = \frac{\sum_{k=1}^n (\hat{\varepsilon}_k - \hat{\varepsilon}_{k-1})^2}{\sum_{k=0}^n \hat{\varepsilon}_k^2},$$

where the least squares residuals are given at stage  $n$ , for all  $1 \leq k \leq n$ , by

$$\hat{\varepsilon}_k = X_k - \hat{\theta}_n X_{k-1}.$$

Our objectives is to establish an MDP for these estimates under the restrictive case where  $(V_n)$  is gaussian and under the more general hypothesis that  $(V_n)$  satisfies a Chen-Ledoux type condition.

Before going further, let us recall here a useful theorem to establish MDP for gaussian martingales, used intensively in the next section.

**Theorem 2.4.** [Worms [31,32]]. *Let  $(Y_n)$  be an adapted sequence with values in  $\mathbb{R}^p$ , and  $(V_n)$  a gaussian noise with variance  $\sigma^2 > 0$ . We suppose that  $(Y_n)$  satisfies, for some invertible square matrix  $C$  of order  $p$ , the exponential convergence, for any  $\delta > 0$ ,*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \left\| \frac{1}{n} \sum_{k=0}^{n-1} Y_k Y_k' - C \right\| > \delta \right) = -\infty.$$

*Then, the sequence  $(n^{-1/2} M_n)_{n \geq 1}$  satisfies an MDP on  $\mathbb{R}^p$  of speed  $1/b_n^2$  and rate function  $I(x) = \frac{1}{2\sigma^2} x' C^{-1} x$ ,*

*where  $(M_n)$  is the martingale given by  $M_n = \sum_{k=1}^n Y_{k-1} V_k$ .*

We also introduce a similar result related to nongaussian martingales which will be useful in the last section.

**Theorem 2.5.** [Puhalskii [29]]. *Let  $(m_j^n)_{1 \leq j \leq n}$  be a triangular array of martingale differences with values in  $\mathbb{R}^d$ , with respect to the filtration  $(\mathcal{F}_n)_{n \geq 1}$ . Let  $(b_n)$  be a sequence of real numbers satisfying (6). Suppose that there exists a symmetric positive-semidefinite matrix  $Q$  such that*

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[ m_k^n (m_k^n)' | \mathcal{F}_{k-1} \right] \xrightarrow[b_n^2]{\text{superexp}} Q.$$

*Suppose that there exists a constant  $c > 0$  such that, for each  $1 \leq k \leq n$ ,  $\|m_k^n\| \leq c \frac{\sqrt{n}}{b_n}$  a.s.*

*Suppose also that, for all  $a > 0$ , we have the exponential Lindeberg's condition*

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[ \|m_k^n\|^2 \mathbf{I}_{\{\|m_k^n\| \geq a \frac{\sqrt{n}}{b_n}\}} | \mathcal{F}_{k-1} \right] \xrightarrow[b_n^2]{\text{superexp}} 0.$$

*Then, the sequence  $(n^{-1/2} \sum_{k=1}^n m_k^n)_{n \geq 1}$  satisfies an MDP on  $\mathbb{R}^d$  with speed  $1/b_n^2$  and good rate function*

$$\Lambda^*(v) = \sup_{\lambda \in \mathbb{R}^d} \left( \lambda' v - \frac{1}{2} \lambda' Q \lambda \right).$$

*In particular, if  $Q$  is invertible  $\Lambda^*(v) = \frac{1}{2} v' Q^{-1} v$ .*

## 2.1. Results on the Durbin-Watson testing procedure

In [3], the following almost sure convergences are established for our estimates,

$$\lim_{n \rightarrow \infty} \widehat{\theta}_n = \theta^* \quad \text{a.s.} \quad \lim_{n \rightarrow \infty} \widehat{\rho}_n = \rho^* \quad \text{a.s.} \quad \lim_{n \rightarrow \infty} \widehat{D}_n = D^* \quad \text{a.s.}$$

where  $\theta^* = (1 + \theta\rho)^{-1} (\theta + \rho)$ ,  $\rho^* = \theta\rho\theta^*$  and  $D^* = 2(1 - \rho^*)$ . The objective of [3] was in particular to establish a statistical procedure for testing  $\mathcal{H}_0 : \rho = 0$  against  $\mathcal{H}_1 : \rho \neq 0$  and we shall recall the associated result in the sequel. To summarize, assume that  $\theta \neq 0$  and  $\theta^* \neq 0$ . Then, under the null  $\mathcal{H}_0$ ,

$$\frac{n}{4\widehat{\theta}_n^2} \left( \widehat{D}_n - 2 \right)^2 \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi_1^2.$$

In addition, under the alternative  $\mathcal{H}_1$ ,

$$\lim_{n \rightarrow \infty} \frac{n}{4\hat{\theta}_n^2} \left( \hat{D}_n - 2 \right)^2 = +\infty \text{ a.s.}$$

On Figure 1 below, the empirical frequencies with which  $\mathcal{H}_0$  is rejected for a large simulation study is represented for different values of  $\rho$ , in comparison with other usual procedures for testing serial correlation (*Durbin's h-test* HT, *Breusch-Godfrey* BG, *Ljung-Box* LB and *Box-Pierce* BP). Figure 1 shows that the aforesaid procedure outperforms all tests on small-size samples, and that it is equally powerful than the BG and HT procedures on large samples.

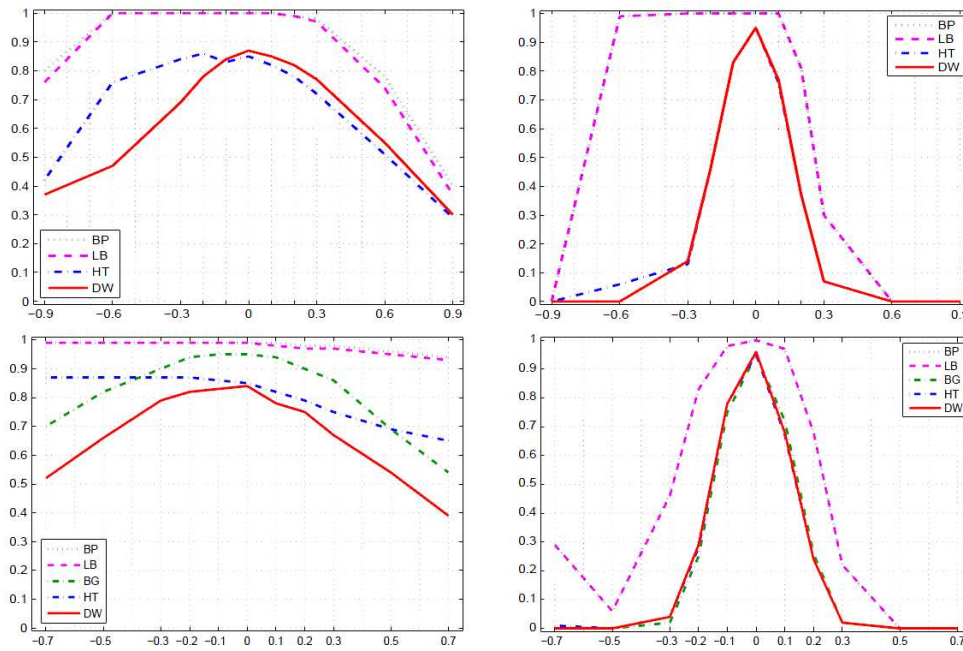


FIGURE 1. Simulation study for  $n = 30$  and  $n = 500$  ( $p = 1, p = 3$ ).

## 2.2. Moderate deviations when $(V_n)$ is gaussian

In this section, we need to introduce the following hypothesis.

- (H1) The driven noise  $(V_n)$  is i.i.d. with a gaussian distribution,  $\mathbb{E}[V_1] = 0$  and  $\mathbb{E}[V_1^2] = \sigma^2$ .
- (H2) There exists  $t > 0$  such that

$$\mathbb{E} \left[ \exp(t\varepsilon_0^2) \right] < \infty \quad \text{and} \quad \mathbb{E} \left[ \exp(tX_0^2) \right] < \infty.$$

### 2.2.1. Moderate deviations for $\hat{\theta}_n$

**Theorem 2.6.** [Bercu-Proia [3]]. Assume that  $\mathbb{E}[V_1^4] < \infty$ . Then, we have the asymptotic normality

$$\sqrt{n} \left( \hat{\theta}_n - \theta^* \right) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \sigma_\theta^2)$$

where

$$\sigma_\theta^2 = \frac{(1-\theta^2)(1-\theta\rho)(1-\rho^2)}{(1+\theta\rho)^3}.$$

**Theorem 2.7.** [Bitseki Penda-Djellout-Proïa [5]]. *Assume that the hypothesis (H1), (H2) are satisfied. Then, the sequence  $\left(\sqrt{n}(\widehat{\theta}_n - \theta^*)\right)_{n \geq 1}$  satisfies an MDP on  $\mathbb{R}$  with speed  $1/b_n^2$  and good rate function  $I_\theta(x) = \frac{x^2}{2\sigma_\theta^2}$ .*

*Proof.* To get an outline of the proof, consider the decomposition

$$\frac{\sqrt{n}}{b_n}(\widehat{\theta}_n - \theta^*) = \frac{\sqrt{n}}{b_n} \left( \frac{\sigma^2}{1+\theta\rho} \right) \frac{M_n}{\langle M \rangle_n} + \frac{n}{S_{n-1}} \left( \frac{1}{1+\theta\rho} \right) \frac{R_n}{b_n\sqrt{n}}$$

where  $M_n = \sum_{k=1}^n X_{k-1}V_k$  is a martingale,  $S_n = \sum_{k=0}^n X_k^2$  and  $R_n$  is a residual. We prove that the first term satisfies an MDP by Theorem 2.4, and that the second term is exponentially negligible. Then, we establish the exponential equivalence

$$\frac{\sqrt{n}}{b_n}(\widehat{\theta}_n - \theta^*) \underset{b_n^2}{\overset{\text{superexp}}{\sim}} \frac{1}{\ell(1+\theta\rho)} \frac{M_n}{b_n\sqrt{n}}$$

where  $\ell$  is the almost sure limit of  $S_n/n$ . We conclude by using the contraction principle. The whole proof may be found in [5].  $\square$

### 2.2.2. Moderate deviations for $\widehat{\rho}_n$

**Theorem 2.8.** [Bercu-Proïa [3]]. *Assume that  $\mathbb{E}[V_1^4] < \infty$ . Then, we have the joint asymptotic normality*

$$\sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, \Gamma)$$

where

$$\Gamma = \begin{pmatrix} \sigma_\theta^2 & \theta\rho\sigma_\theta^2 \\ \theta\rho\sigma_\theta^2 & \sigma_\rho^2 \end{pmatrix}$$

and

$$\sigma_\rho^2 = \frac{(1-\theta\rho)}{(1+\theta\rho)^3} \left( (\theta+\rho)^2(1+\theta\rho)^2 + (\theta\rho)^2(1-\theta^2)(1-\rho^2) \right).$$

**Theorem 2.9.** [Bitseki Penda-Djellout-Proïa [5]]. *Assume that the hypothesis (H1), (H2) are satisfied. Then, as soon as  $\theta \neq -\rho$ , the sequence*

$$\left( \sqrt{n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \right)_{n \geq 1}$$

*satisfies an MDP on  $\mathbb{R}^2$  with speed  $1/b_n^2$  and good rate function  $K(x) = \frac{1}{2}x'\Gamma^{-1}x$ .*

*Proof.* To get an outline of the proof, consider the decomposition

$$\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} = \frac{1}{b_n\sqrt{n}} A_n Z_n + \frac{\sqrt{n}}{b_n} \begin{pmatrix} 1 \\ 1+\theta\rho \end{pmatrix} \begin{pmatrix} R_{1,n} \\ R_{2,n} \end{pmatrix}$$

where

$$A_n = \begin{pmatrix} A_n^{(1,1)} & 0 \\ A_n^{(2,1)} & A_n^{(2,2)} \end{pmatrix}, \quad Z_n = \begin{pmatrix} M_n \\ N_n \end{pmatrix},$$

and where  $M_n = \sum_{k=1}^n X_{k-1}V_k$  and  $N_n = \sum_{k=2}^n X_{k-2}V_k$  are martingales,  $A_n$  is a square matrix, explicitly given, converging almost surely and exponentially to  $A$ , and  $R_{1,n}$  and  $R_{2,n}$  are residuals. We prove that the

first term satisfies an MDP by Theorem 2.4, and that the second term is exponentially negligible. Then, we establish the exponential equivalence

$$\frac{\sqrt{n}}{b_n} \begin{pmatrix} \widehat{\theta}_n - \theta^* \\ \widehat{\rho}_n - \rho^* \end{pmatrix} \underset{b_n^2}{\overset{\text{superexp}}{\rightsquigarrow}} \frac{1}{b_n \sqrt{n}} AZ_n.$$

We conclude by using the contraction principle. The whole proof may be found in [5].  $\square$

### 2.2.3. Moderate deviations for $\widehat{D}_n$

**Theorem 2.10.** [Bercu-Proïa [3]]. *Assume that  $\mathbb{E}[V_1^4] < \infty$ . Then, we have the asymptotic normality*

$$\sqrt{n} \left( \widehat{D}_n - D^* \right) \underset{n \rightarrow \infty}{\xrightarrow{\mathcal{L}}} \mathcal{N}(0, \sigma_D^2)$$

where  $\sigma_D^2 = 4\sigma_\rho^2$ .

**Theorem 2.11.** [Bitseki Penda-Djellout-Proïa [5]]. *Assume that the hypothesis **(H1)**, **(H2)** are satisfied. Then, the sequence  $\left( \sqrt{n} \left( \widehat{D}_n - D^* \right) \right)_{n \geq 1}$  satisfies an MDP on  $\mathbb{R}$  with speed  $1/b_n^2$  and good rate function  $I_D(x) = \frac{x^2}{2\sigma_D^2}$ .*

*Proof.* The result immediately follows from the exponential equivalence

$$\frac{\sqrt{n}}{b_n} \left( \widehat{D}_n - D^* \right) \underset{b_n^2}{\overset{\text{superexp}}{\rightsquigarrow}} -2 \frac{\sqrt{n}}{b_n} \left( \widehat{\rho}_n - \rho^* \right).$$

The whole proof may be found in [5].  $\square$

## 2.3. Moderate deviations when $(V_n)$ satisfies a Chen-Ledoux type condition

In this section, we need to introduce the following hypothesis.

**(H3)** The Chen-Ledoux type condition. For  $a > 0$ ,

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \left( n \mathbb{P}(|V_1|^a > b_n \sqrt{n}) \right) = -\infty.$$

**(H4)** The initial values satisfy

$$\frac{|\varepsilon_0|^a}{b_n \sqrt{n}} \underset{b_n^2}{\overset{\text{superexp}}{\rightsquigarrow}} 0 \quad \text{and} \quad \frac{|X_0|^a}{b_n \sqrt{n}} \underset{b_n^2}{\overset{\text{superexp}}{\rightsquigarrow}} 0.$$

Depending on the result we are currently proving, we need to assume that  $a = 2$  or  $a = 4$ .

**Remark 2.12.** For  $b_n = n^\alpha$  and  $0 < \alpha < 1/2$ , the Chen-Ledoux condition for  $a = 2$  is satisfied if there exists  $t > 0$  and  $0 < \beta < 1$  such that

$$\mathbb{E} \left[ \exp(tV_1^{2\beta}) \right] < \infty.$$

The condition is also satisfied for  $b_n = n^\alpha$  if

$$\frac{|V_1|^a}{b_n \sqrt{n}} \underset{b_n^2}{\overset{\text{superexp}}{\rightsquigarrow}} 0.$$

**Theorem 2.13.** [Bitseki Penda-Djellout-Proïa [5]]. *Assume that the hypothesis **(H3)**, **(H4)** are satisfied. Then, the MDP established in Theorems 2.7, 2.9 and 2.11 still hold.*

*Proof.* We need to use Theorem 2.5 of Puhalskii [29] for nongaussian martingales to establish a related MDP, together with a result of Eichelsbacher and Löwe [16] related to i.i.d. random variables for which we have no information on the log-Laplace transform, that we state below.

**Theorem 2.14.** [Eichelsbacher-Löwe [16]]. *The following results are equivalent.*

(1) *The i.i.d. real-valued random variables  $(Y_k)$  satisfy  $\mathbb{E}[Y_1] < \infty$  and*

$$\limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log (n \mathbb{P}(|Y_1| > b_n \sqrt{n})) = -\infty.$$

(2) *The sequence  $\left( n^{-1/2} \sum_{k=1}^n (Y_k - \mathbb{E}[Y_k]) \right)_{n \geq 1}$  satisfies an MDP with speed  $1/b_n^2$  and good rate function  $I(x) > 0$  for all  $x \neq 0$ , and  $\lim_{x \rightarrow -\infty} I(x) = \lim_{x \rightarrow +\infty} I(x) = +\infty$ .*

Let us just give a sketch of the strategy used to prove our result. We first establish an MDP for our estimates without any gaussianity assumption on the driven noise. To summarize, it is possible to truncate all sequences beyond an unbounded limit, that is, for  $r, R > 0$ ,

$$X_k^{(r)} = X_k \mathbb{I}_{\{|X_k| \leq r \frac{\sqrt{n}}{b_n}\}}, \quad V_k^{(R)} = V_k \mathbb{I}_{\{|V_k| \leq R\}} - \mathbb{E}[V_k \mathbb{I}_{\{|V_k| \leq R\}}], \quad M_n^{(r,R)} = \sum_{k=1}^n X_{k-1}^{(r)} V_k^{(R)}.$$

Then, by Theorem 2.5 applied to the nongaussian martingale  $(M_n^{(r,R)})$ , we establish an MDP on the truncated decomposition. The rate function is given by

$$I_R(x) = \frac{x^2}{2\mathcal{L}\mathbb{E}[(V_1^{(R)})^2]}.$$

Finally, we show that the remaining part of the sequences are exponentially negligible. For all  $r > 0$  and  $\delta > 0$ ,

$$\limsup_{R \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{b_n^2} \log \mathbb{P} \left( \frac{1}{b_n \sqrt{n}} |M_n - M_n^{(r,R)}| > \delta \right) = -\infty.$$

The whole and technical proof may be found in [5]. □

### 3. MODERATE DEVIATION PRINCIPLE FOR BIFURCATING MARKOV CHAINS

The objective of this section is to give deviation inequalities and MDP for Bifurcating Markov Chains (BMC, in short). The results will be obtained under hypothesis of geometric ergodicity or uniform geometric ergodicity of an embedded Markov chain. As statistical applications, we provide deviation inequalities (for either the gaussian setting or the bounded setting), for least square estimators of the parameters of a first order bifurcating autoregressive process.

#### 3.1. BMC's model

Bifurcating Markov chains (BMC) are an adaptation of Markov chains to the data of a regular binary tree. They are appropriate for example in the modeling of cell lineage data when each cell in one generation gives birth to two offspring in the next one. Recently, they have received a great deal of attention because of the experiments of biologists on aging of *Escherichia Coli* (E. Coli in short). E. Coli is a rod-shaped bacterium which reproduces by dividing in the middle, thus producing two cells, one which already existed and that we



call old pole progeny cell, and the other which is new and that we call new pole progeny cell. One question of interest is to know if the new pole progeny cells grow at the same rate that the old pole progeny cells. The answer to this question was one of the main motivation of the introduction of BMC by Guyon [19]. Let us now formally introduce the BMC.

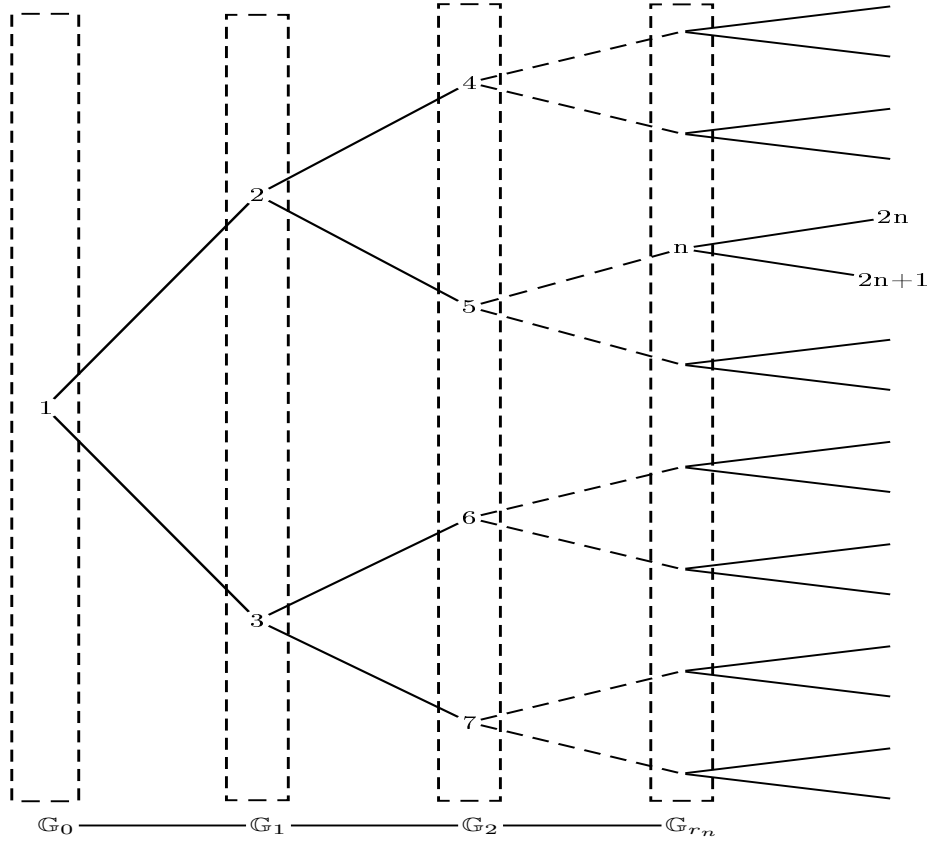


FIGURE 2. The binary tree  $\mathbb{T}$

Let  $\mathbb{T}$  be a binary regular tree, see Fig 2. We shall see  $\mathbb{T}$  as a given population. Each individual (vertex)  $n \in \mathbb{T}$  is seen as a positive integer  $n \in \mathbb{N}^*$ . For  $r \in \mathbb{N}$ , we denote by

$$\mathbb{G}_r = \{2^r, 2^r + 1, \dots, 2^{r+1} - 1\} \quad (\text{resp. } \mathbb{T}_r = \{1, 2, \dots, 2^{r+1} - 1\})$$

the  $r$ -th generation (resp. the first  $r + 1$  generations of the population). Then, the cardinality  $|\mathbb{G}_r|$  of  $\mathbb{G}_r$  (resp.  $|\mathbb{T}_r|$  of  $\mathbb{T}_r$ ) is

$$|\mathbb{G}_r| = 2^r \quad (\text{resp. } |\mathbb{T}_r| = 2^{r+1} - 1).$$

The generation of a given individual  $n$  is  $\mathbb{G}_{r_n}$  with  $r_n = \lfloor \log_2 n \rfloor$ , where  $\lfloor x \rfloor$  denotes the integer part of the real number  $x$ .

Let  $(S, \mathcal{S})$  be a metric space endowed with its Borel  $\sigma$ -field. We call  $\mathbb{T}$ -transition probability any mapping  $P : S \times \mathcal{S}^2 \rightarrow [0, 1]$  such that

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

For a  $\mathbb{T}$ -transition probability  $P$  on  $S \times S^2$ , we denote by  $P_0$ ,  $P_1$  and  $Q$  respectively the first and the second marginal of  $P$ , and the mean of  $P_0$  and  $P_1$ , that is  $P_0(x, B) = P(x, B \times S)$ ,  $P_1(x, B) = P(x, S \times B)$  for all  $x \in S$  and  $B \in \mathcal{S}$  and  $Q = \frac{P_0 + P_1}{2}$ .

For  $p \geq 1$ , we denote by  $\mathcal{B}(S^p)$  (resp.  $\mathcal{B}_b(S^p)$ ), the set of all  $S^p$ -measurable (resp.  $S^p$ -measurable and bounded) mappings  $f : S^p \rightarrow \mathbb{R}$ . For  $f \in \mathcal{B}(S^3)$ , when it is defined, we denote by  $Pf \in \mathcal{B}(S)$  the function

$$x \mapsto Pf(x) = \int_{S^2} f(x, y, z)P(x, dydz).$$

Then, let  $(X_n, n \in \mathbb{T})$  be a family of  $S$ -valued random variables defined on a filtered probability space  $(\Omega, \mathcal{F}, (\mathcal{F}_r)_{r \geq 0}, \mathbb{P})$ . Let  $\nu$  be a probability on  $(S, \mathcal{S})$  and  $P$  be a  $\mathbb{T}$ -transition probability.

**Definition 3.1.** We say that  $(X_n, n \in \mathbb{T})$  is a  $(\mathcal{F}_r)$ -bifurcating Markov chain with initial distribution  $\nu$  and  $\mathbb{T}$ -transition probability  $P$  if

- (a)  $X_n$  is  $\mathcal{F}_{r_n}$ -measurable for all  $n \in \mathbb{T}$ ,
- (b)  $\mathcal{L}(X_1) = \nu$ ,
- (c) for all  $r \in \mathbb{N}$  and for all family  $(f_n, n \in \mathbb{G}_r) \subseteq \mathcal{B}_b(S^3)$

$$\mathbb{E} \left[ \prod_{n \in \mathbb{G}_r} f_n(X_n, X_{2n}, X_{2n+1}) \middle| \mathcal{F}_r \right] = \prod_{n \in \mathbb{G}_r} Pf_n(X_n).$$

**Remark 3.2.** A typical example of bifurcating Markov chain is given by the (stable) first order bifurcating autoregressive process (BAR(1), in short) defined as follows:

$$\mathcal{L}(X_1) = \nu, \quad \text{and} \quad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases}$$

where  $\nu$  is a distribution probability on  $\mathbb{R}$ ,  $\alpha_0, \alpha_1 \in (-1, 1)$ ;  $\beta_0, \beta_1 \in \mathbb{R}$  and  $((\varepsilon_{2n}, \varepsilon_{2n+1}), n \geq 1)$  forms a sequence of centered i.i.d. bivariate random variables with covariance matrix

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$

One can think of a cell of E. Coli “ $n$ ”, that reproduces by dividing into two, thus producing two individuals: one, denoted by  $2n + 1$ , the old pole progeny cell, and the other, denoted by  $2n$ , the new pole progeny cell. For a cell “ $n$ ”,  $X_n$  denote some quantitative value (growth rate, weight...). An issue in this model is for example whether the values associated to new pole progeny and that associated to old pole progeny evolve in the same way. For this purpose, it is advisable to estimate the parameters  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ ,  $\sigma^2$  and  $\rho$  and to test null hypothesis  $H_0 = \{(\alpha_0, \beta_0) = (\alpha_1, \beta_1)\}$  against its alternative  $H_1 = \{(\alpha_0, \beta_0) \neq (\alpha_1, \beta_1)\}$ .

In order, for example, to study the statistics arising for BAR(1) process, we need to define some empirical means related to BMC  $(X_n, n \in \mathbb{T})$ . For all  $f \in \mathcal{B}(S)$  (resp.  $\mathcal{B}(S^3)$ ), we set

$$M_{\mathbb{T}_r}(f) = \sum_{i \in \mathbb{T}_r} f(\tilde{\Delta}_i),$$

with

$$\begin{cases} f(\tilde{\Delta}_i) = f(X_i) & \text{if } f \in \mathcal{B}(S) \\ f(\tilde{\Delta}_i) = f(\Delta_i) & \text{if } f \in \mathcal{B}(S^3) \end{cases} \quad \text{where } \Delta_i = (X_i, X_{2i}, X_{2i+1}),$$

and

$$\overline{M}_{\mathbb{T}_r}(f) = |\mathbb{T}_r|^{-1} M_{\mathbb{T}_r}(f).$$

Under suitable regularity assumptions, Guyon, J. [19] proved laws of large numbers and CLT for empirical averages  $\overline{M}_{\mathbb{T}_r}(f)$ . Our objectives in this section are:

- on the one hand, to specify the order of magnitude in this law of large numbers by given deviation inequalities for  $\overline{M}_{\mathbb{T}_r}(f)$ , that is non asymptotic estimation of the form

$$\mathbb{P}(|\overline{M}_{\mathbb{T}_r}(f) - s| > \delta) \leq h(\delta, r, c),$$

where  $h$ ,  $c$  and  $s$  will be specify later;

- on the other hand to give MDP for

$$\frac{M_{\mathbb{T}_r}(f - Pf)}{b_{|\mathbb{T}_r|}} \quad (\text{for } f \in \mathcal{B}(S^3)),$$

where the sequence  $(b_n)$  will be specify later.

We will work with the subspace  $F$  of  $\mathcal{B}(S)$  which verifies

- $F$  contains the constants,
- $F^2 \subset F$ ,
- $F \otimes F \subset L^1(P(x, \cdot))$  for all  $x \in S$ , and  $P(F \otimes F) \subset F$ ,
- there exists a probability  $\mu$  on  $(S, \mathcal{S})$  such that  $F \subset L^1(\mu)$  and  $\lim_{r \rightarrow \infty} Q^r f(x) = (\mu, f)$  for all  $x \in S$  and  $f \in F$ ,
- for all  $f \in F$ , there exists  $g \in F$  such that for all  $r \in \mathbb{N}$ ,  $|Q^r f| \leq g$ ,
- $F \subset L^1(\nu)$

We introduce the following hypothesis (where  $\mu$  is the probability measure given in above hypothesis (iv)):

**(A1)** Geometric ergodicity of  $Q$ :  $\forall f \in F$  such that  $(\mu, f) = 0$ ,  $\exists g \in F$  such that

$$\forall r \in \mathbb{N} \quad \text{and} \quad \forall x \in S, |Q^r f(x)| \leq \alpha^r g(x) \quad \text{for some } \alpha \in (0, 1).$$

**(A2)** Uniform geometric ergodicity of  $Q$ :  $\forall f \in \mathcal{B}_b(S)$  such that  $(\mu, f) = 0$ ,  $\exists c > 0$  such that

$$|Q^r f(x)| \leq c\alpha^r \quad \text{for some } \alpha \in (0, 1) \quad \text{and for all } x \in S.$$

### 3.2. Main results

**Theorem 3.3.** [Bitseki Penda-Djellout-Guillin [4]]. *Let  $f \in F$  such that  $(\mu, f) = 0$ . We assume hypothesis (A1). Then for all  $r \in \mathbb{N}$*

$$\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(f)| > \delta\right) \leq \begin{cases} \frac{c'}{\delta^4} \left(\frac{1}{4}\right)^{r+1} & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c'}{\delta^4} r^2 \left(\frac{1}{4}\right)^{r+1} & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c'}{\delta^4} \alpha^{4r+4} & \text{if } \alpha^2 > \frac{1}{2}; \end{cases} \quad (8)$$

where the positive constant  $c'$  depends on  $\alpha$  and  $f$ .

When  $f$  depends on the mother-daughters triangle  $(\Delta_i)$ , we have the following.

**Theorem 3.4.** [Bitseki Penda-Djellout-Guillin [4]]. *We assume that **(A1)** is fulfilled. Let  $f \in \mathcal{B}(S^3)$  such that  $Pf$  and  $Pf^2$  exists and belong to  $F$  and  $(\mu, Pf) = 0$ . Then for all  $\delta > 0$  and all  $r \in \mathbb{N}$*

$$\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(f)| > \delta\right) \leq \begin{cases} \frac{c'}{\delta^2} \left(\frac{1}{2}\right)^{r+1} & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c'}{\delta^2} r \left(\frac{1}{2}\right)^{r+1} & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c'}{\delta^2} \alpha^{2(r+1)} & \text{if } \alpha^2 > \frac{1}{2}; \end{cases}$$

where the positive constant  $c'$  depends on  $f$  and  $\alpha$ . Furthermore, if  $Pf = 0$ , we have

$$\mathbb{P}\left(|\overline{M}_{\mathbb{T}_r}(f)| > \delta\right) \leq \frac{c'}{\delta^4} \left(\frac{1}{4}\right)^{r+1}.$$

*Ideas for the proofs.* The proof of Theorems 3.3 and 3.4 is based on two points: Markov inequality and control of fourth order moment (Theorems 3.3) and second order moment (Theorems 3.4) of  $\overline{M}_{\mathbb{T}_r}(f)$ . These controls are done using explicit calculations (via successive conditioning) of fourth and second order moments of  $\overline{M}_{\mathbb{G}_r}(f)$  (where  $\overline{M}_{\mathbb{G}_r}(f) = \frac{1}{|\mathbb{G}_r|} \sum_{i \in \mathbb{G}_r} f(X_i)$ ), hypothesis **(A1)** and hypothesis (i)-(vi) on  $F$ , and the relation  $\overline{M}_{\mathbb{T}_r}(f) = \sum_{q=0}^r \frac{|\mathbb{G}_q|}{|\mathbb{T}_r|} \overline{M}_{\mathbb{G}_q}(f)$ .  $\square$

**Remark 3.5.** Notice that the dichotomy around the value  $\alpha^2 = \frac{1}{2}$  naturally appears in the calculus.

Let  $(\mathcal{H}_n)_{n \geq 1}$  be the filtration defined by

$$\mathcal{H}_0 = \sigma(X_1) \quad \text{and} \quad \mathcal{H}_n = \sigma(\Delta_{\Pi(i)}, \Pi(i+1), 1 \leq i \leq n)$$

where  $\Delta_{\Pi(i)} = (X_{\Pi(i)}, X_{2\Pi(i)}, X_{2\Pi(i)+1})$  and  $\Pi$  is an application which allows to create a random order on the population  $\mathbb{T}$  which preserves the genealogical order (we refer to [4, 19] for more details on this application). Then, we have the following result on MDP.

**Theorem 3.6.** [Bitseki Penda-Djellout-Guillin [4]]. *Let  $(b_n)$  be a sequence of increasing positive real numbers satisfying*

$$\frac{b_n}{\sqrt{n}} \rightarrow +\infty, \quad \frac{b_n}{\sqrt{n \log n}} \rightarrow 0.$$

*Let  $f \in \mathcal{B}(S^3)$  such that  $Pf = 0$ ,  $Pf^2$  and  $Pf^4$  exist and belong to  $F$ . Assume also that*

$$\limsup_{n \rightarrow \infty} \frac{n}{b_n^2} \log \left( n \operatorname{ess\,sup}_{1 \leq k \leq c^{-1}(b_{n+1})} \mathbb{P}\left(|f(\Delta_{\Pi(k)})| > b_n | \mathcal{H}_{k-1}\right) \right) = -\infty$$

where  $c^{-1}(b_{n+1}) := \inf \{k \in \mathbb{N} : \frac{k}{b_k} \geq b_{n+1}\}$ . Then for all  $\delta > 0$ , we have

$$\lim_{r \rightarrow \infty} \frac{|\mathbb{T}_r|}{b_{|\mathbb{T}_r|}^2} \log \mathbb{P}\left(\frac{1}{b_{|\mathbb{T}_r|}} |M_{|\mathbb{T}_r|}(f)| > \delta\right) = -\frac{\delta^2}{2(\mu, Pf^2)}.$$

*Ideas for the proof.* The proof of Theorem 3.6 is based on deviation inequalities (8) and the MDP for the martingale.  $\square$

Now, under the stronger assumption **(A2)**, we have the following more sharp estimations.

**Theorem 3.7.** [Bitseki Penda-Djellout-Guillin [4]]. *Let  $f \in \mathcal{B}_b(S)$  such that  $(\mu, f) = 0$ . Then for all  $\delta > 0$  we have*

$$\mathbb{P}\left(\overline{M}_{\mathbb{T}_r}(f) > \delta\right) \leq \begin{cases} \exp(c'\delta) \exp(-c'\delta^2|\mathbb{T}_r|), & \forall r \in \mathbb{N}, \text{ if } \alpha < \frac{1}{2}; \\ \exp(2c'\delta(r+1)) \exp(-c'\delta^2|\mathbb{T}_r|), & \forall r \in \mathbb{N}, \text{ if } \alpha = \frac{1}{2}; \\ \exp(-c'\delta^2|\mathbb{T}_r|), & \forall r > r_0 - 1, \text{ if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}; \\ \exp\left(-c'\delta^2 \frac{|\mathbb{T}_r|}{r+1}\right), & \forall r > r_0 - 1, \text{ if } \alpha = \frac{\sqrt{2}}{2}; \\ \exp\left(-c'\delta^2 \frac{1}{\alpha^2(r+1)}\right), & \forall r > r_0 - 3, \text{ if } \alpha > \frac{\sqrt{2}}{2}; \end{cases} \quad (9)$$

where  $r_0 := \log(\delta/c_0)/\log(\alpha)$ , and  $c_0, c'$  and  $c''$  are positive constants which depend on  $\alpha, \|f\|_\infty$  and  $c$ .

If  $f \in \mathcal{B}_b(S^3)$  such that  $(\mu, Pf) = 0$  then we have the same conclusions for  $\overline{M}_{\mathbb{T}_r}(f)$ .

*Ideas for the proof.* The proof is based on Chernoff inequality, successive conditioning and successive applications of Azuma-Bennet-Hoeffding using **(A2)**.  $\square$

**Remark 3.8.** Once again, notice that the dichotomy around  $\alpha = \frac{1}{2}$  and  $\alpha^2 = \frac{1}{2}$  in (9) naturally appears from the calculations.

**Theorem 3.9.** [Bitseki Penda-Djellout-Guillin [4]]. *Let  $(b_n)$  be an increasing sequence of positive real numbers such that*

$$(v1) \quad \frac{b_n}{\sqrt{n}} \longrightarrow +\infty,$$

$$(v2) \quad \text{if } \alpha^2 < \frac{1}{2}, \text{ the sequence } (b_n) \text{ is such that } \frac{b_n}{n} \longrightarrow 0,$$

$$(v3) \quad \text{if } \alpha^2 = \frac{1}{2}, \text{ the sequence } (b_n) \text{ is such that } \frac{b_n \log n}{n} \longrightarrow 0,$$

$$(v4) \quad \text{if } \alpha^2 > \frac{1}{2}, \text{ the sequence } (b_n) \text{ is such that } \frac{b_n \alpha^{r_n+1}}{\sqrt{n}} \longrightarrow 0.$$

Let  $f \in \mathcal{B}_b(S^3)$  such that  $Pf = 0$ . Then  $\left(\frac{1}{\sqrt{|\mathbb{T}_r|}} M_{\mathbb{T}_r}(f)\right)$  satisfies a MDP in  $\mathbb{R}$  with the speed  $\frac{|\mathbb{T}_r|}{b^2_{|\mathbb{T}_r|}}$  and rate function  $I(x) = \frac{x^2}{2(\mu, Pf^2)}$ .

**Remark 3.10.** The conditions (v2)-(v4) come from deviation inequalities (9).

*Ideas for the proof.* The proof is based on deviation inequalities (9) and MDP for bounded martingale.  $\square$

### 3.3. Application

We consider the first order bifurcating autoregressive process (BAR(1)).

$$\mathcal{L}(X_1) = \nu, \quad \text{and} \quad \forall n \geq 1, \quad \begin{cases} X_{2n} = \alpha_0 X_n + \beta_0 + \varepsilon_{2n} \\ X_{2n+1} = \alpha_1 X_n + \beta_1 + \varepsilon_{2n+1}, \end{cases} \quad (10)$$

where  $\alpha_0, \alpha_1 \in (-1, 1)$ ;  $\beta_0, \beta_1 \in \mathbb{R}$ ,  $((\varepsilon_{2n}, \varepsilon_{2n+1}))_{n \geq 1}$  forms a sequence of i.i.d. bivariate random variables and  $\nu$  a probability measure on  $\mathbb{R}$ . This model is a typical example of bifurcating markovian dynamics and it has been the motivation for the rigorous mathematical study of BMC in [19]. We assume that  $\nu$  has finite moments

of all orders. The least square estimator  $\hat{\theta}^r$  of  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$  is given by, for  $\eta \in \{0, 1\}$

$$\begin{cases} \hat{\alpha}_\eta^r = \frac{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i X_{2i+\eta} - \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i \right) \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\eta} \right)}{|\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i^2 - \left( |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i \right)^2} \\ \hat{\beta}_\eta^r = |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_{2i+\eta} - \hat{\alpha}_\eta^r |\mathbb{T}_r|^{-1} \sum_{i \in \mathbb{T}_r} X_i. \end{cases}$$

The BAR(1) processes are an adaptation of autoregressive processes, when the data have a binary tree structure. They were first introduced by Cowan and Staudte [18] for cell lineage data where each individual in one generation gives rise to two offspring in the next generation.

In [19], Guyon, after establishing the first results on the theory of BMC, proves laws of large numbers and CLT for the least-square estimators  $\hat{\theta}^r = (\hat{\alpha}_0^r, \hat{\beta}_0^r, \hat{\alpha}_1^r, \hat{\beta}_1^r)$  of the 4-dimensional parameter  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1)$ .

He also gives some statistical tests which allow to check if the model is symmetric or not (roughly  $\alpha_0 = \alpha_1$  or not), and if the new pole and the old pole populations are even distinct in mean, which allows him to conclude a statistical evidence in aging in E. Coli.

### 3.3.1. The gaussian setting

First we consider that  $((\varepsilon_{2n}, \varepsilon_{2n+1}))_{n \geq 1}$  forms a sequence of i.i.d. bivariate random variables with law  $\mathcal{N}_2(0, \Gamma)$  with

$$\Gamma = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}, \quad \sigma^2 > 0, \quad \rho \in (-1, 1).$$

We take  $F = \mathcal{C}_{pol}^1(\mathbb{R})$  where  $\mathcal{C}_{pol}^1(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} / \exists c > 0, \exists m \in \mathbb{N}, |f(x)| + |f'(x)| \leq c(1 + |x|^m)\}$ . Then  $F$  satisfies hypothesis (i)-(vi). **(A1)** are automatically satisfied with  $\alpha = \max(|\alpha_0|, |\alpha_1|)$ .

Let us define two continuous functions  $\mu_1 : \Theta \rightarrow \mathbb{R}$  and  $\mu_2 : \Theta \times \mathbb{R}_+^* \rightarrow \mathbb{R}$  by writing  $(\mu, \mathbf{x}) = \mu_1(\theta)$  and  $(\mu, \mathbf{x}^2) = \mu_2(\theta, \sigma^2)$ , where  $\theta = (\alpha_0, \beta_0, \alpha_1, \beta_1) \in \Theta = (-1, 1) \times \mathbb{R} \times (-1, 1) \times \mathbb{R}$ , and  $\mu$  is the stationary distribution of  $Q$ . Then, we have the following deviation inequalities.

**Proposition 3.11.** [Bitseki Penda-Djellout-Guillin [4]]. *For all  $\delta > 0$ , for all  $r \in \mathbb{N}$  and for all  $\gamma < \min\left(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}}\right)$ , where  $c_1 = c_1(\mu_1) > 0$ , we have*

$$\mathbb{P}\left(\left\|\hat{\theta}^r - \theta\right\| > \delta\right) \leq \begin{cases} \frac{c}{\gamma^{4q}\delta^{4-p}} \left(\frac{1}{4}\right)^{r+1} & \text{if } \alpha^2 < \frac{1}{2}; \\ \frac{c}{\gamma^{4q}\delta^{4-p}} r^2 \left(\frac{1}{4}\right)^{r+1} & \text{if } \alpha^2 = \frac{1}{2}; \\ \frac{c}{\gamma^{4q}\delta^{4-p}} \alpha^{4(r+1)} & \text{if } \alpha^2 > \frac{1}{2}; \end{cases}$$

where  $c = c(\alpha, \mu_1, \mu_2) > 0$ ,  $p = p(\delta) \in \{0, 2, 4\}$  and  $q = q(\delta) \in \{0, 1\}$ .

### 3.3.2. Bounded setting

Now assume that the noise values in a compact set. We set  $F = \mathcal{C}_b^1(\mathbb{R})$  the set of all  $\mathcal{C}^1$  functions bounded on  $\mathbb{R}$ . Then **(A2)** are automatically satisfied with  $\alpha = \max(|\alpha_0|, |\alpha_1|)$ . For all  $\delta > 0$  and for all  $\gamma < \min\left(\frac{c_1 b}{1+\delta}, \frac{c_1 b}{1+\sqrt{\delta}}, \frac{c_1 b}{1+\sqrt[4]{\delta}}\right)$  where  $c_1$  is a positive constant which depends on  $\mu_1$ , let  $r_0 := \frac{\log(\gamma^q \delta^{1-p/2}/c_0)}{\log \alpha}$ , where  $c_0 = c_0(\alpha, c, \gamma)$ ,  $p \in \{0, 1, 3/2\}$  and  $q \in \{0, 1\}$ . Then we have the following deviation inequalities.

**Proposition 3.12.** [Bitseki Penda-Djellout-Guillin [4]]. *We have*

$$\mathbb{P}\left(\left\|\hat{\theta}^r - \theta\right\| > \delta\right) \leq \begin{cases} c_2 \exp\left(c''\gamma^q \delta^{1-p/2}\right) \exp\left(-c'\gamma^{2q} \delta^{2-p} |\mathbb{T}_r|\right), & \forall r \in \mathbb{N}, \text{ if } \alpha < \frac{1}{2}; \\ c_2 \exp\left(c'\gamma^q \delta^{1-p/2}(r+1) - c'\gamma^{2q} \delta^{2-p} |\mathbb{T}_r|\right), & \forall r \in \mathbb{N}, \text{ if } \alpha = \frac{1}{2}; \\ c_2 \exp\left(-c'\gamma^{2q} \delta^{2-p} |\mathbb{T}_r|\right), & \forall r > r_0, \text{ if } \frac{1}{2} < \alpha < \frac{\sqrt{2}}{2}; \\ c_2 \exp\left(-c'\gamma^q \delta^{2-p} \frac{|\mathbb{T}_r|}{r+1}\right), & \forall r > r_0, \text{ if } \alpha = \frac{\sqrt{2}}{2}; \\ c_2 \exp\left(-c'\gamma^{2q} \delta^{2-p} \frac{1}{\alpha^2(r+1)}\right), & \forall r > r_0, \text{ if } \alpha > \frac{\sqrt{2}}{2}; \end{cases}$$

where  $c_2$  is a positive constant,  $c'$  and  $c''$  depend on  $\alpha$ , and  $c$ .

#### 4. LARGE DEVIATIONS OF THE “TRUE” SELF-REPELLING MOTION

In this section, we explain some features related to large deviations of a self-interacting one-dimensional process called the “true” self-repelling motion (TSRM), defined by Bálint Tóth and Wendelin Werner in [30] which were established in [14]. Let us first very briefly recall the intuitive definition of this process and describe the motivations that lead to our study. The TSRM is a continuous real-valued process  $(X_t, t \geq 0)$  that is locally self-interacting with its past occupation-time. More precisely, for each positive time  $t$ , define its occupation-time measure  $\mu_t$  that assigns to each interval  $I \in \mathbb{R}$ , the time spent in it by  $X$  before time  $t$ :

$$\mu_t(I) = \int_0^t 1_{\{X_s \in I\}} ds$$

It turns out that for this particular process  $X$ , almost surely for each  $t$ , the measure  $\mu_t$  has a continuous density  $x \mapsto L_t(x)$  with respect to the Lebesgue measure we will call *local time* by analogy with semi-martingales. Heuristically, the dynamics of  $X_t$  is such that the TSRM is locally pushed in the direction of the negative “gradient” of its local time at its current position. Loosely formulated, one can write  $dX_t = -\nabla_x L_t(X_t) dt$  (even if  $(X_t, t \geq 0)$  is a random process). For more details and comments on this description, we refer to [30].

When a process admits a local time, we may assign a height to the process which simply represents the time spent on the current position. For our process, we will denote by  $H_t$  this height i.e.  $H_t := L_t(X_t)$ . Notice that in this way, one obtains a space-filling 1+1-dimensional curve  $(X_t, H_t)$ .

It turns out that this process is of a very different type than diffusions. For example (see again [30]), its quadratic variation almost surely vanishes whereas its variation of power  $3/2$  is positive and finite. Similarly, it does not have the Brownian scaling property, it has instead a  $2/3$  scaling behavior i.e., for any positive  $\lambda$ ,  $(X_{\lambda t}, t \geq 0)$  has the same law as  $(\lambda^{2/3} X_t, t \geq 0)$ .

The construction of the process  $X_t$  is based on a family of coalescing one-dimensional Brownian motions starting from all points in the plane. Such families had been constructed by Arratia and are now called “Brownian web”. As a consequence, the estimates on the TSRM follow from results concerning this Brownian web.

The TSRM seems at present to be one of the few such “non-diffusive” continuous processes that probabilists can define. This gives us some motivation to study in more detail its behavior. There exists two main versions of the TSRM, a stationary one and a zero-initial conditions one (see [14]). For the sake of simplicity, we only state here the results of [14] about the *stationary* TSRM.

First, let us write both for the process  $(X_t, t \geq 0)$  itself as for the height process  $(H_t, t \geq 0)$  upper and lower bounds for the probability that their value at a given time is very large. More precisely, we have:

**Proposition 4.1** (Dumaz). *When  $x \rightarrow \infty$ ,*

$$\mathbb{P}(X_1 \geq x) \leq \exp\left(-\frac{4|a'_1|^3}{27}x^3 + O(\ln(x))\right)$$

where  $a'_1$  is the first negative root of the derivative of the Airy function.

Moreover, there exist  $c_1, c'_1, c_2$  and  $c'_2$  such that for all  $h > 0$ ,

$$\begin{aligned} \exp(-c_1 h^{3/2}) &\leq \mathbb{P}(H_1 > h) \leq \exp(-c'_1 h^{3/2}) \\ \exp(-c_2 h^{3/2}) &\leq \mathbb{P}(H_1 < -h) \leq \exp(-c'_2 h^{3/2}). \end{aligned}$$

This enables us to derive almost sure fluctuation results (of the type of the law of the iterated logarithm) for these two processes. Note that we need a 0 – 1-law arguments to derive the result about the height process.

**Proposition 4.2** (Dumaz). *Almost surely, we have:*

$$\begin{aligned} \limsup_{t \rightarrow 0} t^{-2/3} (\ln(\ln(1/t)))^{-1/3} X_t &= 3/(2^{2/3}|a'_1|) \\ \limsup_{t \rightarrow +\infty} t^{-2/3} (\ln(\ln(t)))^{-1/3} X_t &= 3/(2^{2/3}|a'_1|). \end{aligned}$$

And there exist two positive constants  $m$  and  $m'$  such that almost surely,

$$\begin{aligned} \limsup_{t \rightarrow 0} t^{-1/3} (\ln(\ln(1/t)))^{-2/3} H_t &= m \\ \limsup_{t \rightarrow +\infty} t^{-1/3} (\ln(\ln(t)))^{-2/3} H_t &= m'. \end{aligned}$$

Let us point out that the process  $X$  does not have independent increments, so that the standard proof of the LIL for Brownian motion can not be adapted directly for the lower bound.

## REFERENCES

- [1] Arcones M. A. The large deviation principle for stochastic processes I. *Theory of Probability and its Applications* 47, 567-583 (2003a).
- [2] Arcones M. A. The large deviation principle for stochastic processes II. *Theory of Probability and its Applications* 48, 19-44 (2003b).
- [3] Bercu B. and Proïa F. A sharp analysis on the asymptotic behavior of the Durbin-Watson statistic for the first-order autoregressive process. *ESAIM Probab. Stat.* 16 (2012).
- [4] Bitseki Penda V., Djellout H. and Guillin A. Deviation inequalities, Moderate deviations and some limit theorems for bifurcating Markov chains with application. *Accepted for publication in Annales of applied probability* (2013).
- [5] Bitseki Penda V., Djellout H., and Proïa F. Moderate deviations for the Durbin-Watson statistic related to the first-order autoregressive process. *Accepted for publication in ESAIM: Probability and Statistics* (2013).
- [6] Bryc W. and Dembo A. Large deviations and strong mixing. *Ann. Inst. H. Poincaré Probab. Stat.* 32, 549-569 (1996)
- [7] Dedecker J., Merlevède F. and Volny D. On the weak invariance principle for non-adapted sequences under projective criteria. *J. Theoret. Probab.* 20 (2007), no. 4, 971-1004.
- [8] Dedecker J. Merlevède F. Peligrad M. and Utev S. (2009). Moderate deviations for stationary sequences of bounded random variables. *Ann. Inst. Henri Poincaré Probab. Stat.* 45, no. 2, 453-476.
- [9] Dedecker J. and Priour C. Coupling for  $\tau$ -dependent sequences and applications. *J. Theoret. Probab.* 17 (2004), no. 4, 861-885.
- [10] Delyon B., Juditsky A. and Liptser R. Moderate deviation principle for ergodic Markov chain. Lipschitz summands. *In From Stochastic Calculus to Mathematical Finance 189-209. Springer, Berlin, 2006.*
- [11] Djellout H Moderate deviations for Martingale Differences and applications to  $\alpha$ -mixing sequences. *Stoch. Stoch. Rep.* 73 (1-2), 37-63 (2002).
- [12] Djellout H. and Guillin A. Moderate deviations for Markov chains with atom. *Stoch. Process. Appl.* 95, 203-217 (2001)
- [13] Dembo A. and Zeitouni O. Large Deviations Techniques and Applications. 2nd edn. Springer, New York (1998)



- [14] Dumaz L. Large deviations and path properties of the true self-repelling motion. *submitted (2011)*, <http://arxiv.org/abs/1105.2948>
- [15] Durbin J. and Watson G. S. Testing for serial correlation in least squares regression. I-II-III. *Biometrika* 37-38-57 (1950-51-71).
- [16] Eichelsbacher P. and Matthias L. Moderate deviations for i.i.d. random variables. *ESAIM Probab. Stat.* 7 (2003), 209-218
- [17] Gao F. Q. Moderate deviations for martingales and mixing random processes. *Stochastic Process. Appl.* 61, 263-275 (1996).
- [18] Guyon J. Bize A. Paul G. Stewart E. J. Delmas J. F. and Taddéi F. Statistical study of cellular aging. *CEMRACS 2004 Proceedings, ESAIM Proceedings, (2005), 14, pp. 100-114.*
- [19] Guyon J. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* 17 (2007), no. 5-6, 1538-1569.
- [20] Heyde C.C. On the central limit theorem for stationary processes. *Z. Wahrsch. verw. Gebiete* 30, 315-320 (1974).
- [21] Maxwell M. and Woodroffe M. Central limit theorems for additive functionals of Markov chains. *Ann. Probab.* 28 (2000), no. 2, 713-724.
- [22] Merlevède F., Peligrad M. and Rio E. (2011). A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields* 151, no. 3-4, 435-474.
- [23] Merlevède F. and Peligrad M. Moderate deviations for linear processes generated by martingale-like random variables. *J. Theoret. Probab.* 23 (2010), no. 1, 277-300.
- [24] Merlevède F. , Peligrad M. and Rio E. Bernstein inequality and moderate deviations under strong mixing conditions. *High dimensional probability V: the Luminy volume, 272-292, Inst. Math. Stat. Collect., 5, Inst. Math. Statist., Beachwood, OH, 2009.*
- [25] Nerlove, M. and Wallis K. F. Use of the Durbin-Watson statistic in inappropriate situations. *Econometrica* 34 (1966).
- [26] Park S. B. On the small-sample power of Durbin's h test. *Journal of the American Statistical Association* 70 (1975).
- [27] Peligrad M. and Utev S. A new maximal inequality and invariance principle for stationary sequences. *Ann. Probab.* 33 (2005), no. 2, 798-815.
- [28] Peligrad M., Utev S. and Wu W. B. A maximal Lp-inequality for stationary sequences and its applications. *Proc. Amer. Math. Soc.* 135 (2007), no. 2, 541-550.
- [29] Puhalskii A. Large deviations of semimartingales via convergence of the predictable characteristics. *Stoch. Stoch. Rep.* 49, 27-85 (1994).
- [30] Tóth, B., Werner W. The true self-repelling motion. *Probab. Theory Related Fields* 111, 375-452 (1998).
- [31] Worms J. Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.* 10 (2001), no. 1, 38-72.
- [32] Worms J. Moderate deviations for stable Markov chains and regression models. *Electron. J. Probab.* 4 (1999), no. 8, 28 pp.