Stochastic Processes and their Applications 41 (1992) 191-202
North-Holland

On large deviations for uniformly strong mixing sequences

Wlodzimierz Bryc
Department of Mathematics, University of Cincinnati, Cincinnati, OH, USA

Received 13 April 1989
Revised 15 August 1990

We prove the large deviation principle for the arithmetic means of a uniform strong mixing stationary sequence which has either fast enough ϕ-mixing rate or is ψ-mixing.

AMS 1980 Subject Classification: 60F10.

large deviations * uniform strong mixing

1. Results

There is an extensive literature on the large deviation principle under various dependence structures. However, in comparison to other limit theorems, few papers deal with strong mixing dependence conditions. In this note we prove the large deviation principle for ϕ-mixing stationary random sequences with fast enough convergence rate and for ψ-mixing stationary random sequences. Our ϕ-mixing results were motivated by Schonmann [17], where a weaker conclusion is proved under a much weaker mixing assumption. The large deviation principle under ψ-mixing was motivated by Orey and Pelikan [15] and Bryc [7]. Chiyonobu and Kusuoka [8] consider large deviations under a mixing condition, different from those ususally considered in the strong mixing approach. Yakimavichyus [19] (cf. also references there) states large deviation theorems under the assumption of uniform strong mixing, but considers one dimensional case only and his conclusions have a different form which makes it difficult to compare his results with ours.

Let $\{X_n\}_{n\in\mathbb{Z}}$ be a stationary sequence. Define σ-fields $\mathcal{F}_{a,b}=\sigma(X_k : a\leq k\leq b)$ and let

$$\phi(n)=\sup\{|P(B\,|\,A)-P(B)|: A\in\mathcal{F}_{-\infty,-n+1},\, B\in\mathcal{F}_{1,\infty},\, P(A)>0\}. \tag{1.1}$$

We shall say that $\{X_n\}_{n\in\mathbb{Z}}$ is ϕ-mixing, if $\phi(n)\to 0$ as $n\to\infty$.

In the Markov case the ϕ-mixing condition is related to the Döeblin condition, see Rosenblatt [16, p. 209]. It is known to hold (see Doob [13, p. 197]), if the following conditions on transition probabilities $P_t(x, A)$ are satisfied:

Correspondence to: Prof. W. Bryc, Department of Mathematics, University of Cincinnati, OH 45221, USA.

There exist $\delta>0$, a probability measure ν, an integer t and a measurable set C such that

(i) $\nu(C)>0$;

(ii) $p_t(x,y)\geqslant\delta$ for each $x\in\mathbb{E}$, $y\in C$,

where $p_t(x,\cdot)=\mathrm{d}P_t(x,\cdot)/\mathrm{d}\nu(\cdot)$ is the ν-density of the absolutely continuous component of $P_t(x,\cdot)$. (Compare also Donsker and Varadhan [11, Assumption A, p. 280].)

We shall need the following hyper-geometric rate of convergence:

$$\mathrm{e}^{Kn}\phi(n)\to 0 \quad \text{as } n\to\infty \quad \text{for each } K\geqslant 0. \tag{1.2}$$

For examples of non-trivial stationary sequences satisfying (1.2), see Bradley [4, Theorem 1]. Note that (i) and (ii) above guarantee geometric ϕ-mixing rate rather than (1.2). We shall also use stronger coefficients of dependence. Define

$$\psi_+(n)=\sup\{P(A\cap B)/(P(A)P(B)):\\ A\in\mathcal{F}_{-\infty,-n+1},\ B\in\mathcal{F}_{1,\infty},\ P(A)P(B)>0\}, \tag{1.3}$$

$$\psi_-(n)=\inf\{P(A\cap B)/(P(A)P(B)):\\ A\in\mathcal{F}_{-\infty,-n+1},\ B\in\mathcal{F}_{1,\infty},\ P(A)P(B)>0\}, \tag{1.4}$$

$$\psi(n)=\psi_+(n)/\psi_-(n). \tag{1.5}$$

It is easily seen that $\psi_-(n)\nearrow$, $\psi_+(n)\searrow$, so $\psi(n)$ is a non-increasing function of n and in particular $\lim_{n\to\infty}\psi(n)$ exists. It is actually known, see Bradley [5], that this limit can take for ergodic-mixing sequences one of the values ∞ or 1 only.

In the Markov case the condition $\psi(1)<\infty$ is equivalent to the condition used e.g. in Stroock [18, Assumption (6.1)]. The condition $\psi(1)<\infty$ can also be verified for Gibbs fields on $\mathbb{Z}^1$ with binary interactions $\Phi_2(k,x,y)$ such that $\sum_{k=1}^{\infty}k\sup_{x,y}|\Phi_2(k,x,y)|<\infty$; this can be seen from Bowen [2, proof of Proposition 1.14].

Theorem 1. *Suppose $\{X_n\}_{n\in\mathbb{Z}}$ is a ϕ-mixing stationary $\mathbb{R}^d$-valued sequence, such that $\|X_1\|\leqslant C<\infty$ and (1.2) holds. Then $Z_n=(X_1+\cdots+X_n)/n$, $n\geqslant 1$, satisfies the large deviation principle, i.e. there is a convex lower semicontinuous rate function $I:\mathbb{R}^d\to[0,\infty]$ with compact level sets $I^{-1}[0,a]$, $a\geqslant 0$, and such that*

$$\limsup_{n\to\infty} n^{-1}\log P(Z_n\in A)\leqslant -\inf_{x\in A} I(x) \tag{1.6}$$

for each closed set $A\subset\mathbb{R}^d$;

$$\liminf_{n\to\infty} n^{-1}\log P(Z_n\in A)\geqslant -\inf_{x\in A} I(x) \tag{1.7}$$

for each open set $A\subset\mathbb{R}^d$.

Moreover, limit

$$\lim_{n\to\infty} n^{-1}\log E\{\exp(n\lambda(Z_n))\}=\mathbb{L}(\lambda) \tag{1.8}$$

exists for each $\lambda \in \mathbb{R}^d$ *and the rate function is given by*

$$I(x) = \sup\{\lambda(x) - \mathbb{L}(\lambda) \colon \lambda \in \mathbb{R}^d\} \tag{1.9}$$

(*here* $\lambda(x) = \sum \lambda_i x_i$). $\square$

The following theorem uses a stronger measure of dependence rather than a mixing rate assumption. Under (1.13) below Theorem 2 extends to dependent case Donsker and Varadhan [12, Theorem 5.3].

Theorem 2. *Let* $(\mathbb{V}, \|\cdot\|)$ *be a separable Banach space. Suppose* $\{X_k\}_{k\in\mathbb{Z}}$ *is a stationary sequence of* $\mathbb{V}$*-valued random variables such that*

$$\lim_{n\to\infty} \psi_-(n) > 0. \tag{1.10}$$

If in addition one of the following conditions holds:

$$X_1 \text{ is compact valued;} \tag{1.11}$$

$$\psi_+(n) < \infty \quad \text{for some } n \geq 1 \quad \text{and} \quad \|X_1\| \text{ is bounded;} \tag{1.12}$$

$$\psi(1) < \infty \quad \text{and} \quad E\{\exp(\theta\|X_1\|)\} < \infty \quad \text{for each } \theta \in \mathbb{R}; \tag{1.13}$$

then $\{(X_1 + \cdots + X_n)/n\}_{n\geq 1}$ *satisfies the large deviation principle, i.e.* (1.6) *and* (1.7) *hold with a rate function* $I(\cdot)$, *which is given by* (1.9) *with* $\lambda \in \mathbb{V}^*$ *and has compact level sets* $I^{-1}[0, a]$, $a \geq 0$.

The following result deals with the empirical measures rather than the sample means. In the statement $\mathscr{P}(\mathbb{E})$ denotes a Polish space of all (countably additive) probability measures on ($\mathbb{E}$, Borel sets) with the weak convergence topology; $C_b(\mathbb{E})$ denotes the bounded continuous functions; if $p \in \mathscr{P}(\mathbb{E})$ and $F \in C_b(\mathbb{E})$, $p(F)$ stands for $\int F(x)\,dp(x)$.

Theorem 3. *Suppose* $(\mathbb{E}, d)$ *is a compact space and* $\{X_n\}_{n\in\mathbb{Z}}$ *is a* ϕ*-mixing stationary* $\mathbb{E}$*-valued sequence such that* (1.2) *holds. Then the sequence of empirical measures* $\mu_n = (\delta_{X_1} + \cdots + \delta_{X_n})/n$ *satisfies the large deviation principle, i.e. there is a convex lower semicontinuous function* $I \colon \mathscr{P}(\mathbb{E}) \to [0, \infty]$ *with compact level sets* $I^{-1}[0, a]$, $a \geq 0$, *and such that*

$$\limsup_{n\to\infty} n^{-1} \log P(\mu_n \in A) \leq -\inf_{p\in A} I(p)$$

for each closed set $A \subset \mathscr{P}(\mathbb{E})$;

$$\liminf_{n\to\infty} n^{-1} \log P(\mu_n \in A) \geq -\inf_{p\in A} I(p)$$

for each open set $A \subset \mathscr{P}(\mathbb{E})$.

Moreover, limit

$$\lim_{n\to\infty} n^{-1} \log E\{\exp(n\mu_n(F))\} = \mathbb{L}(F)$$

exists for all $F \in C_b(\mathbb{E})$ *and*

$$I(p) = \sup\{p(F) - \mathbb{L}(F) \colon F \in C_b(\mathbb{E})\}. \quad \square$$

Remarks. 1. Similarly to Bryc [7, Theorem T.2.2], our proof of Theorem 3 extends to the empirical process level large deviation principle. However considering empirical measures only has an advantage of having both Theorem 1 and Theorem 3 proved simultaneously, see Theorem 4 below.

2. The large deviation principle for an instantaneous function of a Markov chain under minimal hypotheses and with a rate function identified as the entropy can be found in de Acosta [10]. Berbee and Bradley [1] show that there are stationary processes, which cannot be represented as an instantaneous function of an irreducible stationary Markov chain and such that (1.12) holds and the ϕ-mixing rate is geometric. However the question whether either of the conditions (1.2) or (1.10) implies already the Markov representation seems to be left open.

2. Proof of ϕ-mixing results

Recall that a topology τ_1 is called stronger (i.e. non-weaker) than a topology τ_2, if the identity mapping $(\mathbb{X}, \tau_1) \to (\mathbb{X}, \tau_2)$ is continuous. Theorems 1 and 3 are special cases of the following result.

Theorem 4. *Suppose $\mathbb{V}$ is a Hausdorff locally convex topological vector space, $\mathbb{X} \subset \mathbb{V}$ is convex, $(\mathbb{X}, d)$ is metric and compact in the relative topology of $\mathbb{V}$. Let $\|\cdot\| : \mathbb{V} \to \mathbb{R}$ be a measurable norm generating a stronger topology on $\mathbb{V}$. Let $\{X_n\}_{n \in \mathbb{Z}}$ be a stationary ϕ-mixing $\mathbb{X}$-valued random sequence such that $\|X_1\| \leqslant C < \infty$ and (1.2) holds. Then $Z_n = (X_1 + \cdots + X_n)/n$, $n \geqslant 1$, satisfies the large deviation principle, i.e. there is a convex lower semicontinuous function $I : \mathbb{X} \to [0, \infty]$ with compact level sets and such that (1.6) and (1.7) hold (in the $(\mathbb{X}, d)$-topology). Moreover, limit (1.8) exists for all $\lambda \in \mathbb{V}^*$ and formula (1.9) with the supremum taken over all $\lambda \in \mathbb{V}^*$ identifies the rate function.*

Indeed, Theorem 1 is easily seen to be a special case of Theorem 4, with $\mathbb{V} = \mathbb{R}^d$ and $\mathbb{X}$ being a large enough cube. Theorem 3 follows from Theorem 4 applied to the compact and convex subset $\mathbb{X} = \mathscr{P}(\mathbb{E})$ of the locally convex Hausdorff topological vector space $\mathbb{V}$ of all signed measures on $\mathbb{E}$. $\mathbb{V}$ is considered with the topology of weak convergence and linear functionals $\lambda \in \mathbb{V}^*$ are identified with bounded continuous functions $F = C_b(\mathbb{E})$ by $\lambda(\mu) = \mu(F) = \int F(x)\,d\mu(x)$, see e.g. Dunford and Schwartz [14, V.3.9]. The norm $\|\cdot\|$ is the variation of a measure.

Proof of Theorem 4. Since $\mathbb{X}$ is a compact set and $\mathbb{V}^*$ separates points of $\mathbb{V}$, the theorem follows from Bryc [6, C.2.1] and the following two claims, see also Bryc [7, Section 3] for a similar argument.

Let

$$\mathscr{G} = \left\{ \gamma : \gamma(x) = \min_{1 \leqslant j \leqslant d} \{\lambda_j(x) + c_j\},\ d \geqslant 1,\ c_j \in \mathbb{R},\ \lambda_j \in \mathbb{V}^* \right\}.$$

Claim 1. Under the assumptions of Theorem 4, limit

$$\lim_{n\to\infty} n^{-1} \log E\{\exp(n\gamma(Z_n))\} = \mathbb{L}(\gamma) \tag{2.1}$$

exists for all $\gamma \in \mathscr{G}$.

Claim 2. Under the assumptions of Theorem 4, $J : \mathbb{V} \to [0, \infty]$, defined by

$$J(x) = \sup\{\gamma(x) - \mathbb{L}(\gamma) \colon \gamma \in \mathscr{G}\} \tag{2.2}$$

is a convex function.

Proof of Claim 1. Let $\gamma(x) = \min_{1\le j\le d} \{\lambda_j(x) + c_j\}$, where $d \geq 1$, $c_j \in \mathbb{R}$, $\lambda_j \in \mathbb{V}^*$ are fixed. Notice that there is $K < \infty$ such that

$$\sup_n |\gamma(Z_n)| \leq K \quad \text{a.s.} \tag{2.3}$$

Indeed, (2.3) follows from the fact that $\|X_1\| \leq C$ and the $\|\cdot\|$-continuity of linear functionals $\lambda \in \mathbb{V}^*$.

Define $L_n = L_n(\gamma) = \log E\{\exp(n\gamma(Z_n))\}$. We need to show that

$$\liminf_{n\to\infty} n^{-1} L_n \geq \limsup_{n\to\infty} n^{-1} L_n. \tag{2.4}$$

Moreover, since $\gamma_c(x) = \min_{1\le j\le d} \{\lambda_j(x) + c_j + \text{const}\} = \gamma + \text{const}$, we have $L_n(\gamma + \text{const}) = n\,\text{const} + L_n(\gamma) = L_n(\gamma_c)$ and hence by (2.3) without loosing generality we may assume that

$$\gamma(Z_n) \leq 0 \quad \text{for all } n \geq 1. \tag{2.5}$$

It is well known that for each non-negative bounded $\mathscr{N}$-measurable ξ,

$$|E\{\xi \mid \mathscr{M}\} - E\{\xi\}| \leq \phi(\mathscr{M}, \mathscr{N}) \|\xi\|_\infty,$$

where $\phi(\mathscr{M}, \mathscr{N})$ is defined by the right-hand side of (1.1) with the supremum taken over $A \in \mathscr{M}$, $B \in \mathscr{N}$. Therefore by (2.5), for each $N \geq 1$ we have

$$\begin{aligned} &E\{\exp(n\gamma(Z_n)) \mid \mathscr{F}_{-\infty,-N}\} \\ &\quad \geq E\{\exp(n\gamma(Z_n))\} - \phi(N+1) \quad \text{for all } n \geq 1. \end{aligned} \tag{2.6}$$

Fix N, $M \geq 1$ to be chosen later. To prove (2.4), we shall use a standard 'blocking argument'. For $n \geq 1$ write $n = k(M+N) + r$, $0 \leq r \leq M + N$ i.e. put $k = k(n) = [n/(M+N)]$, $r = n - k(M+N)$. Since $\gamma(\cdot)$ is concave and Z_n can be represented as convex combination

$$Z_n = \frac{k(M+N)}{n} Z_{k(M+N)} + \frac{r}{n}\left(\frac{1}{r} \sum_{i=k(N+M)+1}^{n} X_i\right),$$

inequalities (2.3) and $r \leq M + N$ give

$$E\{\exp(n\gamma(Z_n))\} \geq \mathrm{e}^{-K(M+N)} E\{\exp(k(M+N)\gamma(Z_{k(M+N)}))\}. \tag{2.7}$$

Using again the concavity of $\gamma(\cdot)$, the convex combination representation

$$Z_{k(M+N)}=\frac{1}{k}\frac{M}{M+N}\sum_{j=0}^{k-1}\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)+\frac{1}{k}\frac{N}{M+N}\sum_{j=0}^{k-1}\left(\frac{1}{N}\sum_{i=1}^{N}X_{j(N+M)+M+i}\right)$$

and the fact that by (2.3),

$$\left|\gamma\left(\frac{1}{N}\sum_{i=1}^{N}X_{j(N+M)+M+i}\right)\right|\leqslant K,\quad 0\leqslant j\leqslant k-1,$$

we get

$$E\{\exp(k(M+N)\gamma(Z_{k(M+N)}))\}\geqslant \mathrm{e}^{-kKN}E\left\{\exp\left(\sum_{j=0}^{k-1}N\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\right\}. \tag{2.8}$$

We need now to estimate

$$E\left\{\exp\left(\sum_{j=0}^{k-1}M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\right\}$$
$$=E\left\{\exp\left(\sum_{j=0}^{k-2}M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\times E\left\{\exp\left(M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{(k-1)(N+M)+i}\right)\right)\;\middle|\;\mathscr{F}_{-\infty,(k-1)(N+M)-N}\right\}\right\}.$$

Since $\exp(M\gamma((1/M)\sum_{i=1}^{M}X_{(k-1)(N+M)+i}))$ is $\mathscr{F}_{(k-1)(N+M)+1,\infty}$ – measurable, stationarity and (2.6) give

$$E\left\{\exp\left(\sum_{j=0}^{k-1}M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\right\}$$
$$\geqslant E\left\{\exp\left(\sum_{j=0}^{k-2}M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\right\}(E\{\exp(M\gamma(Z_M))\}-\phi(N+1)).$$

This implies recurrently

$$E\left\{\exp\left(\sum_{j=0}^{k-1}M\gamma\left(\frac{1}{M}\sum_{i=1}^{M}X_{j(N+M)+i}\right)\right)\right\}\geqslant E\{\exp(M\gamma(Z_M))\}-\phi(N))^{k}. \tag{2.9}$$

Inequalities (2.7), (2.8) and (2.9) put together give

$$n^{-1}L_n\geqslant -K(M+N)/n-KNk/n+k/n\log(\exp(L_M)-\phi(N)). \tag{2.10}$$

Passing in (2.10) to the limit as $n\to\infty$ we get

$$\liminf_{n\to\infty} n^{-1}L_n\geqslant\frac{M}{M+N}\frac{1}{M}L_M-\frac{KN}{M+N}+(M+N)^{-1}\log(1-\phi(N)\exp(-L_M)).$$

To end the proof of (2.4) it remains to pick N, $M\to\infty$ such that

$$\begin{aligned}&\frac{1}{M}L_M\to\limsup_{n\to\infty} n^{-1}L_n,\\ &N/M\to 0,\\ &\phi(N)\exp(-L_M)\to 0.\end{aligned}\tag{2.11}$$

Note that such a choice of N, M is possible; (2.11) can be satisfied by (1.2), because (2.3) gives $\exp(-L_M)\leqslant e^{KM}$ for each M.

Proof of Claim 2. Since (2.2) defines a lower semicontinuous function, to verify convexity it is enough to show that $J(\frac12(x+y))\leqslant\frac12(J(x)+J(y))$ for all $x,y\in\mathbb{X}$. To this end fix x_0,y_0. By (2.2) it is enough to show that for each $\gamma\in\mathscr{G}$, there are $\gamma_1,\gamma_2\in\mathscr{G}$ such that

$$\gamma(\tfrac12(x_0+y_0))=\tfrac12\gamma_1(x_0)+\tfrac12\gamma_2(y_0),\tag{2.12}$$

$$\mathbb{L}(\gamma)\geqslant\tfrac12\mathbb{L}(\gamma_1)+\tfrac12\mathbb{L}(\gamma_2).\tag{2.13}$$

Indeed, (2.12) and (2.13) imply that $\gamma(\frac12(x_0+y_0))-\mathbb{L}(\gamma)\leqslant\frac12(\gamma_1(x_0)-\mathbb{L}(\gamma_1))+\frac12(\gamma_2(y_0)-\mathbb{L}(\gamma_2))\leqslant\frac12(J(x_0)+J(y_0))$. This proves convexity of $J(\cdot)$, as $\gamma\in\mathscr{G}$ is arbitrary. To construct γ_1, γ_2, write $\gamma(\cdot)=\min_{1\leqslant i\leqslant r}\{\lambda_i(\cdot)+c_i\}$ and put $\gamma_1(\cdot)=\min\{\lambda_i(\cdot)+c_i+d_i\}$, $\gamma_2(\cdot)=\min\{\lambda_i(\cdot)+c_i-d_i\}$, where $d_i=\frac12\lambda_i(y_0-x_0)$, $1\leqslant i\leqslant r$. Without loss of generality we may assume that $\gamma_2(Z_n)\leqslant 0$ for all $n\geqslant 1$. Indeed, as in the proof of claim 1, we may replace γ by γ_c for a suitable constant c.

Then (2.12) is trivially satisfied. To prove (2.13) observe that for all x, y we have

$$\gamma(\tfrac12(x+y))=\min_i\{\tfrac12(\lambda_i(x)+c_i+d_i)+\tfrac12(\lambda_i(y)+c_i-d_i)\}\geqslant\tfrac12\gamma_1(x)+\tfrac12\gamma_2(y).$$

Since

$$Z_{2n}=\tfrac12\left(\frac1n\sum_{i=1}^n X_i+\frac1n\sum_{i=1}^n X_{i+n}\right),$$

the last inequality implies

$$E\{\exp(2n\gamma(Z_{2n}))\}\geqslant E\left\{\exp\left(n\gamma_1\left(\frac1n\sum_{i=1}^n X_i\right)+n\gamma_2\left(\frac1n\sum_{i=1}^n X_{i+n}\right)\right)\right\}.$$

Since $\gamma_1(\cdot)$ is concave,

$$\begin{aligned}&E\left\{\exp\left(n\gamma_1\left(\frac1n\sum_{i=1}^n X_i\right)+n\gamma_2\left(\frac1n\sum_{i=1}^n {}_nX_{i+n}\right)\right)\right\}\\ &\quad\geqslant e^{-KN}E\left\{\exp\left((n-N)\gamma_1\left(\frac{1}{n-N}\sum_{i=1}^{n-N} X_i\right)+\gamma_2\left(\frac1n\sum_{i=1}^n X_{i+n}\right)\right)\right\}\\ &\quad=e^{-KN}E\left\{\exp\left((n-N)\gamma_1\left(\frac{1}{n-N}\sum_{i=1}^{n-N} X_i\right)\right)\right.\\ &\qquad\qquad\left.\times E\left\{\exp\left(n\gamma_2\left(\frac1n\sum_{i=1}^n X_{i+n}\right)\right)\,\middle|\,\mathscr{F}_{-\infty,n-N}\right\}\right\},\end{aligned}$$

where K is a constant given by (2.3) with γ replaced by γ_1. Since γ_2 satisfies (2.5), taking (2.6) into account we obtain therefore

$$
\begin{aligned}
&E\{\exp(2n\gamma(Z_{2n}))\}\\
&\geqslant \mathrm{e}^{-KN} E\left\{\exp\left((n-N)\gamma_1\left(\frac{1}{n-N}\sum_{i=1}^{n-N} X_i\right)\right)\right\}\\
&\quad\times\left(E\left\{\exp\left(n\gamma_2\left(\frac{1}{n}\sum_{i=1}^{n} X_{i+n}\right)\right)\right\}-\phi(N+1)\right)\\
&\geqslant \mathrm{e}^{-(K+C)N}\left(E\left\{\exp\left(n\gamma_2\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)\right)\right\}-\phi(N)\right)\\
&\quad\times E\left\{\exp\left(n\gamma_1\left(\frac{1}{n}\sum_{i=1}^{n} X_{i+n}\right)\right)\right\},
\end{aligned}
$$

where $C=\sup_{1\leqslant i\leqslant r}\sup_{x\in\mathrm{supp}\,X_1}\{|\lambda_i(x)|+|c_i|<\infty$. After choosing $N=N(n)$ such that $N(n)/n\to 0$ and $\phi(N(n))E\{\exp(-n\gamma_2((1/n)\sum_{i=1}^{n}X_i))\}\to 0$ as $n\to\infty$, see (1.2), the last inequality implies (2.13). □

3. Proof of Theorem 2.

The large deviation principle follows from Bryc [6, C.2.1] after the two claims below are established. (Notice that since the convex hull of a compact set is compact, the exponential tightness condition is trivially satisfied for compact valued X_1 and Claim B is not needed in this case. Hypotheses of Claim B follow from either (1.12) or (1.13).)

The proof of the fact that the rate function is convex, which in turn implies (1.9) by Bryc [6, C. 2.1], uses (2.12) and (2.13) again and the details are omitted.

Claim A. *If $\{X_k\}_{k\in\mathbb{Z}}$ is a stationary sequence of $\mathbb{V}$-valued random variables such that either $\psi(1)<\infty$ or $\|X_1\|$ is bounded and* (1.10) *holds, then limit* (2.1) *defining the asymptotic value $\mathbb{L}(\min_i\{\lambda_i+c_i\})$ exists and is finite for each $r\geqslant 1$, $\lambda_1,\ldots,\lambda_r\in\mathbb{V}^*$, $c_1,\ldots,c_r\in\mathbb{R}$.*

Proof. It is easy to verify that for an integrable non-negative $\mathcal{N}$-measurable ξ,

$$E\{\xi\,|\,\mathcal{M}\}\geqslant\psi_-(\mathcal{M},\mathcal{N})E\{\xi\},\tag{3.1}$$

where $\psi_-(\mathcal{M},\mathcal{N})$ is defined by the right-hand side of (1.4) with the supremum taken over $A\in\mathcal{M}$, $B\in\mathcal{N}$. Similarly

$$E\{\xi\,|\,\mathcal{M}\}\leqslant\psi_+(\mathcal{M},\mathcal{N})E\{\xi\},\tag{3.2}$$

where $\psi_+(\mathcal{M}, \mathcal{N})$ is defined by the right-hand side of (1.3) with the supremum taken over $A \in \mathcal{M}$, $B \in \mathcal{N}$ (both the inequalities are trivial for finite σ-fields $\mathcal{M}, \mathcal{N}$). Fix $r \geq 1$, $\lambda_1, \ldots, \lambda_r \in \mathbb{V}^*$, $c_1, \ldots, c_r \in \mathbb{R}$. Define

$$M_n = \operatorname{ess\,inf} E^{\mathcal{F}_{-\infty,0}}\left\{\exp \min_{1 \leq i \leq r} \{nc_i + \lambda_i(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)\}\right\}.$$

Note that $0 < M_n < \infty$ for each $n \geq 1$. Indeed, this is trivial if either (1.11) or (1.12) holds; if (1.13) holds, then for $C = \max_i\{|c_i| \vee \|\lambda_i\|\}$ we have

$$\begin{aligned} M_n &\geq \psi_-(1)\, \mathrm{e}^{-nC} E\left\{\exp\left(-C\left\|\sum_{i=1}^n X_i\right\|\right)\right\} \\ &\geq (\psi_-(1)\, \mathrm{e}^{-C} E\{\exp(-C\|\boldsymbol{X}_1\|)\})^n \end{aligned} \tag{3.3}$$

by (3.1) and

$$\begin{aligned} M_n &\leq \mathrm{e}^{nC} \psi_+(1) E\left\{\exp\left(C\left\|\sum_{i=1}^n X_i\right\|\right)\right\} \\ &\leq (\mathrm{e}^C \psi_+(1) E\{\exp(C\|\boldsymbol{X}_1\|)\})^n \end{aligned} \tag{3.4}$$

by (3.2). By stationarity $\log M_n$ is a super-additive function of n. Indeed, the super-additivity of $\log M_n$ follows from

$$\begin{aligned} M_{n+m} &= \operatorname{ess\,inf} E^{\mathcal{F}_{-\infty,0}}\left\{\exp \min_i \{(m+n)c_i + \lambda_i(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_{m+n})\}\right\} \\ &\geq \operatorname{ess\,inf} E^{\mathcal{F}_{-\infty,0}}\Big\{\exp \min_i \{nc_i + \lambda_i(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)\} \\ &\quad \times \operatorname{ess\,inf} E^{\mathcal{F}_{-\infty,n}}\left\{\exp \min_i \{mc_i + \lambda_i(\boldsymbol{X}_{n+1} + \cdots + \boldsymbol{X}_{n+m})\}\right\}\Big\} = M_m M_n. \end{aligned}$$

Hence $\lim_{n\to\infty} n^{-1} \log M_n$ exists, see e.g. Dunford, Schwartz [14, VIII.1.4] and the limit is easily seen to be finite, see (3.3) and (3.4).

Using ψ_--mixing, we shall show now that $n^{-1} \log M_n$ has the same limit as the sequence defining the asymptotic value $\mathbb{L}(\min_i\{\lambda_i + c_i\})$. The lower bound for $\mathbb{L}(\min_i\{\lambda_i + c_i\})$ is trivial as

$$M_n \leq E\left\{\exp \min_i \{nc_i + \lambda_i(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)\}\right\} \quad \text{for all } n \geq 1. \tag{3.5}$$

To obtain a suitable upper bound, take the smallest $N \geq 1$ such that $\psi_-(N) > 0$. Clearly $N = 1$, if $\psi(1) < \infty$. For each $n \geq N$ we have

$$\begin{aligned} &E^{\mathcal{F}_{-\infty,0}}\left\{\exp \min_i \{nc_i + \lambda_i(\boldsymbol{X}_1 + \cdots + \boldsymbol{X}_n)\}\right\} \\ &\quad \geq C_N E^{\mathcal{F}_{-\infty,0}}\left\{\exp \min_i \{(n-N+1)c_i + \lambda_i(\boldsymbol{X}_N + \boldsymbol{X}_{N+1} + \cdots + \boldsymbol{X}_n)\}\right\}, \end{aligned} \tag{3.6}$$

where $C_N = \exp(-(N-1)\max_{i\leq r}\sup_{x\in\operatorname{supp}\boldsymbol{X}_1}\{|\lambda_i(x)|+|c_i|\})$; under our hypotheses $C_N<\infty$ either because $\lambda_i(\boldsymbol{X}_1)$ is bounded, $1\leq i\leq r$, or because $N=1$. Therefore by (3.1),

$$\begin{aligned}
&E^{\mathscr{F}_{-\infty,0}}\Big\{\exp\min_i\{(n-N+1)c_i+\lambda_i(\boldsymbol{X}_N+\boldsymbol{X}_{N+1}+\cdots+\boldsymbol{X}_n)\}\Big\}\\
&\quad\geq\psi_-(N)E\Big\{\exp\min_i\{(n-N+1)c_i+\lambda_i(\boldsymbol{X}_N+\boldsymbol{X}_{N+1}+\cdots+\boldsymbol{X}_n)\}\Big\}\\
&\quad\geq C_N\psi_-(N)E\Big\{\exp\min_i\{nc_i+\lambda_i(\boldsymbol{X}_1+\boldsymbol{X}_2+\cdots+\boldsymbol{X}_n)\}\Big\}.
\end{aligned}$$

This and (3.6) show that $M_n\geq C_N^2\psi_-(N)\ E\{\exp\min_i\{nc_i+\lambda_i(\boldsymbol{X}_1+\boldsymbol{X}_2+\cdots+\boldsymbol{X}_n)\}\}$ and hence $\mathbb{L}(\min_i\{\lambda_i+c_i\})=\lim_{n\to\infty}n^{-1}\log M_n$ exists and is finite. □

Claim B. *If $\{\boldsymbol{X}_k\}_{k\in\mathbb{Z}}$ is a stationary sequence of $\mathbb{V}$-valued random variables such that $\psi_+(n)<\infty$ for some $n\geq 1$ and $E\{\exp(\theta\|\boldsymbol{X}_1\|)\}<\infty$ for each $\theta\in\mathbb{R}$, then the distributions $\{\mathscr{L}((\boldsymbol{X}_1+\cdots+\boldsymbol{X}_n)/n)\}_{n\geq1}$ are exponentially tight, i.e. for each $\varepsilon>0$ there is a compact set $K\subset\mathbb{X}$ such that $P((\boldsymbol{X}_1+\cdots+\boldsymbol{X}_n)/n\notin K)\leq\varepsilon^n$ for each $n\geq1$.*

Proof. It is well known, see e.g. Stroock [18, Corollary 3.27], that exponential tightness follows if

$$\sup_n\,(E\{\exp(\varepsilon q_0(\boldsymbol{X}_1+\cdots+\boldsymbol{X}_n))\})^{1/n}<\infty \tag{3.7}$$

for some $\varepsilon>0$ and some semi-norm q_0 such that $q_0^{-1}[0,1]$ is compact.By Theorem 3.1 of de Acosta [9], there exist a semi-norm q such that $q^{-1}[0,1]$ is compact and $E\{\exp q(\boldsymbol{X}_1)\}<\infty$. Let $N\geq1$ be such that $\psi_+(N)<\infty$. We shall show that (3.7) holds with $\varepsilon=1/N$ and $q_0=q$. Indeed, for $n\leq N$ by the Hölder inequality we have

$$\begin{aligned}
E\{\exp(q(\boldsymbol{X}_1+\cdots+\boldsymbol{X}_n)/N)\}&\leq E\{\exp(q(\boldsymbol{X}_1)/N+\cdots+q(\boldsymbol{X}_n)/N)\}\\
&\leq E\{\exp(q(\boldsymbol{X}_1)n/N)\}\\
&\leq E\{\exp q(\boldsymbol{X}_1)\}<\infty.
\end{aligned}$$

For $n\geq N$ write $k=[n/N]$ (integer part). Since

$$q(\boldsymbol{X}_1)+\cdots+q(\boldsymbol{X}_n)\leq\sum_{i=1}^{N}\sum_{j=0}^{k}q(\boldsymbol{X}_{i+jN}),$$

therefore

$$\begin{aligned}
E\{\exp(q(\boldsymbol{X}_1\cdots+\boldsymbol{X}_n)/N)\}&\leq E\{\exp(q(\boldsymbol{X}_1)/N+\cdots+q(\boldsymbol{X}_n)/N)\}\\
&\leq E\Big\{\exp\Big(\sum_{i=1}^{N}\sum_{j=0}^{k}q(\boldsymbol{X}_{i+jN})\Big/N\Big)\Big\}.
\end{aligned}$$

By the Hölder inequality and stationarity

$$E\Big\{\exp\Big(\sum_{i=1}^{N}\sum_{j=0}^{k}q(\boldsymbol{X}_{i+jN})/N\Big)\Big\}\leq E\Big\{\exp\Big(\sum_{j=0}^{k}q(\boldsymbol{X}_{jN})\Big)\Big\}. \tag{3.8}$$

Inequality (3.2) applied to the right-hand side of (3.8) k-times gives

$$E\left\{\exp\left(\sum_{j=0}^{k} q(\boldsymbol{X}_{jN})\right)\right\} \leqslant (\psi_+(N))^k (E\{\exp q(\boldsymbol{X}_1)\})^{k+1}.$$

Hence

$$(E\{\exp(q(\boldsymbol{X}_1+\cdots+\boldsymbol{X}_n)/N)\})^{1/n} \leqslant (\psi_+(N))^{k/n}(E\{\exp q(\boldsymbol{X}_1)\})^{(k+1)/n}.$$

Since $k/n \leqslant 1/N \leqslant 1$, (3.7) and the claim are proved. □

Note added in proof

R.C. Bradley (to appear in Stochastic Process. Appl.) gave an example of a strictly stationary sequence which satisfies (1.2) and cannot be represented as an instantaneous function of a strictly stationary real Harris recurrent Markov chain. This answers the ϕ-mixing part of the question in Remark 2 above.

Acknowledgement

I would like to thank the referee for his careful reading of the manuscript and for pointing out numerous corrections needed in the earlier versions of the paper. Comments by R.C. Bradley were also appreciated.

References

[1] H.C.P. Berbee and R.C. Bradley, A limitation of Markov representation for stationary processes, Stochastic Processes Appl. 18 (1984) 33–45.
[2] R. Bowen, Equilibrium states and the ergodic theory of Anosov diffeomorphisms, Lecture Notes in Math. No. 470 (Springer, Berlin, 1975).
[3] R.C. Bradley, Basic Properties of Strong Mixing Conditions, in: E. Eberlein and M. Taqqu eds., Dependence in Probability and Statistics (Birkhaüser, Boston, MA, 1986) pp. 165–192.
[4] R.C. Bradley, On the ϕ-mixing condition for stationary random sequences, Duke Math. J. 47 (1980) 421–433.
[5] R.C. Bradley, On the ψ-mixing condition for stationary random sequences, Trans. Amer. Math. Soc. 276 (1983) 55–66.
[6] W. Bryc, Large deviations by the asymptotic value method, in: M. Pinsky, ed., Proc. of the Conf. on Diffusion Processes (Birkhaüser, Boston, MA, 1990) pp. 447–472.
[7] W. Bryc, On the large deviation principle for stationary weakly dependent random fields, to appear in: Ann. Probab.
[8] T. Chiyonobu and S. Kusuoka, The large deviation principle for hypermixing processes, Probab. Theory Rel. Fields 78 (1988) 627–649.
[9] A. de Acosta, Upper bounds for large deviations of dependent random vectors, Z. Wahrsch. Verw. Gebiete 69 (1985) 551–565.
[10] A. de Acosta, Large deviations for empirical measures of Markov chains, J. Theoret. Probab. 3 (1990) 395–431.

[11] M.D. Donsker and S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, II, Comm. Pure Appl. Math. 28 (1975) 279-301.
[12] M.D. Donsker and S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time, III, Comm. Pure Appl. Math. 29 (1976) 389-461.
[13] J.L. Doob, Stochastic Processes (Wiley, New York, 1953).
[14] N. Dunford and J.T. Schwartz, Linear Operators I (Interscience, New York, 1958).
[15] S. Orey and S. Pelikan, Large deviation principles for stationary processes, Ann. Probab. 16 (1988) 1481-1495.
[16] M. Rosenblatt, Markov Processes. Structure and Asymptotic Behavior (Springer, Berlin, 1971).
[17] R.H. Schonmann, Exponential convergence under mixing, Probab. Theory Rel. Fields 81 (1989) 235-238.
[18] D.W. Stroock, An introduction to the theory of large deviations (Springer, New York, 1984).
[19] D.A. Yakimavichyus, On large deviation theorems for dependent random variables, Soviet Math. Doklady 39 (1989) 35-37.