Moderate deviations for Markov chains with atom

H. Djellout, A. Guillin *

*Laboratoire de Mathématiques Appliquées, CNRS-UMR 6620, Université Blaise Pascal,
24 Avenue des Landais, 63177 Aubière, France*

Abstract

We obtain in this paper moderate deviations for functional empirical processes of general state space valued Markov chains with atom under weak conditions: a tail condition on the first time of return to the atom, and usual conditions on the class of functions. Our proofs rely on the regeneration method and sharp conditions issued of moderate deviations of independent random variables. We prove our result in the nonseparable case for additive and unbounded functionals of Markov chains, extending the work of de Acosta and Chen (J. Theoret. Probab. (1998) 75–110) and Wu (Ann. Probab. (1995) 420–445). One may regard it as the analog for the Markov chains of the beautiful characterization of moderate deviations for i.i.d. case of Ledoux 1992. Some applications to Markov chains with a countable state space are considered. © 2001 Elsevier Science B.V. All rights reserved.

MSC: primary 60F10

Keywords: Moderate deviations; Markov chains; Regeneration chain method; Functional empirical processes; Countable state space

1. Introduction and main result

Let $(E, \mathscr{E})$ be a measurable space and $\mathscr{M}(E)$ be the space of all finite signed measures on $(E, \mathscr{E})$ equipped with the total variation norm $\| \cdot \|_{\text{var}}$. Let $\{X_j\}_{j \geqslant 0}$ be an E-valued irreducible ergodic Markov chain with transition probability P and invariant probability measure π. Throughout the paper, we assume that the chain $\{X_j\}$ has an atom, i.e. $\exists \alpha \subset E$ with $\pi(\alpha) > 0$, v a probability measure such that

$$\forall x \in \alpha, \qquad P(x, \cdot) = v(\cdot), \tag{1.1}$$

α is then called a atom. Note that, when the state space is discrete, every state charged by π is an atom. We introduce the first time of entrance of the chain in this atom

---

* Corresponding author. Tel.: +33-4-73-40-70-50; fax: +33-4-73-40-70-64.

E-mail addresses: djellout@math.univ-bpclermont.fr (H. Djellout), arnaud.guillin@math.univ-bpclermont.fr (A. Guillin).

which will play an important role in the study:

$$\tau = \inf\{n \geqslant 0; \ X_n \in \alpha\},$$

and we will always assume that $\mathbb{E}_\nu \tau^2 < \infty$.

Given a probability measure μ on $(E, \mathcal{E})$, $\mathbb{P}_\mu$ will be the Markovian probability measure on $(E^{\mathbb{N}}, \mathcal{E}^{\otimes \mathbb{N}})$ determined by the transition probability P and the initial law μ. $\{X_j\}_{j \geqslant 0}$ will be then the sequence of coordinates on $E^{\mathbb{N}}$.

Let M_n, $n \geqslant 1$, be random elements of $\mathcal{M}(E)$ defined by

$$M_n := \frac{1}{b_n} \sum_{j=0}^{n-1} (\delta_{X_j} - \pi), \tag{1.2}$$

where b_n is a sequence of positive numbers tending to infinity. We are interested in this paper in the asymptotic behaviour of $\mathbb{P}_\mu(M_n \in \cdot)$.

When $b_n = \sqrt{n}$, it is the Central Limit Theorem obtained first by Nummelin (1978) and Chen (1997) under various conditions. If $b_n = n$, it is the large deviations case extensively studied since the pioneering works of Donsker–Varadhan (see for instance Deuschel and Stroock, 1989; Wu, 1993 for a survey on this topic).

Now assume,

$$\frac{b_n}{\sqrt{n}} \uparrow +\infty, \qquad \frac{b_n}{n} \to 0. \tag{1.3}$$

The estimation of the probabilities $\mathbb{P}_\mu(M_n \in \cdot)$ is usually called the moderate deviation problem. We will suppose moreover the following: $\exists A > 1$, $0 < \delta < 1$ such that

$$\forall n, k \geqslant 1, \qquad b_{nk} \leqslant A k^{1-\delta} b_n. \tag{1.4}$$

It is the usual condition on the speed of moderate deviations in the i.i.d. case (Ledoux, 1992), it means that b_n cannot be too near of n (the scale of large deviations). Sharp results on moderate deviation are quite recent, even for the i.i.d. case: the works of Ledoux (1992) for the upper bound in Banach space (which are largely used in this paper) and results of Wu (1994) for the functional empirical process (nonseparable Banach space case). See also Djellout (2000) for the extension to the martingale differences case and applications to mixing sequences.

The Markovian case has been studied under successively less restrictive conditions (Mogulskii (1984), Gao (2000), Wu (1994)) and recently under weak conditions by de Acosta (1988a,b) and Chen (1997) for the lower bound (under different and non-comparable conditions) and by de Acosta and Chen (1998), and Chen (1997) (under same conditions but different proof) for the upper bound. de Acosta and Chen (1998) have established their results under the assumptions of geometric ergodicity and a regularity condition (de Acosta and Chen, 1998, assumption (1.5)). Very recently, Guillin (2000) extends their results to the uniform trajectorial case, and Guillin, 2001 for Markov processes (continuously indexed).

We will be interested here by the asymptotic behaviour of M_n uniformly over a class of function (context of Wu, 1994).

Given a class of real measurable functions $\mathcal{F}$ such that $\forall f \in \mathcal{F}$, $\pi(f) = 0$, $f \in L_2(\pi)$ and $\mathbb{E}_\nu(\sum_{j=1}^{\tau} \|f(X_j)\|)^2 < \infty$, let $l_\infty(\mathcal{F})$ be the space of all bounded real functions on $\mathcal{F}$ with norm $\|F\|_{\mathcal{F}} = \sup_{f \in \mathcal{F}} |F(f)|$.

If $\mathscr{F}$ is infinite, $l_\infty(\mathscr{F})$ is a nonseparable Banach space. Every $\beta \in \mathscr{M}(E)$ can be regarded as an element $\beta^{\mathscr{F}} \in l_\infty(\mathscr{F})$ given by $\beta^{\mathscr{F}}(f) = \beta(f) = \int_E f \, d\beta$. We will now establish the moderate deviations estimations of $(M_n)^{\mathscr{F}}$ in $l_\infty(\mathscr{F})$.

In the sequel, we will suppose that $\mathscr{F}$ is countable, or that the processes $\{M_n(f); f \in \mathscr{F}\}$ are separable in the sense of Doob, to avoid measurability problems. Let d_2 be the following metric for $\mathscr{F}$: $\forall f, g \in \mathscr{F}$:

$$d_2(f, g) = \sigma(f - g),$$

where

$$\sigma^2(f) = \lim_{n \to \infty} \frac{1}{n} \mathbb{E} \left(\sum_{k=0}^{n-1} f(X_k) \right)^2 = \pi(\alpha) \mathbb{E}_v \left(\sum_{j=1}^{\tau} f(X_j) \right)^2$$

is the associated variance.

For an irreducible Markov chain taking integer values which has a finite second moment for the first return time from some integer to itself, Levental (1990) find necessary and sufficient conditions for the uniform CLT over all subsets of the integers. Tsai (1997) generalized this result to unbounded classes, $\mathscr{F} = \{f: |f| \leqslant F\}$, where F is a non-negative function, say the envelope function, on the countable state space. Tsai (2000) gives sufficient and nearly necessary conditions (weaker than condition of the uniform CLT) for the compact and bounded law of the iterated logarithm for Markov chains with a countable state space.

We will first give the moderate deviation principle in the general framework where an atom is present, and then present some applications on a countable state space, where some conditions can be more explicit.

Here is our main result:

Theorem 1. *Suppose that $(\mathscr{F}, d_2)$ is totally bounded and $(M_n)^{\mathscr{F}} \to 0$ in probability in $l_\infty(\mathscr{F})$. Assume*
(H1) $\limsup_{n \to +\infty} n/b_n^2 \log(n \mathbb{P}_v(\tau \geqslant b_n)) = -\infty$,
(H2) $\limsup_{n \to +\infty} n/b_n^2 \log(n \mathbb{P}_v(\sum_{k=0}^{\tau} \|f(X_k)\|_{\mathscr{F}} \geqslant b_n)) = -\infty$.
Then for every probability measure μ on $(E, \mathscr{E})$ verifying

$$\limsup_{n \to +\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sum_{k=0}^{\tau} \|f(X_k)\|_{\mathscr{F}} \geqslant b_n \right) = -\infty. \tag{1.5}$$

$\mathbb{P}_\mu((M_n)^{\mathscr{F}} \in \cdot)$ *satisfies a moderate deviation principle on $l_\infty(\mathscr{F})$ with speed b_n^2/n and good rate function $J_{\mathscr{F}}$ given by*

$$J_{\mathscr{F}}(F) = \sup\{J_{(f_1,\dots,f_m)}(F(f_1),\dots,F(f_m)); f_1,\dots,f_m \in \mathscr{F}, m \geqslant 1\},$$

where $J_{(f_1,\dots,f_m)}$ is given by

$$J_f(x) = \sup_{\xi \in \mathbb{R}^m} \left[\langle x, \xi \rangle - \frac{1}{2} \sigma^2(\langle f, \xi \rangle) \right]. \tag{1.6}$$

Suppose moreover $\sum_{k=1}^{\infty} \langle \xi, f \rangle P^k \langle f, \xi \rangle \in L_1(\pi)$ *for all* $\xi \in \mathbb{R}^m$, *then*

$$\sigma^2(\langle f, \xi \rangle) = \int \langle f, \xi \rangle^2 \, d\pi + 2 \int \sum_{k=1}^{\infty} \langle f, \xi \rangle P^k \langle f, \xi \rangle \, d\pi.$$

Remarks. (i) Note that when an atom is present, the geometric ergodicity condition is equivalent to

$$\exists \delta > 0 \text{ such that } \mathbb{E}_v(e^{\delta \tau}) < \infty.$$

Condition (H1) is then strictly weaker than the geometric ergodicity imposed in the work de Acosta and Chen (1998). Moreover (H1) can be more explicitly given. For example, in the particular case $b_n = n^{1/p}$ with $1 < p < 2$, for which conditions (1.3) and (1.4) are then obviously verified, then (H1) is easily seen to be implied by

$$\exists \delta > 0 \text{ such that } \mathbb{E}_v(e^{\delta \tau^{2-p}}) < \infty. \tag{1.7}$$

Remark also that we consider here the nonseparable case of the functional empirical process and unbounded functions, cases which are not studied by de Acosta and Chen. To their credit, note however that they suppose neither the existence of an atom nor the condition (1.4) on (b_n) and their sole assumption is the well known geometric ergodicity.

(ii) Still in the context $b_n = n^{1/p}$ with $1 < p < 2$, following Nummelin and Tweedie (1978) and Nummelin and Tuominen (1982) (or Meyn and Tweedie, 1993 for a complete review) one can see that condition (1.7) is equivalent to the following *sub-geometric ergodicity*: there exists $r > 1$ such that for π-a.e. x (with $\| \cdot \|_V$ denoting the total variation norm)

$$\sum_{n=1}^{\infty} r^{n^{2-p}} \|P^n(x, \cdot) - \pi\|_V < \infty, \tag{1.8}$$

which is stronger than ergodicity of degree 2 (see Chen, 1999) but weaker than geometric ergodicity. Such an assertion implies in particular that (1.7) is valid independently of the choice of the atom and so (H1) in this context.

We have not been able to derive the independence of the recurrence condition (H1) on atom nor its characterization by means of some type of ergodicity for general b_n, but fortunately our results are proved if (H1) and (H2) are satisfied by some and then any atom.

(iii) Under (H2), condition (1.5) is verified, for instance, by the invariant measure π of the Markov chain and then by the Dirac measure δ_x for π-a.e. $x \in E$, see the appendix.

2. Applications to Markov chains with a countable state space

We will give in this section some applications where some conditions can be given explicitly, more precisely when the total boundedness of $\mathscr{F}$ with respect to the pseudo-metric d_2 or $(M_n)^{\mathscr{F}} \to 0$ in probability can be proved under satisfying hypotheses. We are much inspired here by the works of Levental (1990) and Tsai (1997).

We then consider the MDP for Markov chains with a countable state space $E = \{1, 2, 3 \ldots\}$. Here τ_i will be the ith hitting time of state 1, i.e.

$$\tau := \tau_1 = \min\{n : n \geqslant 1, X_n = 1\} \quad \text{and for } i > 1, \quad \tau_i = \min\{n : n > \tau_{i-1}, X_n = 1\},$$

and $m_{i,j}$ be the expected minimal number of steps from state i to state j, i.e.

$$m_{i,j} = \mathbb{E}(\min\{n : n \geqslant 1, X_n = j\} \mid X_0 = i).$$

Let us first consider the case where $\mathscr{F}$ is the family of all indicator functions, i.e. $\mathscr{F} = \{1_A - \pi(A) : A \subset E\}$, related with Kolmogorov–Smirnov nonparametrical statistics.

Corollary 2. *Assume that* (H1) *is satisfied and*

$$\sum_{k=1}^{+\infty} \pi(k) \sqrt{m_{1,k}} < \infty, \tag{2.1}$$

for all orderings of E. Then for every probability measure μ satisfying (1.5), *the MDP of Theorem* 1 *holds for the family of all indicator functions on E.*

Proof. Condition (2.1) is the necessary and sufficient condition for the uniform CLT over all subsets of the integers for Markov chains satisfying $\mathbb{E}(\tau_2 - \tau_1)^2 < \infty$ by Leventhal (1990). The uniform CLT implies in particular $M_n^{\mathscr{F}} \to 0$ in probability and $(\mathscr{F}, d_2)$ totally bounded. For this family of indicator functions $\mathscr{F}$, (H2) is identical to (H1). The proof is completed by Theorem 1.

In the particular case of the law of the iterated logarithm $(b_n = \sqrt{2n \log \log n})$, we have the following:

Corollary 3. *Assume that*

$$\mathbb{E}_v(\tau^2 (\log \tau)^a) < \infty, \quad \forall a > 0. \tag{2.2}$$

Suppose moreover

$$\frac{1}{\sqrt{\log \log n}} \sum_{k=1}^{n} \pi(k) \sqrt{m_{1,k}} \to 0, \tag{2.3}$$

for all orderings of E. Then for every probability measure μ satisfying (1.5), *the MDP of Theorem* 1 *holds for the family of all indicator functions on E and for $b_n = \sqrt{2n \log \log n}$.*

Proof. Remark that by Theorem 1, and by the fact that Ledoux (1992, Corollaire 2) shows that (H1) is implied by (2.2), we only have to prove that $(M_n)^{\mathscr{F}} \to 0$ in probability in $l_\infty(\mathscr{F})$ and that $(\mathscr{F}, d_2)$ is totally bounded. But, Tsai (2000) proves that under the square integrability of τ under v (obvious by (2.2)), the compact LIL is implied by (2.3) and that the compact LIL is equivalent to the needed convergence in probability. Note also that the compact LIL implies that $(\mathscr{F}, d_2)$ is totally bounded, so ends our proof. $\quad \square$

We can extend Corollaries 2 and 3 to unbounded classes of functions $\mathscr{F} = \{f : |f| \leqslant F\}$ centred with finite variance, where F is a nonnegative function on E (called envelope of $\mathscr{F}$).

Corollary 4. *Suppose that* (H1) *and* (H2) *hold and that* $(\mathcal{F}, d_2)$ *is totally bounded. Assume either*

(a) $b_n = n^{\alpha+1/2}$ *with* $0 < \alpha \leqslant 1/2$, *and*

$$\exists \alpha' < \alpha \text{ such that } \frac{1}{n^{\alpha'}} \sum_{k=1}^{n} F(k)\pi(k)\sqrt{m_{1,k}} \to 0 \quad \text{as } n \to \infty;$$

or

(b) b_n *general and*

$$\frac{\sqrt{n}}{b_n} \sum_{k=1}^{[\sqrt{n b_n}]} F(k)\pi(k)\sqrt{m_{1,k}} \to 0 \quad \text{as } n \to \infty, \tag{2.4}$$

for all orderings of E. Then for every probability measure μ on $(E, \mathcal{E})$, verifying (1.5), *the MDP of Theorem 1 holds.*

First remark that the particular condition of part (a) is slightly weaker than in (b). By trivial facts, one can show that, for $b_n = \sqrt{2n \log \log n}$, condition (2.4) and (2.3) are equivalent.

Proof. Once again, by Theorem 1, we only have to prove that $(M_n)^{\mathcal{F}} \to 0$ in probability in $l_\infty(\mathcal{F})$. The conditions for cases (a) and (b) are obtained through a rewriting of the proof of Tsai (2000) in our context.

3. Proof

3.1. The separable case

We first need following lemma which gives us the moderate deviation principle when $\mathcal{F}$ is finite, i.e. in the separable case.

Let f be a measurable mapping from E to $\mathbb{R}^d$, suppose moreover that $\pi(f) = 0$ and $\sigma^2(f) < \infty$.

Lemma 5. *Assume that* (H1) *is satisfied and*
(H2′) $\limsup_{n \to +\infty} n/b_n^2 \log(n \mathbb{P}_\nu(\sum_{k=0}^{\tau} \|f(X_k)\| \geqslant b_n)) = -\infty$.
Then for every probability measure μ on $(E, \mathcal{E})$ verifying

$$\limsup_{n \to +\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sum_{k=0}^{\tau} \|f(X_k)\| \geqslant b_n \right) = -\infty, \tag{3.1}$$

$\mathbb{P}_\mu(M_n(f) \in \cdot)$ *verifies a moderate deviation principle with speed b_n^2/n and good rate function J_f given by* (1.6).

Remark. By assumptions (1.3) and (1.4) on the speed b_n, for each $\varepsilon > 0$, we may choose some $l(\varepsilon) > 0$ such that $\varepsilon b_n > b_{[l(\varepsilon)n]}$. Then, using (1.4), it is not hard to conclude that (H1) implies $\forall \varepsilon > 0$

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log(n \mathbb{P}_\mu(\tau > \varepsilon b_n)) = -\infty.$$

The same extension can be made for (H2), (H2′), (3.1) and (1.5), which will be referred in the sequel, with little abuse, again as (H1), (H2), (H2′), (3.1) and (1.5).

Proof of Lemma 5. The proof relies principally on a decomposition into blocks of return to the atom, which by a regeneration argument enables us to reduce the problem to the case of i.i.d. random variables.

We divide the proof of the Lemma into 4 steps: *Step 1* is dedicated to the key decomposition of $M_n(f)$. We give an extended version of Ledoux (1992) moderate deviations of i.i.d.r.v. and we apply it to our setting in *Step 2*. The negligibilities in the decomposition are established in the third and fourth steps.

Step 1: First, introduce by induction the following successive times of return to α:

$$\tau(0) = \tau = \inf\{n \geqslant 0; \ X_n \in \alpha\},$$

$$\tau(k+1) = \inf\{n > \tau(k); \ X_n \in \alpha\}. \tag{3.2}$$

Obviously, $\{\tau(k)\}$ are stopping times w.r.t. $\{X_n\}$, and are almost surely finite. Note that $\mathbb{E}_\nu \tau = \pi(\alpha)^{-1}$.

Here is the classical decomposition of the sum $M_n(f)$; which is again crucial here,

$$M_n(f) = \frac{1}{b_n} \sum_{i=0}^{n-1} f(X_i)$$

$$= M_{\tau \wedge (n-1)}(f) + \frac{1}{b_n} \sum_{k=1}^{i(n)-1} \xi_k(f) + \frac{1}{b_n} \sum_{l(n)+1 \leqslant j \leqslant n-1} f(X_j) \quad \text{p.s.,} \tag{3.3}$$

where the random $\xi_k(f)$ are defined by

$$\xi_k(f) = \sum_{j=\tau(k-1)+1}^{\tau(k)} f(X_j), \tag{3.4}$$

and

$$i(n) = \sum_{k=0}^{n-1} I_\alpha(X_k),$$

and $l(n) = \tau((i(n) - 1) \vee 0)$. Note that, by Nummelin (1984), $\{\xi_k\}(f)$ is a sequence of independent random variables with common law $\mathscr{L}_{\mathbb{P}_\nu}(\sum_{j=0}^{\tau} f(X_j))$.

Let us introduce for all n, $e(n) = [\pi(\alpha)n]$, (3.3) becomes

$$M_n(f) = \frac{1}{b_n} \sum_{k=1}^{e(n)} \xi_k(f) + M_{\tau \wedge (n-1)}(f) + \frac{1}{b_n} \left(\sum_{k=1}^{i(n)-1} \xi_k(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right)$$

$$+ \frac{1}{b_n} \sum_{l(n)+1 \leqslant j \leqslant n-1} f(X_j). \tag{3.5}$$

We control now each term of this decomposition, showing that only the first term contributes to the moderate deviations. It is the decomposition Nummelin used to establish the Central Limit Theorem (Nummelin, 1978, Theorem 7.6).

Step 2: We deal here with the moderate deviations of the first term of (3.5). First note the following.

Lemma 6. *Let* (ω_k) *be a centred and square integrable i.i.d. sequence of* $\mathbb{R}^d$ *such that* $\exists M, \forall u \in \mathbb{R}$

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log(n \mathbb{P}(\|\omega_0\| > u b_n)) \leqslant -\frac{u^2}{M}, \tag{3.6}$$

then, if $a(n)$ *is some positive increasing sequence such that* $a(n)/n \to a < \infty$, $1/b_n \sum_{k=1}^{a(n)} \omega_k$ *satisfies a moderate deviation principle with rate function* $a^{-1}I$ *where*

$$I(x) = \sup_{y \in \mathbb{R}^d} \left\{ \langle x, y \rangle - \frac{1}{2} \mathbb{E} \langle \omega_0, y \rangle^2 \right\}.$$

Ledoux (1992) proves only this result with $a(n)$ substituted by n, but his proof works in this context. So its proof is omitted.

Obviously, $(\xi_k(f))$ verifies condition (3.6) by (H2$'$) (regarded as its extension see remark after Lemma 5). Then $\mathbb{P}_\mu(1/b_n \sum_{k=1}^{e(n)} \xi_k(f) \in \cdot)$ satisfies a moderate deviation principle with speed b_n^2/n and rate function J_f given by

$$J_f(x) = \sup_{\phi \in \mathbb{R}^m} \left\{ \langle \phi, x \rangle - \frac{1}{2} \sigma^2(\langle \phi, f \rangle) \right\}.$$

Let us deal now with the other terms in the summation (3.5).

Step 3: We will prove that $\forall \varepsilon > 0$

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{k=1}^{\tau \wedge (n-1)} f(X_k) \right\| \geqslant \varepsilon b_n \right) = -\infty, \tag{3.7}$$

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{l(n)+1 \leqslant j \leqslant n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) = -\infty, \tag{3.8}$$

In fact,

$$\left\| \sum_{k=0}^{\tau \wedge (n-1)} f(X_k) \right\| \leqslant \sum_{j=0}^{\tau \wedge (n-1)} \left\| f(X_j) \right\|$$

and then (3.7) follows exactly from condition (3.1):

$$\mathbb{P}_\mu \left(\left\| \sum_{l(n)+1 \leqslant j \leqslant n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) \leqslant \mathbb{P}_\mu \left(\sum_{l(n)+1 \leqslant j \leqslant n-1} \| f(X_j) \| \geqslant \varepsilon b_n \right)$$

$$\leqslant \mathbb{P}_\mu \left(\sum_{\tau(i(n)-1)+1 \leqslant j \leqslant \tau(i(n))} \| f(X_j) \| \geqslant \varepsilon b_n \right)$$

$$\leqslant \mathbb{P}_\mu \left(\max_{0 \leqslant k \leqslant n-1} \sum_{j=\tau(k)+1}^{\tau(k+1)} \| f(X_j) \| \geqslant \varepsilon b_n \right).$$

By Nummelin (1984), $\{\sum_{j=\tau(k)+1}^{\tau(k+1)} \|f(X_j)\|\}$ are i.i.d. random variables under $\mathbb{P}_\mu$ with common law $\mathscr{L}_{\mathbb{P}_v}(\sum_{k=0}^{\tau} \|f(X_k)\|)$, so we get

$$\mathbb{P}_\mu \left(\left\| \sum_{l(n)+1 \leqslant j \leqslant n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) \leqslant n \mathbb{P}_v \left(\sum_{k=0}^{\tau} \|f(X_k)\| \geqslant \varepsilon b_n \right),$$

and (3.8) is a straightforward consequence of condition (H2′).

Step 4: We shall prove here

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right\| \geqslant \varepsilon b_n \right) = -\infty. \tag{3.9}$$

This limit needs more effort than the previous negligibilities: let $0 < \delta < \pi(\alpha)$ be fixed but arbitrary, and n be sufficiently large in order that $e(n) \geqslant \delta n$. We have by stationarity

$$\mathbb{P}_\mu \left(\left\| \sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right\| \geqslant \varepsilon b_n \right)$$

$$= \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right\| \geqslant \varepsilon b_n; |i(n) - 1 - e(n)| > \delta n \right)$$

$$+ \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right\| \geqslant \varepsilon b_n; |i(n) - 1 - e(n)| \leqslant \delta n \right)$$

$$\leqslant \mathbb{P}_\mu \left(\max_{e(n)-[\delta n] \leqslant k \leqslant e(n)+[\delta n]} \left\| \sum_{i=e(n)-[\delta n]}^{k} \xi_i(f) \right\| \geqslant \frac{\varepsilon}{2} b_n \right)$$

$$+ \mathbb{P}_\mu(|i(n) - 1 - e(n)| > \delta)$$

$$\leqslant \mathbb{P}_\mu \left(\max_{1 \leqslant k \leqslant 2[\delta n]} \left\| \sum_{i=1}^{k} \xi_i(f) \right\| \geqslant \frac{\varepsilon}{2} b_n \right)$$

$$+ \mathbb{P}_\mu(i(n) - 1 - e(n) > \delta n)$$

$$+ \mathbb{P}_\mu(i(n) < e(n) - \delta n + 1). \tag{3.10}$$

Let us begin with the last two terms of the right side of this last inequality.

$$\mathbb{P}_\mu(i(n) - 1 - e(n) > n\delta) \leqslant \mathbb{P}_\mu(\tau(e(n) + [\delta n]) \leqslant n - 1)$$

$$\leqslant \mathbb{P}_\mu(\tau(e(n) + [\delta n]) - \tau(0) \leqslant n - 1)$$

$$\leqslant \mathbb{P}_\mu \left(\sum_{k=1}^{e(n)+[\delta n]} (\tau(k) - \tau(k-1)) \leqslant n - 1 \right)$$

$$\leqslant \mathbb{P}_\mu \left(\frac{1}{b_{k(n)}} \sum_{k=1}^{k(n)} \left(\tau(k) - \tau(k-1) - \frac{1}{\pi(\alpha)} \right) \leqslant \frac{n - 1 - k(n)\pi(\alpha)^{-1}}{b_{k(n)}} \right)$$

with $k(n) = e(n) + [\delta n]$. We have

$$\frac{n-1-k(n)\pi(\alpha)^{-1}}{b_{k(n)}} = \frac{k(n)}{b_{k(n)}} \left(\frac{n-1}{k(n)} - \frac{1}{\pi(\alpha)} \right).$$

Note now that for sufficiently large n, $(n-1)/k(n) \simeq (\pi(\alpha)+\delta)^{-1}$, and $k(n)/b_{k(n)} \to \infty$. By Ledoux (1992), condition (H1) implies the upper bound of the moderate deviations for the i.i.d. sequence $\{\tau(k) - \tau(k-1) - \pi(\alpha)^{-1}\}$ with rate function I_1 such that $I_1(x) \to \infty$ when $|x| \to \infty$. Therefore, we have $\forall L > 0$,

$$\limsup_{n \to \infty} \frac{k(n)}{b_{k(n)}^2} \log \mathbb{P}_\mu(i(n) - 1 - e(n) > \delta) \leqslant - \inf_{t \leqslant -L} I_1(t).$$

Letting L tend to infinity and noting that $k(n)/n \to \pi(\alpha) + \delta$,

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu(i(n) \geqslant e(n) + [\delta n] - 1) = -\infty.$$

Using the same argument, we obtain also

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu(i(n) \leqslant e(n) - [\delta n] - 1) = -\infty.$$

We have the control of the last two terms of (3.9).

Because $1/b_n \sum_{k=1}^n \xi_k(f) \to 0$ in probability (by CLT for example), we have for sufficiently large n,

$$\max_{k \leqslant 2\delta n} \mathbb{P}_\mu \left(\left\| \sum_{j=k+1}^n \xi_j(f) \right\| \geqslant \frac{\varepsilon}{6} b_n \right) \leqslant \frac{1}{2}.$$

Then, by the Ottavianii's inequality for independent random variables, we get

$$\mathbb{P}_\mu \left(\max_{1 \leqslant k \leqslant 2[\delta n]} \left\| \sum_{j=e(n)-\delta n}^k \xi_j(f) \right\| \geqslant \frac{\varepsilon}{2} b_n \right) \leqslant 2 \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{2[\delta n]} \xi_j(f) \right\| \geqslant \frac{\varepsilon}{6} b_n \right)$$

$$+ 2 \mathbb{P}_\mu \left(\max_{k \leqslant 2\delta n} \| \xi_k(f) \| \geqslant \frac{\varepsilon}{6} b_n \right).$$

Obviously by the same approach as in (3.8), condition (H2′) implies

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\max_{k \leqslant 2\delta n} \| \xi_k(f) \| \geqslant \frac{\varepsilon}{6} b_n \right) = -\infty.$$

Taking $F = \{x; \|x\| \geqslant \varepsilon/6\}$, we have by the results of *Step 2*

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{2[\delta n]} \xi_j(f) \right\| \geqslant \frac{\varepsilon}{6} b_n \right) \leqslant - \pi(\alpha)(2\delta)^{-1} \inf_{\|x\| \geqslant \varepsilon/6} J_f(x).$$

Combining these last results, we obtain

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\left\| \sum_{j=1}^{i(n)-1} \xi_j(f) - \sum_{k=1}^{e(n)} \xi_k(f) \right\| \geqslant \varepsilon b_n \right) \leqslant - \pi(\alpha)(2\delta)^{-1} \inf_{\|x\| \geqslant \varepsilon/6} J_f(x).$$

As $\sigma^2(\langle x, \xi \rangle)$ is differentiable on ξ and $\partial_\xi \sigma^2(\langle x, \xi \rangle)|_{\xi=0} = 0$, $J_f(x) = 0 \Leftrightarrow x = 0$. Thus, by the inf-compactness of J_f, we get

$$\inf_{\|x\| \geqslant \varepsilon/6} J_f(x) > 0.$$

As δ is arbitrary, letting $\delta \to 0^+$, we have then the negligibility (3.9).

Using estimates (3.7)–(3.9) and the moderate deviations of *Step 2*, we get the result by Dembo and Zeitouni (1993, Theorem 4.2.13).

3.2. Proof of Theorem 1

Theorem 1 is established using the line of the proof of Wu (1994) for i.i.d. case. In fact, we reduce our proof to the use of Lemma 5 and to an exponential asymptotic equicontinuity with respect to the pseudometric d_2 associated with $\mathscr{F}$.

Under our hypothesis, by Lemma 5, we have the finite dimensional moderate deviation principle, i.e. for each $f_1, \ldots, f_m \in \mathscr{F}$, $M_n((f_1, \ldots, f_m))$ satisfies the moderate deviation principle with speed b_n^2/n and the good rate function $J_{(f_1, \ldots, f_m)}$. We introduce the following notation $\forall \eta > 0$:

$$\mathscr{F}_\eta = \{f - g; f, g \in \mathscr{F} \text{ and } d_2(f, g) \leqslant \eta\}.$$

We have obviously that $(\mathscr{F}_\eta, d_2)$ is totally bounded. Moreover $(M_n)^{\mathscr{F}_\eta} \to 0$ in probability in $l_\infty(\mathscr{F}_\eta)$ by our assumption. By Wu (1994), for the MDP we have only to verify the following condition: $\forall \varepsilon > 0$,

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu(\|M_n\|_{\mathscr{F}_\eta} \geqslant \varepsilon) = -\infty. \tag{3.11}$$

Or equivalently, $\forall \varepsilon > 0$

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{j=0}^{n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right) = -\infty. \tag{3.12}$$

We use the same decomposition as in the proof of the preceding theorem:

$$\mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{j=0}^{n-1} f(X_j) \right\| \geqslant \varepsilon b_n \right)$$

$$= \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{j=0}^{\tau \wedge (n-1)} f(X_j) + \sum_{i=1}^{i(n)-1} \xi_i(f) + \sum_{l(n)+1 \leqslant i \leqslant n-1} f(X_i) \right\| \geqslant \varepsilon b_n \right)$$

$$\leqslant \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{j=0}^{\tau \wedge (n-1)} f(X_j) \right\| \geqslant \frac{\varepsilon}{3} b_n \right) + \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{i=1}^{i(n)-1} \xi_i(f) \right\| \geqslant \frac{\varepsilon}{3} b_n \right)$$

$$+ \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{l(n)+1 \leqslant i \leqslant n-1} f(X_i) \right\| \geqslant \frac{\varepsilon}{3} b_n \right).$$

We then have to prove the negligibility of all the terms in the right side of this inequality.

Using conditions (1.5) and (H2), we get as for (3.7) and (3.8)

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{j=0}^{\tau \wedge (n-1)} f(X_j) \right\| \geq \frac{\varepsilon}{3} b_n \right) = -\infty,$$

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{l(n)+1 \leq m \leq n-1} f(X_m) \right\| \geq \frac{\varepsilon}{3} b_n \right) = -\infty. \tag{3.13}$$

For the middle term, first note

$$\mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{i(n)-1} \xi_k(f) \right\| \geq \frac{\varepsilon}{3} b_n \right) \leq \mathbb{P}_\mu \left(\max_{1 \leq i \leq n} \sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{i} \xi_k(f) \right\| \geq \frac{\varepsilon}{3} b_n \right).$$

We obviously have that $(1/b_n \sum_{k=1}^n \xi_k(f))^{\mathscr{F}_\eta} \to 0$ in probability in $l_\infty(\mathscr{F}_\eta)$ by the assumption that $(M_n)^{\mathscr{F}} \to 0$ in probability. Therefore, for sufficiently large n, we have

$$\max_{k \leq n} \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{n} \xi_k(f) \right\| \geq \frac{\varepsilon}{9} b_n \right) \leq \frac{1}{2},$$

and we use again the Ottavianii's inequality in Banach space for independent random variables, then for sufficiently large n,

$$\mathbb{P}_\mu \left(\max_{1 \leq i \leq n} \sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{i} \xi_k(f) \right\| \geq \frac{\varepsilon}{3} b_n \right) \leq 2 \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{n} \xi_k(f) \right\| \geq \frac{\varepsilon b_n}{9} \right)$$

$$+ 2 \mathbb{P}_\mu \left(\max_{k \leq n} \sup_{f \in \mathscr{F}_\eta} \| \xi_k(f) \| \geq \frac{\varepsilon b_n}{9} \right).$$

The negligibility of the last term is done as in the proof of Lemma 5. For the first term, we then use Lemma 3 of Wu (1994), an extension of Ledoux moderate deviations in the nonseparable case, we get identifying $B = l_\infty(\mathscr{F}_\eta)$ (in the notations of Wu, 1994, Lemma 3)

$$\limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{n} \xi_k(f) \right\| \geq \frac{\varepsilon b_n}{9} \right) \leq -\frac{\varepsilon^2}{C_0 \sigma^2},$$

where C_0 is some universal positive constant and $\sigma^2 = \sup_{f \in \mathscr{F}_\eta} \mathbb{E}(\xi_1(f))^2$. Remark that $\sigma^2 \leq \eta^2$ and consequently

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \frac{n}{b_n^2} \log \mathbb{P}_\mu \left(\sup_{f \in \mathscr{F}_\eta} \left\| \sum_{k=1}^{n} \xi_k(f) \right\| \geq \frac{\varepsilon b_n}{9} \right) = -\infty. \tag{3.14}$$

Combining (3.3), (3.4), and preceding inequalities, we get (3.2) and then our theorem.

Acknowledgements

Appendix

This section is devoted to the proof of the following lemma:

Lemma 7. *Under hypothesis* (H2), *the invariant measure π satisfies condition* (1.5).

Proof. In fact, we will prove the following strongest assertion:
(i) $\limsup_{n\to\infty} n/b_n^2 \log \mathbb{P}_v(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n) = -\infty$ is equivalent to
(ii) $\limsup_{n\to\infty} n/b_n^2 \log \mathbb{P}_\pi(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n) = -\infty.$

This assertion being proved, the conclusion of the lemma follows.
(ii) $\Rightarrow$ (i): It follows simply from

$$
\pi(\cdot) = \int_E \pi(\mathrm{d}x) P(x, \cdot)
$$

$$
\geqslant \int_\alpha \pi(\mathrm{d}x) P(x, \cdot)
$$

$$
\geqslant \int_\alpha \pi(\mathrm{d}x) v(\cdot)
$$

$$
= \pi(\alpha) v(\cdot).
$$

(i) $\Rightarrow$ (ii): By Nummelin (1984), letting $g(x) = \mathbb{P}_x(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n)$, we have

$$
\mathbb{P}_\pi \left(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n \right) = \int_E \pi(\mathrm{d}x) \mathbb{P}_x \left(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n \right)
$$

$$
= \int_E g(x) \pi(\mathrm{d}x)
$$

$$
= \pi(\alpha) \mathbb{E}_v \left(\sum_{k=1}^{\tau} g(X_k) \right).
$$

Let $\sigma_k = \{n \geqslant k; X_n \in \alpha\}$, using the strong Markov property, we get

$$
\mathbb{P}_\pi \left(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n \right)
$$

$$
= \pi(\alpha) \mathbb{E}_v \left(\sum_{k=1}^{\tau} \mathbb{P}_v \left(\sum_{j=k}^{\sigma_k} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n / \mathscr{F}_k \right) \right)
$$

$$
= \pi(\alpha) \mathbb{E}_v \left(\sum_{k=1}^{\infty} 1_{\{\tau \geqslant k\}} \mathbb{P}_v \left(\sum_{j=k}^{\sigma_k} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n / \mathscr{F}_k \right) \right).
$$

Since $\{\tau \geqslant k\}$ is $\mathscr{F}_k$ measurable, and on $\{\tau \geqslant k\}, \sigma_k = \tau$, we have

$$\mathbb{P}_\pi \left(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n \right) = \pi(\alpha)\mathbb{E}_v \left(\sum_{k=1}^{\infty} 1_{\{\tau \geqslant k\}} 1_{\{\sum_{j=k}^{\sigma_k} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n\}} \right)$$

$$\leqslant \pi(\alpha)\mathbb{E}_v \left(\sum_{k=1}^{\tau} 1_{\{\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n\}} \right)$$

$$= \pi(\alpha)\mathbb{E}_v(\tau 1_{\{\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n\}})$$

$$\leqslant \pi(\alpha)\sqrt{\mathbb{E}_v(\tau^2)}\sqrt{\mathbb{P}_v \left(\sum_{j=0}^{\tau} \|f(X_j)\|_{\mathscr{F}} \geqslant b_n \right)},$$

where the last step is obtained by Cauchy–Schwartz inequality. We now easily derive (ii) from (i).

4. For further reading

The following references are also of interest to the reader: Athreya and Ney, 1978; Chen, 1991; de Acosta, 1990; de Acosta, 1997; Gao, 1994; Wu, 1995.

References

Athreya, K.B., Ney, P., 1978. A new approach to the limit theory of recurrent Markov chains. Trans. Amer. Math. Soc. 245, 493–501.

Chen, X., 1991. Probabilities of moderate deviations for independent random vectors in Banach space. Chinese J. Appl. Probab. Statist. 7, 24–32.

Chen, X., 1997. Moderate deviation for m-dependent random variables with Banach space value. Statist. Probab. Lett. 35, 123–134.

Chen, X., 1999. Limit theorems for functionals of ergodic Markov chains with Banach space value. Memoirs AMS, pp. 139.

de Acosta, A., 1988a. Large deviations for vector-valued functionals of a Markov chain: lower bounds. Ann. Probab. 16, 925–960.

de Acosta, A., 1988b. Moderate deviation for empirical measures of Markov chains: lower bound. Ann. Probab. 25, 259–284.

de Acosta, A., 1990. Large deviations for empirical measures of Markov chain. J. Theoret. Probab. 3, 395–431.

de Acosta, A., 1997. Exponential tightness and projective systems in large deviation theory. Festschrift for Lucien Le Cam. Springer, Berlin, pp. 143–156.

de Acosta, A., Chen, X., 1998. Moderate deviation for empirical measure of Markov chains: Upper bound. J. Theoret. Probab. 4.11, 75–110.

Dembo, A., Zeitouni, O., 1993. Large Deviations Techniques and Their Applications. Jones and Bartlett, Boston, MA.

Deuschel, J.D., Stroock, D.W., 1989. Large Deviations. Academic Press, Boston.

Djellout, H., 2000. Moderate deviations for martingale differences and applications to ϕ-mixing sequences. Submitted for publication.

Gao, F.Q., 1994. Uniform moderate deviations for Markov processes. Research Announcements, Advances in Mathematics, China.

Guillin, A., 2000. Uniform moderate deviations of functional empirical processes of Markov chains. Probab. Math. Stat. 20 (2), 237–260.

Guillin, A., 2001. Moderate deviations of inhomogeneous functional of Markov processes and application to averaging. Stochast. Proc. Appl. 92 (2), 287–313.

Ledoux, M., 1992. Sur les grandes déviations modérées des sommes de variables aléatoires vectorielles indépendantes de même loi. Ann. Inst. H. Poincaré 35, 123–134.

Levental, S., 1990. Uniform CLT for Markov chains with a countable state space. Stochast. Proc. Appl. 34, 245–253.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer-Verlag, Berlin.

Mogulskii, A.A., 1984. On moderately large deviation from the invariant measure. In: Borokov, A.A. (Ed.), Advances in Probability Theory: Limit Theorems and Related Problems. Optimization Software, New York.

Nummelin, E., 1978. A splitting technique for Harris recurrent chains. Z. Wahrs. nerw Gebiete 43, 309–318.

Nummelin, E., 1984. General Irreducible Markov Chains and Non-negative Operators. Cambridge University Press, Cambridge, England.

Nummelin, E., Tuominen, P., 1982. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. Stochast. Proc. Appl. 12, 187–202.

Nummelin, E., Tweedie, R.L., 1978. Geometric ergodicity and R-positivity for general Markov chains. Ann. Probab. 6, 404–420.

Tsai, T., 1997. Uniform CLT for Markov chains with a countable state space. Taiwanese J. Math. 1 (4), 481–498.

Tsai, T., 2000. Empirical law of the iterated logarithm for Markov chains with a countable state space. Stochast. Proc. Appl. 89, 175–191.

Wu, L.M., 1993. Habilitation à diriger des recherches. Université PARIS 6.

Wu, L.M., 1994. Large deviations, moderate deviations and LIL for empirical processes. Ann. Probab. 22, 17–27.

Wu, L.M., 1995. Moderate deviations of dependent random variables related to CLT. Ann. Probab. 23, 420–445.