# Large Deviations and Strong Mixing

Włodzimierz Bryc*
Department of Mathematics
University of Cincinnati
Cincinnati, OH 45 221
bryc@uc.edu

Amir Dembo†
Department of Mathematics and Department of Statistics
Stanford University
Stanford, CA 94 305
amir@playfair.stanford.edu

Annales de l'I.H.P. Probabilitiés & Statistiques, 32(1996), 549–569

**Résumé** Nous prouvons le propriétés de grandes déviations (P.G.D.) pour les mesures empiriques en $\tau$-topologie, dans les cas de suites stationnaires sous conditions de mélange $\alpha(n) \ll \exp(-n(\log n)^{1+\delta})$ pour certain $\delta > 0$, ou $\phi(n) \ll \exp(-n\ell(n))$ avec $\ell(n) \to \infty$. Les examples de chaînes de Markov récurrentes au sense de Doeblin montrent que ces conditions ne permettrent pas de amélioration substantielle, et que l'existence même du P.G.D. dépend du choix de la mesure initiale.

**Abstract** The Large Deviation Principle (LDP) with respect to the $\tau$-topology holds for the empirical measure of any $\alpha$-mixing or any $\phi$-mixing stationary process with a hyper-exponential mixing rate of at least $\alpha(n) \ll \exp(-n(\log n)^{1+\delta})$, for some $\delta > 0$ or at least $\phi(n) \ll \exp(-n\ell(n))$ with $\ell(n) \to \infty$. Positive recurrent Doeblin Markov chain examples for which the LDP does not hold demonstrate the tightness of these rates and the relationship between the LDP for the empirical means of all bounded $\mathbb{R}^d$-valued functionals and the LDP for the empirical measure.

## 1 Introduction.

We shall say that a sequence of probability measures $\{\mu_n\}$ on a topological space $\mathcal{X}$ equipped with a $\sigma$-field $\mathcal{B}$ satisfies the Large Deviation Principle (LDP), if there is a lower semicontinuous rate function $I : \mathcal{X} \to [0, \infty]$, with compact level sets $I^{-1}([0, a])$ for all $a > 0$, and

*Partially supported by C.P. Taft Memorial Fund
†Partially supported by NSF DMS92-09712 grant and by a US-ISRAEL BSF grant.
**Key Words:** large deviations, empirical measure, strong mixing, bounded additive functionals.
**AMS (1991) Subject Classification:** 60F10

such that for all $\Gamma \in \mathcal{B}$,

$$- \inf_{x \in \Gamma^o} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq - \inf_{x \in \overline{\Gamma}} I(x) \,,$$

where $\overline{\Gamma}$ denotes the closure of $\Gamma$, $\Gamma^o$ the interior of $\Gamma$, and the infimum of a function over an empty set is interpreted as $\infty$. The left-most inequality is called the (large deviations) lower bound and the right-most is called the upper bound, while the *weak* LDP corresponds to the upper bound holding (only) for pre-compact $\Gamma \in \mathcal{B}$ and with $I^{-1}([0,a])$ required to be (only) closed sets.

Note that $\mathcal{B}$ need not necessarily be $\mathcal{B}_{\mathcal{X}}$ – the Borel $\sigma$-field of $\mathcal{X}$.

Let $\{Y_i\}_{i=0}^{\infty}$ be a stationary sequence of random variables which take values in a Polish space $\Sigma$, and consider the empirical measures

$$L_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{Y_i} \,,$$

where $\delta_y$ denote the probability measure degenerate at $y \in \Sigma$. Let $M_1(\Sigma)$ denote the space of (Borel) probability measures on $\Sigma$ equipped with the $\tau$-topology generated by the collection

$$\{\nu \in M_1(\Sigma) \,:\, |\int_{\Sigma} f d\nu - x| < \delta\} \,,$$

where $x \in \mathbb{R}$, $\delta > 0$ and $f \in B(\Sigma, \mathbb{R})$ – the vector space of all bounded, real-valued, Borel measurable functions on $\Sigma$. Let $\mathcal{B}^{cy}$ be the cylinder $\sigma$-field on $M_1(\Sigma)$, i.e., the smallest $\sigma$-field that for any fixed $f \in B(\Sigma, \mathbb{R})$ makes $\nu \mapsto \int_{\Sigma} f d\nu$ a measurable map from $M_1(\Sigma)$ to $\mathbb{R}$. With $L_n$ measurable with respect to $\mathcal{B}^{cy}$, let $\mu_n$ be the law of $L_n$ on $(M_1(\Sigma), \mathcal{B}^{cy})$. With $\Sigma$ Polish, $\mathcal{B}^{cy}$ is merely the Borel $\sigma$-field associated with weak convergence in $M_1(\Sigma)$ (see for example [2, Lemma 2.1]).

Sanov's well known theorem states that $\mu_n$ satisfies the Large Deviation Principle when $Y_i$ are i.i.d. random variables (see [13, Section 3.2] or [12, Section 6.2] and the references therein, see also [11] for extension to arbitrary measurable space). This result has been extended to a rather large class of Markov chains (see [10, 12, 13, 20] and the references therein for earlier works, most notably of Donsker and Varadhan). In a different direction, this LDP has been shown to hold for stationary processes satisfying the hypermixing conditions of [8], for $\psi$-mixing processes and in compact spaces also for $\phi$-mixing processes of hyper-exponential mixing rate (see [6]). On the other hand, [1] show that the LDP might fail for empirical measures of Doeblin chains which are $\phi$-mixing with exponential rate, and an example of a $\beta$-mixing process such that the Large Deviation Principle fails for a bounded linear functional of $L_n$ (with values in $\{-1, 0, 1\}$) is provided in [7], see also Section 4.

A natural question, addressed here is therefore "is there an $\alpha$-mixing rate that suffices to guarantee the Large Deviation Principle for the empirical measures ?" (recall the following chain of implications

$$\psi - \text{mixing} \implies \phi - \text{mixing} \implies \beta - \text{mixing} \implies \alpha - \text{mixing}$$

and see [3] for a review of the relations between these mixing conditions).

The sufficient rate is $\alpha(n) = \exp(-n(\log n)^{1+\delta})$ for some $\delta > 0$ (see Proposition 2 below).

Theorem 1, the main result of this note, states that if $\{Y_i\}$ possesses the mixing condition (S) (see definition below) then $\{\mu_n\}$ satisfies the Large Deviation Principle in $(M_1(\Sigma), \mathcal{B}^{cy})$ equipped with the $\tau$-topology. Condition (S) bellow suffices for the empirical process LDP to hold, though it seems that the rate function might in general be strictly less than the specific entropy function.

The analog of Theorem 1 for continuous time processes is stated in Theorem 2.

In Section 3 we bring the definitions of the various mixing conditions mentioned above and show in particular that either fast $\alpha$-mixing or hypermixing or $\psi$-mixing suffice for (S) (see also Lemma 4 for (S) in the context of Markov chains). The method of proof of Theorem 1 also allows for extending the result of [6, Theorem 3] to the non-compact setting, i.e., proving that the LDP holds for the empirical measure (and the empirical process) provided that

$$\frac{1}{n}\log(1/\phi(n)) \to \infty \,. \tag{1}$$

Doeblin recurrent, irreducible, countable state Markov chains are studied in Section 4. First, we show that for these chains the LDP does hold for the empirical means of *every* bounded $\mathbb{R}^d$-valued function of the state, and for all finitely supported initial measures. In contrast, modifying [7, Proposition 4.1] we provide such a chain and a set $\mathcal{D} \subset \Sigma$ such that under the invariant measure of the chain the LDP for $\{n^{-1}\sum_{i=1}^n 1_{\mathcal{D}}(Y_i)\}$, the empirical frequency of visits to $\mathcal{D}$, does not hold. This example demonstrates the tightness of (1) and also shows that the *mere existence* of the LDP is sensitive to the choice of the initial measure even for *irreducible* Markov chains. Then, we provide an example of such a chain with the LDP holding for the empirical means of every bounded $\mathbb{R}^d$-valued function of the state even under its invariant measure, yet *for any initial distribution* the LDP fails for its empirical measures.

# 2 Large Deviation Principle under (S) Mixing

**Definition:** We shall say that condition (S) is satisfied if for every $C < \infty$ there is a non-decreasing sequence $\ell(n) \in \mathbb{N}$ with

$$\sum_{n=1}^{\infty} \frac{\ell(n)}{n(n+1)} < \infty \tag{2}$$

such that

$$(S_-):\ \sup\{P(A)P(B) - e^{\ell(n)}P(A\cap B)\ :\ A \in \mathcal{F}_0^{k_1},\, B \in \mathcal{F}_{k_1+\ell(n)}^{k_1+k_2+\ell(n)}\ k_1, k_2 \in \mathbb{Z}_+\} \leq e^{-Cn},$$

$$(S_+):\ \sup\{P(A\cap B) - e^{\ell(n)}P(A)P(B)\ :\ A \in \mathcal{F}_0^{k_1},\, B \in \mathcal{F}_{k_1+\ell(n)}^{k_1+k_2+\ell(n)}\ k_1, k_2 \in \mathbb{Z}_+\} \leq e^{-Cn},$$

where $\mathcal{F}_a^b = \sigma(Y_i : a \leq i \leq b)$.

**Remark:** Conditions $(S_-)$ and $(S_+)$ are reminiscent of the double-mixing condition of [5] and can be similarly interpreted as arising from $\psi_-$ and $\psi_+$ mixing (see Section 3 below) except on small sets.

The connection with moment estimates is given by the following.

**Lemma 1** *Fix two $\sigma$-fields $\mathcal{F}$ and $\mathcal{G}$. If*

$$\sup_{A\in\mathcal{F},B\in\mathcal{G}}[P(A)P(B)-aP(A\cap B)]\le b\ ,$$

*then for all non-negative random variables $W\in L_\infty(\mathcal{F})$, $Z\in L_\infty(\mathcal{G})$*

$$E(W)E(Z)-aE(WZ)\le b\|W\|_\infty\|Z\|_\infty \tag{3}$$

*If*

$$\sup_{A\in\mathcal{F},B\in\mathcal{G}}[P(A\cap B)-aP(A)P(B)]\le b\ ,$$

*then for all non-negative random variables $W\in L_\infty(\mathcal{F})$, $Z\in L_\infty(\mathcal{G})$*

$$E(WZ)-aE(W)E(Z)\le b\|W\|_\infty\|Z\|_\infty \tag{4}$$

**Proof.** By Fubini's theorem for the non-negative $W$, $Z$ we have

$$E(W)E(Z)-aE(WZ)=\int_0^{\|W\|_\infty}\int_0^{\|Z\|_\infty}[P(W>t)P(Z>s)-aP(W>t,Z>s)]dtds$$

and (3) follows. The proof of (4) is done analogously. ■

Our main result is the following theorem.

**Theorem 1** *If (S) holds for the stationary sequence $\{Y_i\}$, then $\{\mu_n\}$ satisfies the Large Deviation Principle with respect to the $\tau$-topology in $M_1(\Sigma)$ and with the (convex) rate function*

$$I(\nu)=\sup_{f\in B(\Sigma,\mathbb{R})}\{\int_\Sigma fd\nu-\Lambda(f)\}\ , \tag{5}$$

*where*

$$\Lambda(f)=\lim_{n\to\infty}\frac{1}{n}\Lambda_n(f)=\lim_{n\to\infty}\frac{1}{n}\log E(\exp(\sum_{i=1}^n f(Y_i)))\ . \tag{6}$$

*In particular the limit in (6) exists for every $f\in B(\Sigma,\mathbb{R})$. The weaker condition ($S_-$) suffices for the LDP to hold for the empirical means $\{n^{-1}\sum_{i=1}^n f(Y_i)\}$, for every $f\in B(\Sigma,\mathbb{R}^d)$.*

Clearly, if (S) holds for $\{Y_i\}$ it follows that for each $r\in\mathbb{N}$, (S) also holds for the process $\{(Y_i,\dots,Y_{i+r-1})\}$ which takes values in the product space $\Sigma^r$. Consequently, by Theorem 1, the $r$-empirical measures

$$L_{n,r}=\frac{1}{n}\sum_{i=1}^n\delta_{Y_i,\dots,Y_{i+r-1}}\ ,$$

then satisfy the LDP in $M_1(\Sigma^r)$ equipped with the $\tau$-topology and with a convex rate function $\Lambda_r^*(\cdot)$ which is the Fenchel–Legendre transform (on $B(\Sigma^r,\mathbb{R})$) of

$$\Lambda^{(r)}(f)=\lim_{n\to\infty}n^{-1}\log E(\exp(\sum_{i=1}^n f(Y_i,\dots,Y_{i+r-1})))\ .$$

Let $M_1(\Sigma^{\mathbb{N}})$ be equipped with the cylinder $\sigma$-field $\mathcal{B}^{cy}$, and with the projective limit of the $\tau$-topologies of $M_1(\Sigma^r)$, i.e., the weakest topology making $Q \mapsto \int_{\Sigma^r} f d(\pi_r Q)$ continuous for every $f \in B(\Sigma^r, \mathbb{R})$ and all $r \in \mathbb{N}$, where $\pi_r : M_1(\Sigma^{\mathbb{N}}) \to M_1(\Sigma^r)$ is the projection mapping of $Q$ to its marginal on the first $r$ coordinates (compare with [13, Exercise 5.4.40]). It now follows by [9, Theorem 3.3] that the process level empirical measure

$$L_{n,\infty} = \frac{1}{n} \sum_{i=1}^{n} \delta_{T^i \mathbf{Y}}$$

(where $T^i \mathbf{Y} = (Y_i, Y_{i+1}, \ldots)$ ) satisfies the LDP with respect to the latter topology, and with the rate function

$$I_\infty(Q) = \sup_{r \in \mathbb{N}} \Lambda_r^*(\pi_r Q)$$

(a similar derivation is presented in [12, Section 6.5.3] in the context of Markov chains). Following the argument of [13, Equations (5.4.15) and (5.4.19)] one can verify that $I_\infty(Q) = \infty$ when $Q$ is not shift-invariant and that $I_\infty(Q)$ is bounded above by the specific entropy function (see [13, Equation (5.4.8)] for definition). It seems however that in general $I_\infty(Q)$ might be strictly less than the specific entropy function for some $Q \in M_1(\Sigma^{\mathbb{N}})$.

The proof of Theorem 1 builds on [12, Section 6.4], with the following application of the Dawson–Gärtner projective limit theorem [9, Theorem 3.3] serving as a useful tool (see also [10] where it is applied when proving upper bounds for empirical measures of Markov chains).

**Lemma 2** *Let $\mathcal{P}_a(\Sigma)$ denote the space of all finitely additive non-negative set functions on $\mathcal{B}_\Sigma$ assigning unit mass to $\Sigma$, equipped with the $B(\Sigma, \mathbb{R})$-topology and the (relevant) cylinder $\sigma$-field $\mathcal{B}^{cy}$.*
*(a) With $\mu_n(A) = \mu_n(A \cap M_1(\Sigma))$ for all $A \in \mathcal{B}^{cy}$, $A \subset \mathcal{P}_a(\Sigma)$, the sequence $\{\mu_n\}$ satisfies the Large Deviation Principle in $\mathcal{P}_a(\Sigma)$ iff for every $f \in B(\Sigma, \mathbb{R}^d)$ the sequence $\{\hat{S}_n = n^{-1} \sum_{i=1}^{n} f(Y_i)\}$ satisfies the LDP in $\mathbb{R}^d$.*
*(b) If for every $\eta > 0$ and any sequence $A_j \in \mathcal{B}_\Sigma$ decreasing monotonically to the empty set, $\lim_j \Lambda(\eta 1_{A_j}) = 0$ (with $\Lambda(\cdot)$ given in (6)), then $\{\mu_n\}$ also satisfies the LDP with respect to the $\tau$-topology in $M_1(\Sigma)$. The rate function for this LDP is given by (5) provided the rate functions for the LDP of $\{\hat{S}_n\}$ are all convex.*

**Remark 1** *Clearly, the LDP in $\mathcal{P}_a(\Sigma)$ with a rate function $J$ implies that $\{\mu_n\}$ satisfies the weak LDP in $M_1(\Sigma)$ for the function $J$ restricted to $M_1(\Sigma)$.*

**Proof.** (a) Since $\hat{S}_n = \int_\Sigma f dL_n$, the LDP in force for every $f \in B(\Sigma, \mathbb{R}^d)$ and every $d < \infty$ implies by [9, Theorem 3.3] that $\{\mu_n\}$ satisfies the LDP in $\mathcal{X}$ – the algebraic dual of $B(\Sigma, \mathbb{R})$ equipped with the $B(\Sigma, \mathbb{R})$-topology. The LDP in $\mathcal{P}_a(\Sigma)$ follows by the identification of $\mathcal{P}_a(\Sigma)$ with a closed subset of $\mathcal{X}$ via $\langle \omega, 1_A \rangle \mapsto \omega(A)$ (where $\langle \omega, f \rangle$ denotes the value that the linear functional $\omega$ assigns to $f \in B(\Sigma, \mathbb{R})$). The converse implication (that LDP in $\mathcal{P}_a(\Sigma)$ yields LDP for all $f \in B(\Sigma, \mathbb{R}^d)$) is trivial.
(b) The LDP in $\mathcal{X}$ implies that $\Lambda(f)$ exists and is finite everywhere with

$$\Lambda^*(\omega) = \sup_{f \in B(\Sigma, \mathbb{R})} \{\langle \omega, f \rangle - \Lambda(f)\} ,$$

being the affine regularization of the rate function $I(\cdot)$ (see [12, Theorem 4.5.10]). In particular, $I(\omega) \geq \Lambda^*(\omega)$ for all $\omega \in \mathcal{X}$. Fixing $\omega \in \mathcal{P}_a(\Sigma)$ which is not countably additive, there exists $\epsilon > 0$ and a sequence $A_j \in \mathcal{B}_\Sigma$ decreasing monotonically to the empty set, such that $\omega(A_j) = \langle \omega, 1_{A_j} \rangle \geq \epsilon$ for all $j$. By our assumptions, for every $\eta > 0$,

$$I(\omega) \geq \Lambda^*(\omega) \geq \limsup_{j\to\infty} \{\langle \omega, \eta 1_{A_j} \rangle - \Lambda(\eta 1_{A_j})\} \geq \eta\epsilon ,$$

and taking $\eta \to \infty$ we conclude that $I(\omega) = \infty$. The LDP in $M_1(\Sigma)$ now follows by observing that the $\tau$-topology is the relative topology induced by $\mathcal{P}_a(\Sigma)$ on $M_1(\Sigma)$ with the latter supporting $\{\mu_n\}$ and containing the set of points of finite rate. If the rate functions for $\{\hat{S}_n\}$ are convex then by [9, Theorem 3.3] so is $I(\cdot)$ and necessarily $I = \Lambda^*$ is given by (5). ■

In particular, we shall first prove the existence of limits like (6) for suitably chosen functions based on Hammersley's [19, Theorem 2] approximate sub-additivity lemma, quoted for completeness.

**Lemma 3 (Approximate sub-additivity)** *Assume $h : \mathbb{N} \to \mathbb{R}$ is such that for all $n, m \geq 1$,*

$$h(n+m) \leq h(n) + h(m) + \Delta(n+m) ,$$

*where $\Delta(n)$ is a non-decreasing sequence satisfying*

$$\sum_{r=1}^{\infty} \frac{\Delta(r)}{r(r+1)} < \infty .$$

*Then $\lambda = \lim_{n\to\infty} [h(n)/n]$ exists, $\lambda < \infty$ and for all $m \in \mathbb{N}$,*

$$\lambda \leq \frac{h(m)}{m} - \frac{\Delta(m)}{m} + 4 \sum_{r=2m}^{\infty} \frac{\Delta(r)}{r(r+1)} . \tag{7}$$

**Proof of Theorem 1.** First we show that for every $f \in B(\Sigma, \mathbb{R}^d)$, the sequence $\hat{S}_n = \int_\Sigma f dL_n$ satisfies the Large Deviation Principle in $\mathbb{R}^d$ with a convex rate function. This result is proved by adapting the proof of [12, Theorem 6.4.4], with $(S_-)$ replacing Assumption 6.4.1 of [12, page 253] and Lemma 3 replacing [12, Lemma 6.4.10]. Indeed, $\hat{S}_n$ is the empirical mean of the stationary process $X_i = f(Y_i)$ taking values in a compact (convex) subset of $\mathbb{R}^d$. Assumption 6.4.1 is used twice in the proof of Theorem 6.4.4 of [12]; first for any $B < \infty$ and every fixed concave continuous $g : \mathbb{R}^d \to [-B, 0]$ to uniformly bound below the quantity

$$\frac{1}{\ell} \log \left\{ \frac{E(\exp(ng(\hat{S}_n)) \exp(mg(m^{-1} \sum_{j=1}^m X_{j+n+\ell})))}{E(\exp(ng(\hat{S}_n)))E(\exp(mg(\hat{S}_m)))} \right\} , \tag{8}$$

with $\ell = \ell(n+m)$ any non-decreasing sequence satisfying (2). Choosing the sequence $\ell(n)$ corresponding to $C = B + 1$ in $(S_-)$ and applying (3) for $W = \exp(ng(\hat{S}_n))$ and $Z = \exp(mg(m^{-1} \sum_{j=1}^m X_{j+n+\ell}))$ (with $\|W\|_\infty = \|Z\|_\infty = 1$ and $E(W)E(Z) \geq e^{-B(n+m)}$), it follows that (8) is bounded below by $-1+\log(1-e^{-2})$. Assumption 6.4.1 is needed once more in [12] for showing that if for some $M<\infty$ and fixed open sets $G$, $G'$

$$P(\hat{S}_n \in G)P(\hat{S}_n \in G') \geq \exp(-Mn)$$

for all $n$ large enough, then

$$\liminf_{\eta\downarrow 0} \liminf_{n\to\infty} \rho(n,\eta) \geq 0 \ , \tag{9}$$

where

$$\rho(n,\eta) = \frac{1}{n}\log\left\{\frac{P(\hat{S}_n \in G\ ,\ \sum_{j=1}^{n} X_{j+n+\eta n} \in G')}{P(\hat{S}_n \in G)P(\hat{S}_n \in G')}\right\} \ .$$

Choosing the sequence $\ell(n)$ corresponding to $C=M+1$ in $(S_-)$ we have $\ell(n)\leq \eta n$ for all $n$ large enough and hence applying $(S_-)$ to $A=\{\hat{S}_n \in G\}$ and $B=\{\sum_{j=1}^n X_{j+n+\eta n}\in G'\}$ yields the bound $\rho(n,\eta)\geq -2\eta$ for all $n\geq n_0(\eta)$ and (9) follows.

With the LDP for $\{\hat{S}_n\}$ with a convex rate function holding for every $d<\infty$ and any $f\in B(\Sigma,\mathbb{R}^d)$, we fix $\eta>0$ and a sequence $A_j\in\mathcal{B}_\Sigma$ monotonically decreasing to the empty set. Note that for every $m,n,\ell,j\in\mathbb{N}$

$$\sum_{i=1}^{n+m}(1_{A_j}(Y_i)-1) \leq \sum_{i=1}^{n}(1_{A_j}(Y_i)-1) + \sum_{i=1}^{m}(1_{A_j}(Y_{i+n+\ell})-1) + 2\ell \ .$$

Choosing the sequence $\ell=\ell(n+m)$ corresponding to $C=\eta$ in $(S_+)$ and applying (4) to $W=\exp(\eta\sum_{i=1}^n(1_{A_j}(Y_i)-1))$ and $Z=\exp(\eta\sum_{i=1}^m(1_{A_j}(Y_{i+n+\ell})-1))$ (with $\|W\|_\infty = \|Z\|_\infty = 1$ and $E(W)E(Z)\geq e^{-\eta(n+m)}$), it follows that for every $\eta>0$

$$\Lambda_{n+m}(\eta(1_{A_j}-1)) \leq \Lambda_n(\eta(1_{A_j}-1)) + \Lambda_m(\eta(1_{A_j}-1)) + 2(\eta+1)\ell \ .$$

Therefore, Lemma 3 applies with the sequence $\Delta(n)=2(\eta+1)\ell(n)$ which is *independent* of $j$. With $m^{-1}\Lambda_m(\eta 1_{A_j}) = m^{-1}\Lambda_m(\eta(1_{A_j}-1))+\eta$, we consequently have by (7) that for every $m,j\in\mathbb{N}$ and any $\eta>0$

$$\Lambda(\eta 1_{A_j}) = \Lambda(\eta(1_{A_j}-1))+\eta \leq m^{-1}\Lambda_m(\eta 1_{A_j}) + 8(\eta+1)\sum_{r=2m}^{\infty}\frac{\ell(r)}{r(r+1)} \ .$$

Note that

$$\Lambda_m(\eta 1_{A_j}) \leq \log(1+e^{\eta m}\sum_{i=1}^{m} E(1_{A_j}(Y_i))) \ ,$$

implying by the countable additivity of the laws of $Y_i$ that $\Lambda_m(\eta 1_{A_j})\to 0$ as $j\to\infty$. Taking now $m\to\infty$ it follows that $\Lambda(\eta 1_{A_j})\to 0$.

The proof is now completed by applying Lemma 2. ■

Following the treatment of the i.i.d. case in [17, Section 5] we deduce from Theorem 1 that if condition (S) holds for $\Sigma$-valued stationary sequence $\{Y_i\}$ in separable Banach space $\Sigma=(E,\|\cdot\|)$ such that $\|Y\|\leq K$, then the sequence $\{\hat{S}_n = n^{-1}\sum_{i=1}^n Y_i\}$ satisfies the LDP in $E$ with the convex rate function

$$I'(x) = \sup_{\lambda\in E^*}\{\langle\lambda,x\rangle - \Lambda(\lambda)\} \tag{10}$$

(where $\Lambda(\lambda)$ is as in (6) but with $\langle \lambda, Y_i \rangle$ replacing $f(Y_i)$). Indeed, $L_n(B_{0,K}) = 1$ with $B_{0,K} = \{x : \|x\| \leq K\}$ closed, so by Theorem 1 we know that $\{\mu_n\}$ satisfies the LDP with a convex rate function in $\mathcal{X} = M_1(B_{0,K})$ equipped with the topology of convergence in law. For $\nu \in \mathcal{X}$ let $m(\nu)$ be the unique element of $E$ such that

$$\langle \lambda, m(\nu) \rangle = \int_E \langle \lambda, x \rangle \nu(dx) \quad \forall \lambda \in E^* \ ,$$

where by [13, Lemma 3.3.8], the map $m : \mathcal{X} \to E$ is well defined and continuous. Plainly, $\hat{S}_n = m(L_n)$ and the LDP for $\{\hat{S}_n\}$ follows by the contraction principle (see [12, Theorem 4.2.1]); the rate function for the latter LDP is convex and (10) follows (see [12, Theorem 4.5.10]). The above observation is applicable whenever the LDP holds for the empirical measures of bounded $(E, \| \cdot \|)$-valued random variables, for example in the context of Proposition 3. It can also be extended to unbounded $Y_i$ provided that $E(\exp(s\|Y\|)) < \infty$ for all $s \in \mathbb{R}$ and the hypermixing condition (H–1) holds (this is done by [13, Lemma 3.3.10 and Exercise 2.1.20 (i)]).

Theorem 1 is now extended to the following continuous time setting. Let $\Sigma$ be a Polish space, and denote by $\Omega$ the space $D([0,\infty), \Sigma)$ of right-continuous paths $\omega : [0,\infty) \to \Sigma$ which have a left-limit at each $t > 0$, equipped with a $\sigma$-field $\mathcal{B}$ such that $Y_t(\omega) = \omega_t$ are measurable with respect to $\mathcal{B}_\Sigma$ for each fixed $t$. The occupation time $L_T(\omega, \cdot)$ is the measurable map from $(\Omega, \mathcal{B})$ to $(M_1(\Sigma), \mathcal{B}^{cy})$ such that for every $A \in \mathcal{B}_\Sigma$

$$L_T(\omega, A) = T^{-1} \int_0^T 1_A(Y_t(\omega)) dt \ .$$

For each $T > 0$, let $\mu_T$ denote the law of $L_T$ induced by a (fixed) stationary probability measure $P$ on $(\Omega, \mathcal{B})$.

Modifying the definition of the mixing condition (S) to have $k_1, k_2 \in [0, \infty)$ and $\ell : [0,\infty) \to (0,\infty)$ non-decreasing such that $\int_{r=\xi}^\infty \ell(r) dr/r^2 < \infty$ for some $\xi > 0$, we have the following analog of Theorem 1.

**Theorem 2** *If (S) holds, then $\{\mu_T\}$ satisfies the Large Deviation Principle with respect to the $\tau$-topology in $M_1(\Sigma)$ and with the convex rate function given by (5) where now*

$$\Lambda(f) = \lim_{T \to \infty} \frac{1}{T} \log E(\exp(\int_0^T f(Y_t) dt)) \ .$$

*If only ($S_-$) holds, then the LDP holds for $\{T^{-1} \int_0^T f(Y_t) dt\}$ for every $f \in B(\Sigma, \mathbb{R}^d)$.*

**Proof.** The analog of Lemma 3 for subadditive functions $h : [0, \infty) \to \mathbb{R}$ is also given in [19, Theorem 2], where the condition $\sum_{r=1}^\infty \Delta(r)/(r(r+1)) < \infty$ and the term $\sum_{r=2m}^\infty \Delta(r)/(r(r+1))$ in (7) are replaced by $\int_{r=\xi}^\infty \Delta(r) dr/r^2 < \infty$ for some $\xi > 0$ and $\int_{r=2m}^\infty \Delta(r) dr/r^2$, respectively. Theorem 2 is now proved by adapting the proof of Theorem 1. ■

## 3 Hypermixing, $\psi$-mixing, $\alpha$-mixing, $\phi$-mixing and (S)

It is shown in [6, Theorem 2] that the empirical means of bounded stationary separable Banach space valued random variables satisfy the LDP as soon as $\inf_\ell \psi(\ell) < \infty$, where

$\psi(\ell) = \psi_+(\ell)/\psi_-(\ell)$ and

$$\begin{aligned}\psi_+(\ell) &= \sup\{\frac{P(A\cap B)}{P(A)P(B)} : P(A)P(B)>0,\, A\in\mathcal{F}_0^k,\, B\in\mathcal{F}_{k+\ell}^\infty, k\in\mathbb{Z}_+\}\\ \psi_-(\ell) &= \inf\{\frac{P(A\cap B)}{P(A)P(B)} : P(A)P(B)>0,\, A\in\mathcal{F}_0^k,\, B\in\mathcal{F}_{k+\ell}^\infty, k\in\mathbb{Z}_+\}\ ,\end{aligned}$$

Noting that $\psi_-(\ell)\le 1\le\psi_+(\ell)$ one can readily check that $\psi(m)<\infty$ implies that (S) holds with any constant $\ell=\ell(n)>\log\psi(m)$ (independent of $C$). Hence, Theorem 1 yields the corresponding LDP for the empirical measures. Similarly, condition $(S_-)$ which is weaker than $\sup_\ell \psi_-(\ell)>0$ suffices for the LDP to hold for the empirical means of $f(Y_i)$ for every $f\in B(\Sigma,\mathbb{R}^d)$ (see also [24, Corollary 4.1] where the same LDP is obtained for a different weakening of the condition $\psi_-(1)>0$).

The LDP for $\{\mu_n\}$ as in Theorem 1 is proved in [12, Lemma 6.4.18] assuming that the sequence $\{Y_i\}$ satisfies a slight modification of the hypermixing conditions (H–1) and (H–2) of [8, 13]. The hypermixing condition (H–2) of [12] is as follows:

**(H–2)** *There exist $\beta(\ell)\in[1,\infty]$ and $\gamma(\ell)\ge 0$ such that for all $k_1,k_2\in\mathbb{Z}_+$, and every $W\in L_\infty(\mathcal{F}_0^{k_1})$, and $Z\in L_\infty(\mathcal{F}_{k_1+\ell}^{k_1+k_2+\ell})$,*

$$|E(W)E(Z)-E(WZ)|\le\gamma(\ell)E\left(|W|^{\beta(\ell)}\right)^{1/\beta(\ell)}E\left(|Z|^{\beta(\ell)}\right)^{1/\beta(\ell)}\ , \tag{11}$$

*where $\lim_{\ell\to\infty}\gamma(\ell)=0$, while for some $\delta>0$*

$$\lim_{\ell\to\infty}(\beta(\ell)-1)\ell(\log\ell)^{1+\delta}=0\ . \tag{12}$$

The next proposition relates (H–2) with (S).

**Proposition 1** *If the hypermixing condition (H–2) holds then (S) holds as well.*

**Remark:** Proposition 1 shows that Theorem 1 improves the results of [12, Section 6.4] by not requiring the condition (H–1). This is to be compared with the LDP for $n^{-1}\sum_{i=1}^n f(Y_i)$ which when both (H–1) and (H–2) are satisfied holds also for *unbounded* real-valued Borel measurable $f$ if $E(\exp(sf(Y)))<\infty$ for all $s\in\mathbb{R}$.

The proof of Proposition 1 also shows that it is enough to assume (H-2) for indicator functions $W=1_A, Z=1_B$ only. This fact is not obvious in light of [4, Proposition 2.5].

**Proof.** Note that (H–2) implies that for some finite $\ell_0\ge 1$ and all $\ell\ge\ell_0$ both $\beta(\ell)\le 2$ and $\gamma(\ell)\le 1/2$. If $\beta(m)=1$ for some $m\ge\ell_0$ then (11) implies in particular that $\psi(m)\le 3$ and consequently condition (S) holds. Assuming otherwise, fix $C<\infty$, and let $\ell(n)\ge\ell_0$ be the smallest integer such that

$$e^{-Cn}\ge e^\ell(2(1-e^{-1}))^{-1/(\beta(\ell)-1)}$$

(the sequence $\ell(n)$ is well defined in view of (12)). For $W=1_A$ and $Z=1_B$ it follows from (11) that $P(A\cap B)\in[x-\gamma(\ell)x^{1/\beta(\ell)}, x+\gamma(\ell)x^{1/\beta(\ell)}]$, where $x=P(A)P(B)\in[0,1]$. Hence, for all $\ell\ge\ell_0$

$$\begin{aligned}\max\{P(A\cap B)-e^\ell P(A)P(B) &, \quad P(A)P(B)-e^\ell P(A\cap B)\}\\ &\le \sup_{x\in[0,1]}\{e^\ell\gamma(\ell)x^{1/\beta(\ell)}-(e^\ell-1)x\}\\ \le e^\ell\sup_{x\ge 0}\{\beta(\ell)x^{1/\beta(\ell)}-2(1-e^{-1})x\} &\le e^\ell(2(1-e^{-1}))^{-1/(\beta(\ell)-1)}\ ,\end{aligned}$$

and in particular both $(S_-)$ and $(S_+)$ hold for $\ell = \ell(n)$. By (12) we have for all $n$ large enough $\ell(n) \le (C+1)n/(\log n)^{1+\delta}$. Therefore, $\ell(n)$ satisfies (2) and the mixing condition (S) holds as claimed. ∎

Recall that the $\alpha$-mixing rate is

$$\alpha(\ell) = \sup\{|P(A \cap B) - P(A)P(B)| \,:\, A \in \mathcal{F}_0^k,\, B \in \mathcal{F}_{k+\ell}^{\infty}, k \in \mathbb{Z}_+\} \,.$$

The following proposition states the $\alpha$-mixing rate which suffices for (S) to hold, and hence also for the LDP to hold for $\{\mu_n\}$.

**Proposition 2** *(S) holds if for some $\delta > 0$,*

$$\lim_{n\to\infty} \frac{\log \alpha(n)}{n(\log n)^{1+\delta}} = -\infty \,. \tag{13}$$

**Proof.** Fix $C < \infty$ and let $\ell = \ell(n)$ be the smallest integer such that $e^{-Cn} \ge \alpha(\ell)e^{\ell}$ (note that (13) implies that $\alpha(\ell)e^{\ell} \to 0$ as $\ell \to \infty$). It is easy to check that both $(S_-)$ and $(S_+)$ are satisfied for this sequence. Plainly, (13) implies that for $n$ large enough $\ell(n) \le (C+1)n/(\log n)^{1+\delta}$. Hence, $\ell(n)$ satisfies (2) and (S) holds as claimed. ∎

Recall that the $\phi$-mixing rate is

$$\phi(\ell) = \sup\{|P(B|A) - P(B)| \,:\, A \in \mathcal{F}_0^k, P(A) > 0,\, B \in \mathcal{F}_{k+\ell}^{\infty}, k \in \mathbb{Z}_+\} \,.$$

The next proposition extends [6, Theorem 3] to non-compact spaces, thus settling the issue of critical $\phi$-mixing rate for empirical measures LDP – hyper-exponential rate implies LDP, while an exponential rate does not imply LDP even in the weak topology and for Markov chains (see also Section 4 below). This proposition is not based on the mixing condition (S) of Theorem 1. Similar result was found independently by Hu [22].

**Proposition 3** *If $\{Y_i\}$ is a $\phi$-mixing stationary $\Sigma$-valued sequence such that (1) holds, then $\{\mu_n\}$ satisfies the Large Deviation Principle with respect to the $\tau$-topology in $M_1(\Sigma)$ and with the (convex) rate function of (5).*

**Proof.** By [6, Theorem 1], for every $d < \infty$ and any $f \in B(\Sigma, \mathbb{R}^d)$, the LDP with a convex rate function holds for $\hat{S}_n = \int_\Sigma f dL_n$. Therefore, the limit $\Lambda(f)$ in (6) exists for all $f \in B(\Sigma, \mathbb{R})$, and by Lemma 2 suffices to show that $\Lambda(\eta 1_{A_j}) \to 0$ for any fixed $\eta > 0$ and every sequence $A_j \in \mathcal{B}_\Sigma$ decreasing monotonically to the empty set. To this end, following the proof of [6, Claim 1] taking $n = k(M+N)$ and reasoning analogously to [6, (2.8) to (2.10)] we get the inequality

$$n^{-1}\Lambda_n(\eta(1_{A_j} - 1)) \le (M+N)^{-1} \log[\exp(\Lambda_M(\eta(1_{A_j} - 1))) + \phi(N)]$$

Since $\Lambda_n(\eta 1_{A_j}) = \Lambda_n(\eta(1_{A_j} - 1)) + n\eta$ it follows that

$$\Lambda(\eta 1_{A_j}) \le \eta + (M+N)^{-1} \log[\exp(\Lambda_M(\eta 1_{A_j}) - M\eta) + \phi(N)] \,.$$

Plainly, $\lim_{j\to\infty} \Lambda_M(\eta 1_{A_j}) = 0$ (for every fixed $\eta > 0$ and $M < \infty$), and hence

$$\lim_{j\to\infty} \Lambda(\eta 1_{A_j}) \le \eta + (M+N)^{-1} \log[\exp(-M\eta) + \phi(N)] \, .$$

This bound holds for every $M, N \in \mathbb{N}$, and it goes to zero by choosing $N, M$ such that $N/M \to 0$ and $M^{-1} \log \phi(N) \to -\infty$ which is possible by the hyper-exponential $\phi-$mixing rate assumed. ∎

# 4 LDP and mixing in the context of Markov chains

We start by specializing the mixing condition (S) to the context of stationary $\Sigma$-valued Markov chains with the transition kernel denoted by $\pi_x(\cdot)$ and using the notations $\pi_x^\ell(D) = P_x(Y_\ell \in D)$ and $(\pi u)(x) = \int_\Sigma \pi_x(dy)u(y)$.

**Lemma 4** *For a stationary Markov chain $\{Y_i\}$, i.e., with the distribution $p$ of $Y_0$ an invariant measure for $\pi$, $(S_-)$ and $(S_+)$ are equivalent to*

$$(S_-): \sup_{D\subset\Sigma} \int_\Sigma \max\{p(D) - e^\ell \pi_x^\ell(D), 0\}\, p(dx) \ \le e^{-Cn}$$

$$(S_+): \sup_{D\subset\Sigma} \int_\Sigma \max\{\pi_x^\ell(D) - e^\ell p(D), 0\}\, p(dx) \ \le e^{-Cn} \, ,$$

*where for every fixed $C < \infty$, $\ell = \ell(n)$ is a non-decreasing sequence satisfying (2).*

**Proof.** Note that by the Markov property, Lemma 1 and stationarity it suffices to take $k_1 = k_2 = 0$ in (S), i.e., to consider $A = \{Y_0 \in F\}$ and $B = \{Y_\ell \in D\}$ for some $D, F \subset \Sigma$. Fixing the set $D$ it is readily checked that the supremum in $(S_-)$ is obtained for $F = \{x : p(D) \ge e^\ell \pi_x^\ell(D)\}$, and in $(S_+)$ for $F = \{x : \pi_x^\ell(D) \ge e^\ell p(D)\}$. ∎

**Remarks:** (a) In particular, all hypercontractive Markov chains satisfy the hypermixing condition (H–2) of [12] (see [13, Theorems 5.5.12 and 5.5.17]), and hence by Proposition 1 these chains also satisfy condition (S).
(b) The results of Theorem 1 apply for a Markov chain with distribution $q$ of $Y_0$ which is not an invariant measure of $\pi$, provided in $(S_-)$ and $(S_+)$ of Lemma 4 we replace $p(D)$ by $q(D)$ and $p(dx)$ by $q\pi^k(dx)$ where the supremum is now over $D \subset \Sigma$ and $k = 1, \ldots, n-1$.

Let $m$ denote the counting measure. It is well known that the LDP holds for the empirical measure of $m$-irreducible finite state Markov chains, with the rate function independent of the distribution $p$ of the initial state $Y_0$ (see [16, Theorem 1]). It is not hard to convince oneself that in this context the LDP holds for *all finite state* chains, but examples with transient states are given in [10, Example 1] [14, Example 4.1] and [20, Example 3.6] in which the rate function either depends on $p$ or is non-convex.

The following lemma summarizes results which apply for every $m$-irreducible countable-valued Markov chain and for any distribution of $Y_0$.

**Lemma 5** *Let $\Sigma$ be a countable set equipped with the discrete topology.*
*(a) The empirical measures $L_n$ of an $m$-irreducible $\Sigma$-valued Markov chain with transition kernel $\pi$ satisfy the weak LDP in $M_1(\Sigma)$ with the convex function*

$$I(\nu) = \sup_{u \in B(\Sigma, \mathbb{R}), u \geq 1} [-\int_\Sigma \log[(\pi u)/u] d\nu] \ , \tag{14}$$

*which does not depend on the distribution $p$ of $Y_0$.*
*(b) If the LDP for $L_n$ holds in $\mathcal{P}_a(\Sigma)$ then its rate function coincides with $I$ on $M_1(\Sigma)$.*
*(c) If the large deviations lower and upper bounds for $L_n$ hold in $M_1(\Sigma)$ for some function $J(\nu)$ then they have to hold for $I(\nu)$ as well.*

**Proof.** Being the supremum of continuous linear functionals of $\nu$, $I(\cdot)$ is convex and lower semicontinuous.
(a) Let $\mu_n$ denote the law of $L_n$ in $M_1(\Sigma)$. For any law $p$ of $Y_0$ and any $m$-irreducible kernel $\pi$ on $\Sigma$, by [10, Theorem 6 and Remark 4] $\{\mu_n\}$ satisfies the large deviations lower bound in $M_1(\Sigma)$ with $I(\nu)$ of (14). For every fixed $u \in B(\Sigma, \mathbb{R})$ and $A \in \mathcal{B}^{cy}$ the proof of [16, (2.3)] is easily adapted to accommodate any transition kernel $\pi$ and any distribution $p$ of $Y_0$ yielding the upper bound

$$\limsup_{n\to\infty} n^{-1} \log \mu_n(A) \leq \inf_{u \in B(\Sigma, \mathbb{R}), u \geq 1} \sup_{\nu \in A} [\int_\Sigma \log[(\pi u)/u] d\nu]$$

The large deviations upper bound with $I(\nu)$ now follows for all pre-compact measurable sets as in [16, proof of (1.7)].
(b) Assuming the LDP for $L_n$ holds in $\mathcal{P}_a(\Sigma)$ with the rate function $J$, by part (a) of the lemma and Remark 1, the weak LDP holds in $M_1(\Sigma)$ for both $I(\nu)$ of (14) and the function $J$ restricted to $M_1(\Sigma)$. Here, the $\tau$-topology of $M_1(\Sigma)$ coincides with the metric topology of weak convergence, hence it follows by [25, Proposition 1.5 and Theorem 2.3] that $I(\nu) = J(\nu)$ for all $\nu \in M_1(\Sigma)$.
(c) Suppose the large deviations lower and upper bounds for $L_n$ hold in $M_1(\Sigma)$ for some function $J(\nu)$. Note that then these bounds also hold for the lower semicontinuous function $J_1(\mu) = \sup_{\{G \text{ open}, \ \mu \in G\}} \inf_{\nu \in G} J(\nu)$. In particular, the weak LDP in $M_1(\Sigma)$ holds for both $J_1$ and $I$ of (14). Consequently, $I = J_1$ by the argument we used in proving part (b), i.e., the large deviations lower and upper bounds must also hold for the function $I$. ■

Recall that a countable-valued Markov chain is a Doeblin chain iff $\inf_{x \in \Sigma} P(Y_{k_0} = x_0 | Y_0 = x) = \rho > 0$ for some $x_0 \in \Sigma$ and $k_0 \in \mathbb{N}$ (see [18, page 192]). The results of [21] imply that all countable-valued, Doeblin $m$-irreducible chains satisfy the LDP in $\mathcal{P}_a(\Sigma)$ for any finitely supported law $p$ of $Y_0$, as summarized in the following proposition.

**Proposition 4** *The empirical measures of countable-valued, $m$-irreducible Markov chain, such that for some $M \in \mathbb{N}$ and $x_0 \in \Sigma$*

$$\inf_{x \in \Sigma} \sum_{k=1}^{M} P(Y_k = x_0 | Y_0 = x) > 0 \ , \tag{15}$$

*satisfy the LDP in* $\mathcal{P}_a(\Sigma)$ *with the same convex rate function, for any law* $p$ *of* $Y_0$ *such that for some* $K = K(p) \in \mathbb{N}$ *and* $y_0 = y_0(p) \in \Sigma$

$$\inf_{x\in\Sigma} \sum_{k=1}^{K} P(Y_k = x | Y_0 = y_0)/p(x) > 0 \,. \tag{16}$$

**Remark:** Condition (16) is a slight extension of requiring $p$ to be an $s$-measure for the Markov chain (see [23, page 15] for definition).

For an $m$-irreducible, aperiodic positively recurrent Doeblin chain the condition $(S_-)$ is equivalent to $\sum_x \max\{p(x) - e^{\ell}\pi^{\ell}_{x_0}(x), 0\} \le e^{-Cn}$ for some non-decreasing $\ell(n)$ satisfying (2). In particular, $(S_-)$ holds whenever (16) holds for the invariant measure.

**Proof.** By Lemma 2 it suffices to prove the LDP with a convex rate function for the empirical means of every $f \in B(\Sigma, \mathbb{R}^d)$. Fixing $f \in B(\Sigma, \mathbb{R}^d)$, since the chain is $m$-irreducible the minorization condition of [21, equation (2.2)] holds. Our assumption (15) implies that $\Sigma$ is an $s$-set in the sense of [21, page 606], and hence by [21, Theorems 1 and 2] the LDP holds for all degenerate initial measures $p = \delta_x$ with the convex rate function independent of $x$. The large deviations lower bound with the same rate function then holds for any initial measure $p$, while the upper bound easily follows from (16). ■

The next example provides an $m$-irreducible, aperiodic positively recurrent Doeblin chain for which the LDP in $\mathcal{P}_a(\Sigma)$ fails to hold when $p$ is the invariant distribution of the chain. Since such Doeblin chains have exponential $\phi$-mixing rate (see [18, page 221] or [26, page 209]), hence at least exponential $\alpha$-mixing rate, this example demonstrates the tightness of the hyper-exponential $\phi$-mixing rate assumed in Proposition 3 (and the necessity of at least hyper-exponential $\alpha$-mixing rate in Proposition 2). In view of Proposition 4, this example also demonstrates the necessity of (16) and the sensitivity of the mere existence of the LDP to the choice of the initial distribution. It is thus related also to large deviations for sequences of mixtures and in particular to [15, Theorem 2.2].

**Example 1** *Consider a stationary countable-valued Markov chain* $Y_n$ *with the state space* $\Sigma = \{(k,j) : k = 0,1,2,\ldots, \quad 1 \le j \le 2n_k\}$*, where parameters are* $n_k = 3^k$ *for* $k \ge 1$ *and* $n_0 = 1/2$*. Transition probabilities are defined by*

$$\pi_{(k,j)}(0,1) = 1 - \pi_{(k,j)}(k, j+1) = \rho \textit{ for } j \ne 2n_k$$

$$\pi_{(k,2n_k)}(0,1) = 1 \textit{ for } k > 0$$

$$\pi_{(0,1)}(0,1) = \rho$$

$$\pi_{(0,1)}(k,1) = q_k, \quad \textit{for } k > 0$$

*where* $\rho \in (0,1)$ *is fixed but arbitrary, and* $q_k = Ce^{-\alpha n_k}$ *with* $\alpha = 1 - 4\log(1-\rho) > 0$ *and* $C = (1-\rho)/\sum_{k=1}^{\infty} e^{-\alpha n_k}$.

*Since* $\inf_{(k,j)\in\Sigma} P(Y_1 = (0,1)|Y_0 = (k,j)) = \rho$ *this* $m$*-irreducible Markov chain is aperiodic Doeblin recurrent and its invariant probability measure is* $p(k,j) = Kq_k(1-\rho)^{j-1}$*, where* $q_0 = 1$ *and* $K = \rho/(1 - \sum_{k=1}^{\infty} q_k(1-\rho)^{2n_k})$.

The following proposition improves over [7, Proposition 4.1], as it applies to the Doeblin chain of Example 1 (instead of a Harris chain in [7]). In particular, Example 1 satisfies also [7, (2.3)], i.e., $p(A) \leq \rho < K$ implies that $\pi_x(A) \leq (1-\rho)$ for all $x \in \Sigma$ and all $A \subset \Sigma$, which *does not result* with the empirical measures satisfying the LDP in $\mathcal{P}_a(\Sigma)$.

**Proposition 5** *Consider the stationary sequence $\{Y_i\}$ corresponding to Example 1 with $Y_0$ distributed according to the invariant measure $p$. The empirical measures of $\{Y_i\}$ fail to satisfy the LDP in $\mathcal{P}_a(\Sigma)$; in particular for the set $\mathcal{D} = \{(k,j) : j \geq n_k + 1\}$ the sequence $\{n^{-1}\sum_{i=1}^n 1_{\mathcal{D}}(Y_i)\}$ does not satisfy the LDP in $\mathbb{R}$.*

**Remark:** Note that the *weak* LDP with the convex function $I$ of (14) does hold in $M_1(\Sigma)$ by part (a) of Lemma 5, and by Proposition 4 the LDP in $\mathcal{P}_a(\Sigma)$ holds for every $p$ which is supported on $\{(k,j) : j \leq K\}$ for some $K \in \mathbb{N}$ with a convex rate function which is independent of $p$.

**Proof.** Let $S_n = 2\sum_{i=1}^n (1_{\mathcal{D}}(Y_i) - 0.5)$. By Varadhan's integral lemma (see [12, Theorem 4.3.1]) it suffices to prove that the sequence $n^{-1}\log E[\exp(3\alpha S_n)]$ does not have a limit as $n \to \infty$. Since for $k \geq 1$

$$E[e^{3\alpha S_{n_k}}] \geq e^{3\alpha n_k} P(Y_j = (k, n_k + j),\ j = 0, \ldots, n_k) = K e^{3\alpha n_k} q_k (1-\rho)^{2n_k - 1}$$

it follows that

$$\limsup_{n\to\infty} n^{-1} \log E[e^{3\alpha S_n}] \geq \limsup_{k\to\infty} {n_k}^{-1} \log E[e^{3\alpha S_{n_k}}] \geq 2\alpha + 2\log(1-\rho)\ . \qquad (17)$$

If $Y_0 \in \mathcal{A}_k = \{(\ell, j) : \ell \leq k\}$ then $T = \inf\{i \geq 0 : Y_i = (0,1)\} \leq 2n_k$ and $S_T \leq n_k$. Since $S_n \leq S_T$ for all $n \geq T$, this and the trivial bound $S_n \leq n$ yield

$$E[e^{3\alpha S_{2n_k}}] \leq e^{3\alpha n_k} + e^{6\alpha n_k} P(Y_0 \notin \mathcal{A}_k)\ .$$

Since $n_{k+1} = 3n_k$ and

$$P(Y_0 \notin \mathcal{A}_k) \leq \frac{K}{\rho} \sum_{\ell=k+1}^{\infty} q_\ell \leq \frac{KC}{\rho(1-e^{-\alpha})} e^{-\alpha n_{k+1}}\ ,$$

it follows that

$$\liminf_{n\to\infty} n^{-1} \log E[e^{3\alpha S_n}] \leq \liminf_{k\to\infty} (2n_k)^{-1} \log E[e^{3\alpha S_{2n_k}}] \leq 1.5\alpha < 2\alpha + 2\log(1-\rho)\ .$$

By (17) this ends the proof. ■

The next example, which is inspired by [1], illustrates the relation between LDP in $\mathcal{P}_a(\Sigma)$ and $M_1(\Sigma)$ under invariant initial measure. It also points out that condition $(S_-)$ alone is not strong enough to imply the LDP in $M_1(\Sigma)$.

**Example 2** *Consider the $\mathbb{Z}_+$-valued Markov chain $Y_k$ with transition probabilities:*

$$\pi_0(k) = \frac{C}{(k+1)^2}\left(1 + \frac{\rho 1_{\{k=1\}}}{(1-\rho)}\right),\quad k = 0, 1, 2, \ldots,$$

$$\pi_k(k+1) = 1 - \pi_k(0) = \rho(\frac{k+1}{k+2})^2, \; k = 1, 2, \ldots$$

*where $\rho \in (0,1)$ and $C = C(\rho) > 0$ is the normalizing constant. This is an m-irreducible, aperiodic Doeblin-recurrent Markov chain and has the finite invariant measure*

$$p(k) = \frac{C}{(2-\rho)(k+1)^2}, k \geq 1, \quad \text{and } p(0) = \frac{1-\rho}{2-\rho}.$$

Plainly, for every $j, k \in \mathbb{Z}_+$

$$P(Y_2 = k | Y_0 = j) \geq \pi_j(0)\pi_0(k) \geq \min\{C, 1-\rho\}\pi_0(k) \geq \min\{C, 1-\rho\}^2 \frac{p(k)}{p(0)}$$

implying that when $Y_0$ is distributed according to the invariant measure, $\psi_-(2) > 0$ and hence $(S_-)$ is satisfied. In particular, by Theorem 1, $\{\mu_n\}$ satisfies the LDP in $\mathcal{P}_a(\mathbb{Z}_+)$ (see also Lemma 2 part (a)).

**Proposition 6** *The empirical measures corresponding to Example 2 satisfy the LDP in $\mathcal{P}_a(\mathbb{Z}_+)$ for any distribution $p$ of $Y_0$ for which $\sup_{x \in \mathbb{Z}_+} x^2 p(x) < \infty$. However, there is no initial measure $q$ for $Y_0$, and no function $J$ for which both the large deviations lower and upper bounds hold in $M_1(\mathbb{Z}_+)$ for these empirical measures.*

**Proof.** For this Markov chain, (16) follows from our assumption on $p$, hence by Proposition 4 the empirical measures satisfy the LDP in $\mathcal{P}_a(\mathbb{Z}_+)$.

By part (c) of Lemma 5, if the large deviations lower and upper bound are to hold in $M_1(\mathbb{Z}_+)$ for any function $J$ then necessarily the upper bound also holds for the function $I(\cdot)$ of (14), and in particular it holds for the closed (actually, discrete) set

$$\mathcal{C} = \{\nu : \nu(j+1) = \ldots = \nu(j+m) = m^{-1} \qquad m \in \mathbb{N}, j \in \mathbb{Z}_+\} .$$

Note that if $\nu(j+1) = \ldots = \nu(j+m) = m^{-1}$ for some $j \geq 0$, $m \geq 1$, then taking all $u(k) = 1$ except $u(j+1) = \eta$, we get

$$I(\nu) = m^{-1} \sup_{u(\cdot)>0} \sum_{k=j+1}^{j+m} \log(\frac{u(k)}{\rho(\frac{k+1}{k+2})^2 u(k+1) + (1-\rho(\frac{k+1}{k+2})^2)u(0)}) \geq m^{-1} \log \eta .$$

Consequently, $\inf_{\nu \in \mathcal{C}} I(\nu) = \infty$. Since for every $j \in \mathbb{Z}_+$

$$P(L_n \in \mathcal{C} | Y_0 = j) \geq P(Y_1 = j+1, \ldots, Y_n = j+n | Y_0 = j) \geq \min\{\rho, C\}\rho^{n-1}(\frac{j+1}{n+j+1})^2$$

it follows that for any initial distribution $q$ we have

$$\mu_n(\mathcal{C}) = \sum_{j=0}^{\infty} q(j) P(L_n \in \mathcal{C} | Y_0 = j) \geq \min\{\rho, C\}(n+1)^{-2}\rho^{n-1}$$

and hence $\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\mathcal{C}) \geq \log \rho$, in contradiction with the large deviations upper bound for the set $\mathcal{C}$. ■

## References


[1] Baxter, J. R., Jain, N.C. and Varadhan, S. R. S. Some familiar examples for which the large deviation principle does not hold. Commun. Pure Appl. Math. 34, (1991) 911–923.

[2] Bolthausen, E. and Schmock, U. On the maximum entropy principle for uniformly ergodic Markov chains. Stochastic Processes Appl. 33, (1989) 1–27.

[3] Bradley, R. C. Basic properties of strong mixing conditions. In E. Eberlein and M. Taqqu, editors, *Dependence in Probability and Statistics*, Birkhäuser, Basel, Switzerland, (1986) 165–192.

[4] Bradley, R. C., Bryc, W. and Janson, S. On dominations between measures of dependence, J. Multivar. Anal. 23, (1987) 312–329.

[5] Bradley, R. C. and Peligrad, M. Invariance principles under a two-part mixing assumption, Stochastic Processes Appl. 22, (1986) 271–289.

[6] Bryc, W. On large deviations for uniformly strong mixing sequences, Stochastic Processes Appl. 41, (1992) 191–202.

[7] Bryc, W. and Smolenski, W. On the convergence of averages of mixing sequences, J. Theor. Probab. 6 (1993) 473–483.

[8] Chiyonobu, T. and Kusuoka, S. The large deviation principle for hypermixing processes. Probab. Theory Relat. Fields 78, (1988) 627–649.

[9] Dawson, D. and Gärtner, J. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. Stochastics. 20, (1987) 247–308.

[10] De Acosta, A. Large deviations for empirical measures of Markov chains. J. Theor. Probab. 3, (1990) 395–431.

[11] De Acosta, A. On large deviations of empirical measures in $\tau$ topology. To appear in the special issue of Journal of Applied Probability in honor of L. Takacs.

[12] Dembo, A. and Zeitouni, O. *Large Deviations Techniques and Applications.* Jones and Bartlett, Boston, MA, 1993.

[13] Deuschel, J. D. and Stroock, D. W. *Large Deviations*, Academic Press, Boston, MA, 1989.

[14] Dinwoodie, I. H. Identifying a large deviation rate function. Ann. Probab. 21, (1993) 216–231.

[15] Dinwoodie, I. H. and Zabell, S. L. Large deviations for exchangeable random vectors. Ann. Probab. 20, (1992) 1147–1166.

[16] Donsker, M. D. and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time, I. Commun. Pure Appl. Math. 28, (1975) 1–47.

[17] Donsker, M. D. and Varadhan, S. R. S. Asymptotic evaluation of certain Markov process expectations for large time, III. Commun. Pure Appl. Math. 29, (1976) 389–461.

[18] Doob, J. L. *Stochastic Processes*, Wiley. New-York, NY, 1953.

[19] Hammersley, J. M. Generalization of the fundamental theorem on subadditive functions, Math. Proc. Camb. Philos. Soc. 58, (1962) 235–238.

[20] Jain, N. C. Large deviation lower bounds for additive functionals of Markov processes. Ann. Probab. 18, (1990) 1071–1098.

[21] Ney, P. and Nummelin, E. Markov additive processes II: large deviations. Ann. Probab. 15, (1987) 593–609.

[22] Hu, Y. J. Large deviations for stationary $\phi$-mixing sequences in $\tau$-topology, preprint 1993.

[23] Nummelin, E. General irreducible Markov chains and non-negative operators. Cambridge Tracts in Mathematics, 83. Cambridge University Press, 1984.

[24] Nummelin, E. Large deviations for functionals of stationary processes. Probab. Theory Relat. Fields 86, (1990) 387–401.

[25] O'Brien, G. L. Sequences of capacities, with connections to large-deviation theory. Preprint 1993.

[26] Rosenblatt, M. *Markov Processes, Structure and Asymptotic Behavior*. Springer-Verlag, Berlin, 1971.