EURANDOM PREPRINT SERIES
2013-007
March 18, 2013

**Large deviation principles for words drawn from correlated letter sequences**

F. den Hollander, J. Poisat
ISSN 1389-2355

# LARGE DEVIATION PRINCIPLES FOR WORDS DRAWN FROM CORRELATED LETTER SEQUENCES

F. DEN HOLLANDER AND J. POISAT

Abstract. When an i.i.d. sequence of letters is cut into words according to i.i.d. renewal times, an i.i.d. sequence of words is obtained. In the *annealed* LDP (large deviation principle) for the empirical process of words, the rate function is the specific relative entropy of the observed law of words w.r.t. the reference law of words. In Birkner, Greven and den Hollander [3] the *quenched* LDP (= conditional on a typical letter sequence) was derived for the case where the renewal times have an *algebraic* tail. The rate function turned out to be a sum of two terms, one being the annealed rate function, the other being proportional to the specific relative entropy of the observed law of letters w.r.t. the reference law of letters, obtained by concatenating the words and randomising the location of the origin. The proportionality constant equals the tail exponent of the renewal process.

The purpose of the present paper is to extend both LDP's to letter sequences that are not i.i.d. It is shown that both LDP's carry over when the letter sequence satisfies a mixing condition called *summable variation*. The rate functions are again given by specific relative entropies w.r.t. the reference law of words, respectively, letters. But since neither of these reference laws is i.i.d., several approximation arguments are needed to obtain the extension.

*MSC* 2010: 60F10, 60G10.
*Key words:* Letters and words, renewal times, empirical process, annealed vs. quenched large deviation principle, rate function, specific relative entropy, mixing.
*Acknowledgment:* FdH and JP were supported by ERC Advanced Grant 267356 VARIS.

## 1. Introduction and main results

1.1. **Notation.** Let $E$ be a finite set of *letters* and $\widetilde{E} = \cup_{\ell\in\mathbb{N}} E^\ell$ the set of finite *words* drawn from $E$. Write $E^\mathbb{Z}$ and $\widetilde{E}^\mathbb{Z}$ for the sets of two-sided sequences of letters and words, and let $\theta$ and $\widetilde{\theta}$ denote the left-shifts acting on these sets, respectively. The set of probability laws on $E^\mathbb{Z}$ and $\widetilde{E}^\mathbb{Z}$ that are shift-invariant, respectively, shift-invariant and ergodic w.r.t. $\theta$ and $\widetilde{\theta}$ are denoted by $\mathcal{P}^{\mathrm{inv}}(E^\mathbb{Z})$ and $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^\mathbb{Z})$, respectively, $\mathcal{P}^{\mathrm{inv,erg}}(E^\mathbb{Z})$ and $\mathcal{P}^{\mathrm{inv,erg}}(\widetilde{E}^\mathbb{Z})$, and are endowed with the topology of weak convergence.

Let $X = (X_k)_{k\in\mathbb{Z}}$ be a two-sided *random sequence of letters* sampled according to a shift-invariant probability distribution $\nu$ on $E^\mathbb{Z}$. Let $\tau = (\tau_i)_{i\in\mathbb{Z}}$ be a two-sided i.i.d. sequence of *renewal times* drawn from a common probability law $\varrho$ on $\mathbb{N}$, independent of $X$. The latter form a renewal process $T = (T_i)_{i\in\mathbb{Z}}$ given by

$$T_0 = 0, \qquad T_i = T_{i-1} + \tau_i, \quad i \in \mathbb{Z}. \tag{1.1}$$

Let $Y = (Y_i)_{i\in\mathbb{Z}}$ be the two-sided *random sequence of words* cut out from $X$ according to $\tau$, i.e.,

$$Y_i = X_{(T_{i-1},T_i]} = (X_{T_{i-1}+1}, \dots, X_{T_i}), \qquad i \in \mathbb{Z}. \tag{1.2}$$

The joint law of $X$ and $\tau$ is denoted by P. Write $|Y_i|$ to denote the length of word $i$.

The reverse of cutting is glueing. The *concatenation operator* $\kappa$: $\widetilde{E}^\mathbb{Z} \to E^\mathbb{Z}$ glues a word sequence into a letter sequence. In particular, $\kappa(Y) = X$. Given $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^\mathbb{Z})$ with $m_Q =$

*Date*: March 18, 2013.

$E_Q(|Y_1|) < \infty$, let $\Psi_Q \in \mathcal{P}^{\mathrm{inv}}(E^{\mathbb{Z}})$ be defined by

$$\Psi_Q(A) = \frac{1}{m_Q} E_Q\left(\sum_{k=0}^{|Y_1|-1} 1_{\{\theta^k \kappa(Y) \in A\}}\right), \qquad A \subset E^{\mathbb{Z}}, \tag{1.3}$$

i.e., the law of $\kappa(Y)$ when $Y$ is drawn from $Q$, turned into a stationary law by randomizing the location of the origin.

For $n \in \mathbb{N}$, let $(Y_{(0,n]})^{\mathrm{per}} \in \widetilde{E}^{\mathbb{Z}}$ denote the $n$-periodized version of $Y$. We are interested in the *empirical distribution of words*

$$R_n = \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\widetilde{\theta}^i (Y_{(0,n]})^{\mathrm{per}}}, \tag{1.4}$$

both under P (= annealed law) and under $\mathrm{P}(\cdot \mid X)$ for $\nu$-a.a. $X$ (= quenched law).

1.2. **Large deviation principles.** If $\nu$ is i.i.d., then P is i.i.d. and the annealed LDP is standard, with the rate function given by the specific relative entropy of the observed law of words w.r.t. P. The quenched LDP, however, is not standard. The quenched LDP was obtained in Birkner [2] for the case where $\varrho$ has an exponentially bounded tail, and in Birkner, Greven and den Hollander [3] for the case where $\varrho$ has a polynomially decaying tail:

$$\lim_{\substack{m\to\infty \\ \varrho(m)>0}} \frac{\log \varrho(m)}{\log m} = -\alpha, \qquad \alpha \in [1,\infty). \tag{1.5}$$

(No condition on the support of $\varrho$ is needed other than that it is infinite.) In the latter case, the quenched rate function turns out to be a sum of two terms, one being the annealed rate function, the other being proportional to the specific relative entropy of the observed law of letters w.r.t. $\nu$, obtained by concatenating the words and randomising the location of the origin. The proportionality constant equals $\alpha - 1$ times the average word length.

The goal of the present paper is to extend both LDP's to the situation where $\nu$ is no longer i.i.d., but satisfies a mixing condition called *summable variation*, which will be defined in Section 3. In what follows, $H(\cdot \mid \cdot)$ denotes specific relative entropy (see Dembo and Zeitouni [4], Chapter 6, for the definition and key properties).

**Theorem 1.1** (**Annealed LDP**). *If $\nu$ has summable variation, then the family of probability laws $\mathrm{P}(R_n \in \cdot)$, $n \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})$ with rate $n$ and with rate function $I^{\mathrm{ann}}\colon \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}}) \mapsto [0,\infty]$ given by the specific relative entropy*

$$I^{\mathrm{ann}}(Q) = H(Q \mid \mathrm{P}). \tag{1.6}$$

*$I^{\mathrm{ann}}$ is lower semi-continuous, has compact level sets, is affine, and has a unique zero at $Q = \mathrm{P}$.*

**Theorem 1.2** (**Quenched LDP**). *If $\nu$ has summable variation, then for $\nu$-a.a. $X$ the family of conditional probability laws $\mathrm{P}(R_n \in \cdot \mid X)$, $n \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})$ with rate $n$ and with rate function $I^{\mathrm{que}}\colon \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}}) \mapsto [0,\infty]$ given by the sum of specific relative entropies*

$$I^{\mathrm{que}}(Q) = H(Q \mid \mathrm{P}) + (\alpha - 1) m_Q H(\Psi_Q \mid \nu). \tag{1.7}$$

*$I^{\mathrm{que}}$ is lower semi-continuous, has compact level sets, is affine, and has a unique zero at $Q = \mathrm{P}$.*

**Theorem 1.3.** *Both LDPs remain valid when $E$ is a Polish space.*

**Remark:** If $m_Q = \infty$, then the second term in (1.7) is defined to be $\alpha - 1$ times the *truncation limit* $\lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid [\nu]_{\mathrm{tr}})$, where tr is the operator that truncates all the words to length $\leq$ tr. See Birkner, Greven and den Hollander [3] for details.

**Remark:** Both rate functions are the same as for the i.i.d. case, even though the reference laws $P$ and $\nu$ are no longer i.i.d. This lack of independence will require us to go through several approximation arguments. Both LDP's can be applied to the problem of pinning of a polymer chain at an interface carrying correlated disorder. This application, which is our main motivation for extending the LDP's, will be discussed in a future paper.

1.3. **Outline.** In Section 2 we collect some basic facts, introduce the relevant mixing coefficients, and define summable variation. We give examples where this mixing condition holds, respectively, fails. In Section 3 we prove the annealed LDP by applying a result from Orey and Pelikan [14]. In Section 4 we prove the quenched LDP by going over the proof in Birkner, Greven and den Hollander [3] for i.i.d. letter sequences and checking which parts have to be adapted. In Section 5 we extend the LDP's from finite $E$ to Polish $E$ by using the Dawson-Gärtner projective limit LDP.

## 2. Basic facts, mixing coefficients and summable variation

2.1. **Basic facts.** Throughout the paper we abbreviate

$$X_{(m,n]} = (X_{m+1}, \dots, X_n), \quad Y_{(m,n]} = (Y_{m+1}, \dots, Y_n), \qquad -\infty \le m \le n \le \infty. \tag{2.1}$$

The associated sigma-algebra's are written as

$$\mathcal{F}_{(m,n]} = \sigma(X_{(m,n]}), \qquad \mathcal{G}_{(m,n]} = \sigma(Y_{(m,n]}). \tag{2.2}$$

Since $X$ is no longer i.i.d., the distribution of a word in $Y$ depends on the outcome of all the previous words. However, since the word lengths are still i.i.d., when we condition on the past of the word sequence only the past of the letter sequence is relevant, as is stated in the next lemma.

**Lemma 2.1.** $\mathrm{P}(A \mid \mathcal{G}_{(-\infty,0]}) = \mathrm{P}(A \mid \mathcal{F}_{(-\infty,0]})$ *a.s. for all* $A \in \mathcal{G}_{(0,\infty)}$.

*Proof.* Fix $r \in \mathbb{N}$ and $y_1, \dots, y_r \in \widetilde{E}$, and pick $A = \{Y_{(0,r]} = y_{(0,r]}\}$. Write

$$\mathrm{P}(A \mid \mathcal{G}_{(-\infty,0]}) = \mathrm{P}\Big(T_{(0,r]} = |y|_{(0,r]},\, X_{\left(0, \sum_{i=1}^r |y_i|\right]} = \kappa(y_{(0,r]}) \mid \mathcal{G}_{(-\infty,0]}\Big), \tag{2.3}$$

where $|y_i|$ is the length of word $y_i$. Since $\sigma(\tau_{(0,r]})$ is independent of $\mathcal{G}_{(-\infty,0]}$, we have

$$\mathrm{P}(A \mid \mathcal{G}_{(-\infty,0]}) = \mathrm{P}\Big(X_{\left(0, \sum_{i=1}^r |y_i|\right]} = \kappa(y_{(0,r]}) \mid \mathcal{G}_{(-\infty,0]}\Big) \prod_{i=1}^r \varrho(|y_i|). \tag{2.4}$$

But $X$ and $\tau$ are independent as well, and so

$$\mathrm{P}(A \mid \mathcal{G}_{(-\infty,0]}) = \nu\Big(X_{\left(0, \sum_{i=1}^r |y_i|\right]} = \kappa(y_{(0,r]}) \mid \mathcal{F}_{(-\infty,0]}\Big) \prod_{i=1}^r \varrho(|y_i|), \tag{2.5}$$

which yields the claim after we argue backwards. □

Write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. Let $(\nu_{x^-}(\cdot); x^- \in E^{-\mathbb{N}_0})$ be a regular version of $\nu(\cdot \mid X_{(-\infty,0]})$, i.e.,

$$\nu(A) = \int_{x^- \in E^{-\mathbb{N}_0}} \nu_{x^-}(A)\, \mathrm{d}\nu(x^-), \qquad A \in \mathcal{F}_{(0,\infty)}. \tag{2.6}$$

From the regular conditional probabilities of $\nu$ we obtain regular conditional probabilities of P as follows.

**Lemma 2.2.** *The collection* $(\mathrm{P}_{y^-}(\cdot), y^- \in \widetilde{E}^{-\mathbb{N}_0})$ *of probability laws on* $\widetilde{E}^{\mathbb{N}}$ *defined by*

$$\mathrm{P}_{y^-}(A) = \int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\, \mathrm{d}\nu_{\kappa(y^-)} \qquad \forall\, A \in \mathcal{G}_{(0,\infty)}, \tag{2.7}$$

*constitute a regular version of the conditional probability* $\mathrm{P}(\cdot \mid \mathcal{G}_{(-\infty,0]})$.

*Proof.* For every $y^- \in \widetilde{E}^{-\mathbb{N}_0}$, $\mathrm{P}_{y^-}(\cdot)$ defined in (2.7) is a probability measure. It therefore is enough to prove (2.7) for cylinder sets. Let $m \in \mathbb{N}$, $(y_i)_{1\le i\le m} \in \widetilde{E}^m$ and $A = \bigcap_{1\le i\le m}\{Y^{(i)} = y_i\}$. Then

$$\begin{aligned}\int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\,\mathrm{d}\nu_{\kappa(y^-)} &= \int_{E^{\mathbb{Z}}} \mathrm{d}\nu_{\kappa(y^-)}\, 1_{\{X\in\kappa(A)\}} \prod_{1\le i\le m} \varrho(|y_i|)\\ &= \nu_{\kappa(y^-)}(X \in \kappa(A)) \prod_{1\le i\le m} \varrho(|y_i|).\end{aligned} \tag{2.8}$$

Since $\int_{\widetilde{E}^{-\mathbb{N}_0}} \mathrm{dP}(y^-)\,\nu_{\kappa(y^-)}(\cdot) = \int_{E^{-\mathbb{N}_0}} \mathrm{d}\nu(x^-)\,\nu_{x^-}(\cdot) = \nu(\cdot)$, we have

$$\int_{\widetilde{E}^{-\mathbb{N}_0}} \mathrm{dP}(y^-) \int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\,\mathrm{d}\nu_{\kappa(y^-)} = \nu(X \in \kappa(A)) \prod_{1\le i\le m} \varrho(|y_i|) = \mathrm{P}(A), \tag{2.9}$$

which proves the claim. □

2.2. **Mixing coefficients.** We need the following mixing coefficients for letters and words:

**Definition 2.3.** *(a) For $\Lambda_1 \subset -\mathbb{N}_0$ and $\Lambda_2 \subset \mathbb{N}$, let*

$$\varphi(\Lambda_1,\Lambda_2) = \sup_{\substack{x^-,\hat{x}^- \in E^{-\mathbb{N}_0}\\ (x^-)_{\Lambda_1} = (\hat{x}^-)_{\Lambda_1}}} \ \sup_{\substack{A\in\mathcal{F}_{\Lambda_2}:\\ \nu_{x^-}(A)>0}} \left|\log \nu_{x^-}(A) - \log \nu_{\hat{x}^-}(A)\right|. \tag{2.10}$$

*(b) For $\Lambda \subset \mathbb{N}$, let*

$$\psi(\Lambda) = \sup_{y^-,\hat{y}^- \in \widetilde{E}^{-\mathbb{N}_0}} \ \sup_{\substack{A\in\mathcal{G}_\Lambda\\ P_{y^-}(A)>0}} \left|\log \mathrm{P}_{y^-}(A) - \log \mathrm{P}_{\hat{y}^-}(A)\right|. \tag{2.11}$$

The restrictions $\nu_{x^-}(A) > 0$ and $\mathrm{P}_{\hat{y}^-}(A) > 0$ are put in to avoid $\infty - \infty$. Nonetheless, (2.10) and (2.11) may be infinite. Note that if $\Lambda_1 = \emptyset$, then the supremum in Definition 2.3(a) is taken over all $x^-, \hat{x}^- \in E^{-\mathbb{N}_0}$ without any restriction ($(x^-)_\Lambda$ denotes the restriction of $x^-$ to $\Lambda$). We will use the following abbreviations:

$$\varphi(k,\cdot) = \varphi((-k,0],\cdot),\quad k\in\mathbb{N},\qquad \varphi(0,\cdot) = \varphi(\emptyset,\cdot),\qquad \varphi(\cdot,\ell) = \varphi(\cdot,(0,\ell]),\quad \ell\in\mathbb{N}. \tag{2.12}$$

**Lemma 2.4.** *Let $0 \le m < n$, $y_{(m,n]} \in \widetilde{E}^{n-m}$ and $A = \{Y_{(m,n]} = y_{(m,n]}\}$. For all $y^-, \hat{y}^- \in \widetilde{E}^{-\mathbb{N}_0}$,*

$$\mathrm{P}_{y^-}(A) \le \mathrm{E}\left[\exp\left\{\varphi\Big(0, \Big(T_m, T_m + \sum_{k=m+1}^n |y_k|\Big]\Big)\right\} \mathrm{P}_{\hat{y}^-}(A \mid T_m)\right]. \tag{2.13}$$

*Proof.* Using Definition 2.3(a), we have

$$\begin{aligned}\mathrm{P}_{y^-}(A) &= \mathrm{E}\left[\nu_{\kappa(y^-)}\Big(X_{\left(T_m, T_m+\sum_{k=m+1}^n |y_k|\right]} = \kappa(y_{(m,n]})\Big)\right]\\ &\le \mathrm{E}\left[\exp\left\{\varphi\left(0, \Big(T_m, T_m + \sum_{k=m+1}^n |y_k|\Big]\right)\right\} \nu_{\kappa(\hat{y}^-)}\Big(X_{\left(T_m, T_m+\sum_{k=m+1}^n |y_k|\right]} = \kappa(y_{(m,n]})\Big)\right]\\ &= \mathrm{E}\left[\exp\left\{\varphi\left(0, \Big(T_m, T_m + \sum_{k=m+1}^n |y_k|\Big]\right)\right\} \mathrm{P}_{\hat{y}^-}(A \mid T_m)\right].\end{aligned} \tag{2.14}$$

□

**Lemma 2.5.** *For all $k \in \mathbb{N}_0, \ell \in \mathbb{N}$,*

$$\varphi(k,\ell) \le \sum_{m=0}^{\ell-1} \varphi(k+m), \tag{2.15}$$

*where $\varphi(k) = \varphi(k,1)$, $k \in \mathbb{N}_0$.*

*Proof.* We show that, for all $m \in \mathbb{N}_0$ and $k, \ell \in \mathbb{N}$,

$$\varphi(m, k+\ell) \le \varphi(m,k) + \varphi(m+k,\ell), \tag{2.16}$$

which yields the claim via iteration. To prove (2.16), pick $x_{(0,k+\ell]} \in E^{k+\ell}$ and $x^-, \hat{x}^- \in E^{-\mathbb{N}_0}$ with $(x^-)_{[-m,0]} = (\hat{x})_{[-m,0]}$, and consider the events

$$A_{(0,k+\ell]} = \{X_{(0,k+\ell]} = x_{(0,k+\ell]}\}, \quad A_{(0,k]} = \{X_{(0,k]} = x_{(0,k]}\}, \quad A_{(k,k+\ell]} = \{X_{(k,k+\ell]} = x_{(k,k+\ell]}\}. \tag{2.17}$$

Estimate

$$\begin{aligned} \nu_{x^-}(A_{(0,k+\ell]}) &= \nu_{x^-}(A_{(0,k]})\, \nu_{x^- x_{(0,k]}}(A_{(k,k+\ell]}) \\ &\le e^{\varphi(m,k)}\, \nu_{\hat{x}^-}(A_{(0,k]})\, e^{\varphi(m+k,\ell)}\, \nu_{\hat{x}^- x_{(0,k]}}(A_{(k,k+\ell]}) \\ &= e^{\varphi(m,k)+\varphi(m+k,\ell)}\, \nu_{\hat{x}^-}(A_{(0,k+\ell]}), \end{aligned} \tag{2.18}$$

where $\hat{x}^- x_{(0,k]}$ is the concatenation of $\hat{x}^-$ and $x_{(0,k]}$. Insert this estimate into (2.3) and take the supremum over $x_{(0,k+\ell]}$ and $x^-, \hat{x}^-$ to get (2.16). □

Note that $k \mapsto \varphi(k)$ is non-increasing on $\mathbb{N}_0$.

2.3. **Summable variation.** The key mixing condition in our LDP's is *summable variation*:

$$\text{(SV)} \qquad \sum_{n \in \mathbb{N}_0} \varphi(n) < \infty. \tag{2.19}$$

The term summable variation is borrowed from the theory of Gibbs measures, where logarithms of probabilities play the role of interaction potentials, and coefficients similar to our $\varphi(n)$'s are used to measure the absolute summability of these interaction potentials.

(I) Random processes (with finite alphabet) that satisfy (SV) include i.i.d. processes ($\varphi(n) = 0$ for all $n \in \mathbb{N}_0$), Markov chains of order $m$ ($\varphi(0) < \infty$ and $\varphi(n) = 0$ for all $n \ge m$), and chains with complete connections whose one-letter forward conditional probabilities have summable variation. Ledrappier [12, Example 2, Proposition 4] shows that such chains have a unique invariant measure and are Weak Bernoulli under (SV). Berbee [1, Theorem 1.1] shows that they have a unique invariant measure and are Bernoulli when $\sum_{n \in \mathbb{N}} \exp[-\sum_{m=1}^n \varphi(m)] = \infty$, a condition slightly weaker than (SV). (Uniqueness of the invariant measure has been proved more recently by Johansson and Öberg [10] and by Johansson, Öberg and Pollicott [11] under the even weaker condition $\sum_{n \in \mathbb{N}} \varphi(n)^2 < \infty$.) Yet other examples satisfying (SV) include Ising spins labeled by $\mathbb{Z}$ with a ferromagnetic pair potential that has a sufficiently thin tail.

(IIa) A class of random processes that fail to satisfy (SV) is the following. Let $E = \{0,1\}$, and let $p$ be any probability law on $\mathbb{N}$ such that $p(\ell) \sim C\ell^{-\gamma}$ for some $\gamma > 2$. Since $\sum_{\ell \in \mathbb{N}} \ell p(\ell) < \infty$, there exists a stationary Markov chain $(A_k)_{k \in \mathbb{Z}}$ on $\mathbb{N}_0$ with the following transition probabilities:

$$\mathrm{P}(A_1 = n+1 \mid A_0 = n) = \frac{\sum_{\ell > n+1} p(\ell)}{\sum_{\ell > n} p(\ell)}, \quad \mathrm{P}(A_1 = 0 \mid A_0 = n) = \frac{p(n+1)}{\sum_{\ell > n} p(\ell)}, \quad n \in \mathbb{N}_0. \tag{2.20}$$

The process $(X_k)_{k\in\mathbb{Z}}$ defined by $X_k = 1_{\{A_k=0\}}$ fails to satisfy (SV). Indeed, pick $n \in \mathbb{N}$ and $x, x[n] \in E^{-\mathbb{N}_0}$ be such that $x_i = 1$ for $i \in -\mathbb{N}_0$, $x[n]_i = 0$ for $i \in (-n, 0]$ and $x[n]_i = 1$ for $i \in (-\infty, -n]$. Then

$$\varphi(1) \geq \log \nu_x(X_1 = 1) - \log \nu_{x[n]}(X_1 = 1) = \log p(1) - \log\left(\frac{p(n+1)}{\sum_{\ell>n} p(\ell)}\right). \tag{2.21}$$

Since this lower bound holds for all $n \in \mathbb{N}$, we conclude by letting $n \to \infty$ that $\varphi(1) = \infty$.

(IIb) Another class of random processes that fail to satisfy (SV) is random walk in random scenery. Let $S = (S_n)_{n\in\mathbb{Z}}$ be a simple random walk on $\mathbb{Z}^d$, $d \geq 1$, i.e., $S_0 = 0$ and $S_n - S_{n-1} = X_n$ with $(X_n)_{n\in\mathbb{Z}}$ i.i.d. random variables uniformly distributed on $\{e \in \mathbb{Z}^d\colon \|e\| = 1\}$. Let $\xi = (\xi(x))_{x\in\mathbb{Z}^d}$ be i.i.d. random variables taking the values 0 and 1 with probability $\frac{1}{2}$ each, and define $Z_n = (X_n, \xi(S_n))$. Then $Z = (Z_n)_{n\in\mathbb{Z}}$ is stationary and ergodic, but not i.i.d. In den Hollander and Steif [9, Theorems 2.4 and 2.5] it is shown that $Z$ is *Weak Bernoulli* if and only if $d \geq 5$. Since (SV) implies Weak Bernoulli (Ledrappier [12, Proposition 4]), $Z$ does not satisfy (SV) when $1 \leq d \leq 4$.

## 3. Annealed LDP

The annealed LDP in Theorem 1.1 is a process-level LDP. Such LDP's were proven by Donsker and Varadhan [6, 7] for reference processes that are Markov or Gaussian. Orey [13] and Orey and Pelikan [14] gave a proof for *ratio-mixing* processes (see below), using the observation that any random process can be viewed as a Markov process by keeping track of its past.

**Proposition 3.1.** (Orey and Pelikan [14, Theorem 2.1]) *Suppose that* P *has the following ratio-mixing property:*

(RM) *There exists a non-decreasing function $n \mapsto m(n)$ such that*

$$0 \leq m(n) < n, \quad \lim_{n\to\infty} m(n)/n = 0, \quad \lim_{n\to\infty} \psi((m(n), n])/n = 0. \tag{3.1}$$

*Then the family of probability laws $\mathrm{P}(R_n \in \cdot)$, $n \in \mathbb{N}$, satisfies the LDP on $\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})$ with rate $n$ and with rate function given by the specific relative entropy*

$$Q \mapsto H(Q \mid \mathrm{P}) = \int_{y^-\in\widetilde{E}^{-\mathbb{N}_0}} Q(\mathrm{d}y^-) \int_{y\in\widetilde{E}} Q_{y^-}|_1(\mathrm{d}y) \log\left(\frac{dQ_{y^-}|_1}{d\mathrm{P}_{y^-}|_1}(y)\right). \tag{3.2}$$

The specific relative entropy $H(Q \mid \mathrm{P})$ is defined to be infinite when $Q_{y^-}|_1 \ll \mathrm{P}_{y^-}|_1$ fails on a set of $y^-$'s with a strictly positive $Q$-measure. An alternative form of (3.2) is

$$H(Q \mid \mathrm{P}) = \int_{y^-\in\widetilde{E}^{-\mathbb{N}_0}} Q(\mathrm{d}y^-)\, H\big(Q_{y^-}(Y_1 \in \cdot\,) \mid \mathrm{P}_{y^-}(Y_1 \in \cdot\,)\big). \tag{3.3}$$

The latter can be viewed as the specific relative entropy of the laws of two Markov processes, namely, the laws of the *past processes* $Y^* = (Y^{(n),*})_{n\in\mathbb{N}}$ with $Y^{(n),*} = (Y^{(n-m)})_{m\in\mathbb{N}}$, $n \in \mathbb{N}$, when $Y$ is distributed according to $Q$, respectively, $P$. The regular conditional probability laws $(\mathrm{P}_{y^-}(Y_1 \in \cdot\,), y^- \in \widetilde{E}^{-\mathbb{N}_0})$ play the role of transition probabilities for $Y^*$, and regularity translates into the Feller property.

We are now ready to prove Theorem 1.1.

*Proof.* From Lemma 2.4 and the fact that $\ell \mapsto \varphi(0, \ell)$ is non-decreasing, we get $P_{y^-}(A) \leq e^{\varphi(0,\infty)} P_{\hat{y}^-}(A)$. Hence Definition 2.3(b) gives $\psi((m, n]) \leq \varphi(0, \infty)$ for all $0 \leq m < n$. From Lemma 2.5 we get

$$\varphi(0, \infty) \leq \sum_{n\in\mathbb{N}_0} \varphi(n). \tag{3.4}$$

Hence, if (SV) holds, then (RM) holds for $m(n) = 0$, and so we can apply Proposition 3.1. $\square$

## 4. Quenched LDP

In Sections 4.1–4.3 we prove several lemmas that are needed in Section 4.4 to give the proof of Theorem 1.2. This proof is an extension of the proof in [3] for i.i.d. $\nu$. We focus on those ingredients where the lack of independence of $\nu$ requires modifications.

### 4.1. **Decoupling inequalities.** Abbreviate

$$C(\varphi) = \exp\left[\sum_{n\in\mathbb{N}_0} \varphi(n)\right] < \infty. \tag{4.1}$$

**Lemma 4.1.** *For all $x^-$, $\hat{x}^- \in E^{-\mathbb{N}_0}$, $A \in \mathcal{F}_{(0,\infty)}$ and $n \in \mathbb{N}$,*

$$C(\varphi)^{-1}\nu_{\hat{x}^-}(A) \le \nu_{x^-}(A) \le C(\varphi)\nu_{\hat{x}^-}(A), \tag{4.2}$$

$$C(\varphi)^{-1}\nu_{\hat{x}^-}(A) \le \nu\big(A \mid X_{(-n,0]} = x^-_{(-n,0]}\big) \le C(\varphi)\nu_{\hat{x}^-}(A). \tag{4.3}$$

*Proof.* To prove (4.2), pick $k \in \mathbb{N}$ and $A \in \mathcal{F}_{(0,k)}$. If $\nu_{\hat{x}^-}(A) = 0$ then $\nu_{x^-}(A) = 0$ as well because $\varphi(k) < \infty$ and there is nothing to prove, so we can assume $\nu_{\hat{x}^-}(A) > 0$. Then, by the definition of $\varphi(k)$ and Lemma 2.5,

$$e^{-C(\varphi)} \le e^{-\varphi(0,k)} \le \frac{\nu_{x^-}(A)}{\nu_{\hat{x}^-}(A)} \le e^{\varphi(0,k)} \le e^{C(\varphi)}. \tag{4.4}$$

To prove (4.3), write

$$\begin{aligned}
\nu\big(A \mid X_{(-n,0]} = x^-_{(-n,0]}\big) &= \frac{\nu(\{X_{(-n,0]} = x^-_{(-n,0]}\} \cap A)}{\nu(X_{(-n,0]} = x^-_{(-n,0]})}\\
&= \frac{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(\{X_{(0,n]} = x^-_{(-n,0]}\} \cap \theta^{-n}A)}{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(X_{(0,n]} = x^-_{(-n,0]})}\\
&= \frac{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(X_{(0,n]} = x^-_{(-n,0]})\nu_{\widetilde{x}^-,x^-_{(-n,0]}}(A)}{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(X_{(0,n]} = x^-_{(-n,0]})}\\
&\le \frac{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(X_{(0,n]} = x^-_{(-n,0]})\, e^{C(\varphi)}\, \nu_{\hat{x}^-}(A)}{\int_{\widetilde{x}^-\in \mathrm{E}^{-\mathbb{N}_0}} d\nu(\widetilde{x}^-)\, \nu_{\widetilde{x}^-}(X_{(0,n]} = x^-_{(-n,0]})}\\
&= e^{C(\varphi)}\, \nu_{\hat{x}^-}(A),
\end{aligned} \tag{4.5}$$

where the inequality uses (4.2). The reverse inequality is obtained in a similar manner. □

**Lemma 4.2.** *Let $m \in \mathbb{N}$, and let $(i_1, \dots, i_m)$, $(j_1, \dots, j_m)$ be two collections of integers satisfying $i_1 < j_1 \le i_2 < j_2 \le \dots < i_{m-1} < j_{m-1} \le i_m < j_m$. For $1 \le k \le m$, let $A_k \in \mathcal{F}_{(i_k,j_k]}$ and $p_k = \nu(A_k)$. Suppose that $\nu$ satisfies condition* (SV)*. Then*

$$\nu\left(\cap_{1\le k\le m} A_k\right) \le C(\varphi)^{m-1} \prod_{1\le k\le m} p_k. \tag{4.6}$$

*Proof.* We give the proof for $m=2$. The general case can be handled by induction. Let $i_1<j_1\le i_2<j_2$, $A_1\subset E^{j_1-i_1}$ and $A_2\subset E^{j_2-i_2}$. For all $x^-\in \mathrm{E}^{-\mathbb{N}_0}$,

$$\begin{aligned}
&\nu\big(X_{(i_1,j_1]}\in A_1, X_{(i_2,j_2]}\in A_2\big)\\
&=\sum_{\substack{x_{(i_1,j_1]}\in A_1\\ x_{(i_2,j_2]}\in A_2}} \nu\big(X_{(i_1,j_1]}=x_{(i_1,j_1]}, X_{(i_2,j_2]}=x_{(i_2,j_2]}\big)\\
&=\sum_{\substack{x_{(i_1,j_1]}\in A_1\\ x_{(i_2,j_2]}\in A_2}} \nu\big(X_{(i_1-j_1,0]}=x_{(i_1,j_1]}, X_{(i_2-j_1,j_2-j_1]}=x_{(i_2,j_2]}\big)\\
&=\sum_{\substack{x_{(i_1,j_1]}\in A_1\\ x_{(i_2,j_2]}\in A_2}} \nu\big(X_{(i_1-j_1,0]}=x_{(i_1,j_1]}\big)\,\nu\big(X_{(i_2-j_1,j_2-j_1]}=x_{(i_2,j_2]}\mid X_{(i_1-j_1,0]}=x_{(i_1,j_1]}\big)\\
&\le C(\varphi)\sum_{\substack{x_{(i_1,j_1]}\in A_1\\ x_{(i_2,j_2]}\in A_2}} \nu\big(X_{(i_1-j_1,0]}=x_{(i_1,j_1]}\big)\,\nu_{x^-}\big(X_{(i_2-j_1,j_2-j_1]}=x_{(i_2,j_2]}\big)\\
&=C(\varphi)p_1\sum_{x_{(i_2,j_2]}\in A_2}\nu_{x^-}\big(X_{(i_2-j_1,j_2-j_1]}=x_{(i_2,j_2]}\big),
\end{aligned}\tag{4.7}$$

where the inequality uses (4.3) in Lemma 4.1. Averaging $x^-$ w.r.t. $\nu$, we get

$$\nu(X_{(i_1,j_1]}\in A_1, X_{(i_2,j_2]}\in A_2)\le C(\varphi)p_1p_2. \tag{4.8}$$

□

4.2. **Successive occurrences of patterns.**

**Lemma 4.3.** *Fix $m\in\mathbb{N}$ and let $A\in\mathcal{F}_{(0,m]}$ be such that $\nu(A)>0$. Let $(\sigma_n)_{n\in\mathbb{Z}}$ be defined by*

$$\begin{aligned}
&\sigma_0=\inf\{k\ge 0\colon\ \theta^kX\in A\}+m,\\
\forall \ell\in\mathbb{N},\qquad &\sigma_\ell=\inf\{k\ge \sigma_{\ell-1}\colon\ \theta^kX\in A\}+m,\\
\forall \ell\in-\mathbb{N},\qquad &\sigma_\ell=\sup\{k\le \sigma_{\ell+1}-2m\colon\ \theta^kX\in A\}+m.
\end{aligned}\tag{4.9}$$

*If $\nu$ satisfies condition* (SV)*, then $\nu$-a.s.,*

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{1\le\ell\le n}\log[\sigma_\ell-\sigma_{\ell-1}]\le C(\varphi)E_\nu[\log\sigma_1]. \tag{4.10}$$

*Proof.* The strategy of proof consists in writing the sum in (4.10) as an additive functional of an ergodic process and to use Birkhoff's ergodic theorem. First note that the sequence $(\sigma_n)_{n\in\mathbb{N}_0}$ cuts blocks out of the letter sequence $X$, which we denote by

$$B_n=X_{(\sigma_{n-1},\sigma_n]}\in\widetilde{E},\qquad n\in\mathbb{N}. \tag{4.11}$$

Each of these blocks belongs to the following subset of words:

$$\widetilde{E}_A=\big\{y\in\widetilde{E}\colon\ |y|\ge m;\ \forall\, 0\le k<|y|-m\colon\ y_{(k,k+m]}\notin A;\ y_{(|y|-m,|y|]}\in A\big\}. \tag{4.12}$$

Define the process $B^\star=(B^\star_n)_{n\in\mathbb{N}_0}$ in $E^{-\mathbb{N}_0}$ by putting $B^\star_n=X_{(-\infty,\sigma_n]}$. This process is Markovian and its transition kernel is given by

$$\mathrm{P}^A_x(\hat{x})=\mathrm{P}(B^\star_{n+1}=\hat{x}\mid B^\star_n=x)=\sum_{y\in\widetilde{E}_A}1_{\{\hat{x}=x,y\}}\nu_x(X_{(0,|y|]}=y),\qquad x,\hat{x}\in E^{-\mathbb{N}_0}. \tag{4.13}$$

For the collection $(\mathrm{P}^A_x(\cdot),x\in E^{-\mathbb{N}_0})$ to be a proper transition kernel, $\sigma_1$ must be $\nu_x$-a.s. finite for all $x\in E^{-\mathbb{N}_0}$. Since $\nu(A)>0$, we know from the Recurrence Theorem in Halmos [8] that

$\sigma_1$ is $\nu$-a.s. finite. Since $\nu$ and $(\nu_x)_{x\in E^{-\mathbb{N}_0}}$ are equivalent under condition (SV) (note that $C(\varphi)^{-1}\nu(\cdot) \le \nu_x(\cdot) \le C(\varphi)\nu(\cdot)$ as a consequence of (4.2) in Lemma 4.1), $\sigma_1$ is $\nu_x$-a.s. finite for all $x \in E^{-\mathbb{N}_0}$. Since (with a slight abuse of notation) the $B_n^\star$'s are also in $E^{-\mathbb{N}_0} \times \widetilde{E}_A$, we can write

$$\sum_{1\le\ell\le n} \log[\sigma_\ell - \sigma_{\ell-1}] = \sum_{1\le\ell\le n} \log|\pi(B_\ell^\star)|, \tag{4.14}$$

where $\pi$ is defined by $\pi\colon\ (u,v) \in E^{-\mathbb{N}_0} \times \widetilde{E}_A \mapsto v$. We next apply Birkhoff's ergodic theorem to the sum in the right-hand side, i.e., to the process $B^\star$. This process has a stationary distribution, which we denote by $P_A$. It is easy to check that $P_A$ is the law of $X_{(-\infty,\sigma_0]}$ conditional on the event $\cap_{\ell\in-\mathbb{N}_0}\{\sigma_\ell > -\infty\}$, which has probability one according to the Recurrence Theorem. Again using (4.2) in Lemma 4.1, we see that for all sets $\mathcal{A}$ and $\mathcal{B}$ that are measurable w.r.t. $\sigma(B^\star_{(-\infty,0]})$ and $\sigma(B^\star_{(0,\infty)})$, respectively,

$$C(\varphi)^{-1} P_A(\mathcal{A}) P_A(\mathcal{B}) \le P_A(\mathcal{A}\cap\mathcal{B}) \le C(\varphi) P_A(\mathcal{A}) P_A(\mathcal{B}). \tag{4.15}$$

Therefore $P_A$ is Weak Bernoulli (Ledrappier [12]), and hence is ergodic. Thus, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{1\le\ell\le n} \log[\sigma_\ell - \sigma_{\ell-1}] = E_{P^A}(\log[\sigma_1 - \sigma_0]). \tag{4.16}$$

Moreover, for all $\hat{x}^- \in E^{-\mathbb{N}_0}$,

$$E_{P^A}(\log[\sigma_1 - \sigma_0]) = \int E_{\nu_{x^-}}(\log[\sigma_1 - \sigma_0]) dP_A(x^-) \le C(\varphi) E_{\nu_{\hat{x}^-}}(\log[\sigma_1 - \sigma_0]), \tag{4.17}$$

which gives $E_{P^A}(\log[\sigma_1 - \sigma_0]) \le C(\varphi) E_\nu(\log[\sigma_1 - \sigma_0])$ and completes the proof. □

4.3. **Decomposition of relative entropy.**

**Lemma 4.4.** *Suppose that $\varphi(0) < \infty$. Then, for all $Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})$,*

$$\begin{aligned} H(Q \mid \mathrm{P}) &= -H(Q) - E_Q[\log \varrho(\tau_1)] - m_Q E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)], \\ H(\Psi_Q \mid \nu) &= -H(\Psi_Q) - E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)]. \end{aligned} \tag{4.18}$$

*Proof.* To get the first relation, write $H(Q \mid \mathrm{P}) = -H(Q) - \mathrm{E}_Q[\log \mathrm{P}_{Y_{(-\infty,0]}}(Y_1)]$,

$$\mathrm{E}_Q[\log \mathrm{P}_{Y_{(-\infty,0]}}(Y_1)] = \mathrm{E}_Q[\log \varrho(\tau_1)] + \mathrm{E}_Q[\log \nu_{X_{(-\infty,0]}}(X_{(0,\tau_1]})] \tag{4.19}$$

and (recall (1.3))

$$\mathrm{E}_Q[\log \nu_{X_{(-\infty,0]}}(X_{(0,\tau_1]})] = \mathrm{E}_Q\left[\sum_{k=0}^{\tau_1-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1})\right] = m_Q \mathrm{E}_{\psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)], \tag{4.20}$$

where we use the abbreviation $\nu_{x^-}(x_\Lambda) = \nu_{x^-}(X_\Lambda = x_\Lambda)$, $\Lambda \subset \mathbb{N}$. The second relation follows in a similar manner. □

All terms in the right-hand side of (4.18), except possibly $H(Q)$, are finite because $E$ is finite, $\varrho$ satisfies (1.5), and $\varphi(0) < \infty$.

**Lemma 4.5.** *If $\nu$ satisfies condition* (SV)*, then for all $Q \in \mathcal{P}^{\mathrm{inv,erg}}(\widetilde{E}^{\mathbb{N}})$,*

$$\lim_{n\to\infty} \frac{1}{n} \log \nu(X_{(0,T_n]}) = m_Q E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)] \quad Q-a.s. \tag{4.21}$$

*Proof.* First observe that (4.3) in Lemma 4.1 gives

$$C(\varphi)^{-1}\nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) \leq \nu(X_{(0,T_n]}) \leq C(\varphi)\nu_{X_{(-\infty,0]}}(X_{(0,T_n]}). \tag{4.22}$$

Next write

$$\log \nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) = \sum_{k=0}^{T_n-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1}) = \sum_{i=0}^{n-1}\sum_{k=T_i}^{T_{i+1}-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1}). \tag{4.23}$$

Use (4.23) and the ergodicity of $Q$ to obtain, for $Q$-a.s. $Y$,

$$\lim_{n\to\infty}\frac{1}{n}\log \nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) = \mathrm{E}_Q\left[\sum_{k=0}^{\tau_1-1}\log \nu_{X_{(-\infty,k]}}(X_{k+1})\right] = m_Q \mathrm{E}_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)]. \tag{4.24}$$

Combine (4.22–4.24) to get the claim. □

4.4. **Proof of quenched LDP.** We are now ready to give the proof of Theorem 1.2.

*Proof.* The proof is an extension of the proof in [3] for i.i.d. $\nu$. Since the latter is rather long, it is not possible to repeat all the ingredients here. Below we restrict ourselves to indicating the necessary *modifications*, which are based on the results in Sections 4.1–4.3. We leave it to the reader to go over the full proof in [3] and check that, indeed, these are the only modifications needed.

*Decomposition of relative entropies.* Replace Eq.(1.25) and Eq.(1.26) of [3] by the relations in Lemma 4.4.

*Upper bound.* Fix $\varepsilon_1, \delta_1 > 0$. Replace the fourth line in the definition of the event defined in Eq. (3.4) of [3] by

$$\left\{\frac{1}{M}\log \nu(X_{(0,T_M]}) \in m_Q E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)] + [-\varepsilon_1, \varepsilon_1]\right\}. \tag{4.25}$$

By Lemma 4.5, the event in Eq.(3.4) of [3] has probability at least $1-\delta_1/4$ for $M$ large enough. Parts 3.2 and 3.3 of [3] are unchanged. The next (harmless) modification is in Eq.(3.39) of [3], which has to be replaced by

$$\mathrm{P}\left(\cap_{1\leq k\leq n}\{A_k = a_k\}\right) \leq [C(\varphi)p]^{\sum_{1\leq k\leq n} a_k}, \tag{4.26}$$

where $A_k$ is the indicator defined in Eq.(3.36) and Eq.(3.37) of [3], and $a_k \in \{0,1\}$. This relation can be proved via Lemma 4.2.

*Lower Bound.* One modification is needed to go from Eq.(4.7) to Eq.(4.8) of [3], since the increments of the $\sigma_\ell^{(M)}$, $\ell \in \mathbb{N}$, defined in Eq.(4.6) of [3] are no longer i.i.d. Use Lemma 4.3 instead. □

## 5. Extension to Polish spaces

In this section we prove Theorem 1.3, i.e., we extend the LDP's in Theorems 1.1–1.2 from a finite letter space to a Polish letter space. We first prove the LDP's for a sequence of *coarse-grained* finite letter spaces associated with a sequence of nested finite partitions of the Polish letter space. After that we apply the *Dawson-Gärtner projective limit LDP* (see Dembo and Zeitouni [4], Lemma 4.6.1). A somewhat delicate point is that (SV) for the full process does not necessarily imply (SV) for the coarse-grained process. Indeed, the first supremum in (2.10) decreases under coarse-graining while the second supremum increases. The way out is to use (SV) for the full process to prove the decoupling inequalities in Section 4.1 for the coarse-grained process.

Let $X = (X_k)_{k\in\mathbb{Z}}$ be a stationary process on a Polish space $(E, d)$, with $(\nu_{x^-}(\cdot),\ x^- \in E^{-\mathbb{N}_0})$ a regular version of the conditional probability $\nu(\cdot \mid X_{(-\infty,0]})$ satisfying condition (SV), i.e.,

$$C(\varphi) = \exp\left[\sum_{n\in\mathbb{N}_0} \varphi(n)\right] < \infty, \tag{5.1}$$

where

$$\varphi(n) = \sup_{\substack{x^-,\hat{x}^-\in E^{-\mathbb{N}_0}:\\ d(x^-,\hat{x}^-)\le 2^{-n}}} \ \sup_{\substack{A\in\mathcal{F}_1:\\ \nu_{x^-}(A)>0}} |\log \nu_{x^-}(A) - \log \nu_{\hat{x}^-}(A)| \tag{5.2}$$

with

$$d(x^-, \hat{x}^-) = \sum_{k\in\mathbb{N}_0} 2^{-(k+1)} \left[1 \wedge d\big(x^-_{-k}, \hat{x}^-_{-k}\big)\right]. \tag{5.3}$$

We assume that, for any $x^-, \hat{x}^- \in E^{-\mathbb{N}_0}$, the measures $\nu_{x^-}|_1 = \nu_{x^-}(X_1 \in \cdot\,)$ and $\nu_{\hat{x}^-}|_1 = \nu_{\hat{x}^-}(X_1 \in \cdot\,)$ are equivalent, so that the Radon-Nikodym derivative $\mathrm{d}\nu_{x^-}|_1/\mathrm{d}\nu_{\hat{x}^-}|_1$ exists and

$$\sup_{\substack{A\in\mathcal{F}_1:\\ \nu_{x^-}(A)>0}} \Big[\log \nu_{x^-}(A) - \log \nu_{\hat{x}^-}(A)\Big] = \operatorname{supess}\left[\log \frac{\mathrm{d}\nu_{x^-}|_1}{\mathrm{d}\nu_{\hat{x}^-}|_1}\right], \tag{5.4}$$

leading to the alternative definition

$$\varphi(n) = \sup_{\substack{x^-,\hat{x}^-\in E^{-\mathbb{N}_0}:\\ d(x^-,\hat{x}^-)\le 2^{-n}}} \operatorname{supess}\left[\log \frac{\mathrm{d}\nu_{x^-}|_1}{\mathrm{d}\nu_{\hat{x}^-}|_1}\right]. \tag{5.5}$$

Similarly as in Section 2.3, we note that (SV) holds for i.i.d. processes, for Markov chains of finite order with $\varphi(0) < \infty$, and a subclass of chains with complete connections whose letter space is countable (Berbee [1]). Other examples are rotators that are labelled by $\mathbb{Z}$, take values in the unit circle, and interact with each other according to a Hamiltonian with long-range potentials that have a sufficiently thin tail.

The following lemma generalizes (4.2) in Lemma 4.1.

**Lemma 5.1.** *For all $x^-$,$\hat{x}^- \in E^{-\mathbb{N}_0}$ and $A \in \mathcal{F}_{(0,\infty)}$,*

$$C(\varphi)^{-1}\nu_{\hat{x}^-}(A) \le \nu_{x^-}(A) \le C(\varphi)\nu_{\hat{x}^-}(A). \tag{5.6}$$

*Proof.* For all $x^-, \hat{x}^- \in E^{-\mathbb{N}_0}$ and $n \in \mathbb{N}$,

$$\frac{\mathrm{d}\nu_{x^-}|_n}{\mathrm{d}\nu_{\hat{x}^-}|_n}(x_1,\dots,x_n) = \frac{\mathrm{d}\nu_{x^-}|_1}{\mathrm{d}\nu_{\hat{x}^-}|_1}(x_1) \times \frac{\mathrm{d}\nu_{x^-x_1}|_1}{\mathrm{d}\nu_{\hat{x}^-x_1}|_1}(x_2) \times \cdots \times \frac{\mathrm{d}\nu_{x^-x_1\cdots x_{n-1}}|_1}{\mathrm{d}\nu_{\hat{x}^-x_1\cdots x_{n-1}}|_1}(x_n) \tag{5.7}$$

$$\le \exp[\varphi(0) + \varphi(1) + \cdots + \varphi(n-1)] \le C(\varphi), \tag{5.8}$$

which proves the claim. □

Let $\mathcal{E}_c = \{E_1, \dots, E_c\}$, $c \in \mathbb{N}$, be a finite partition of $E$. Identify $\mathcal{E}_c^{\mathbb{Z}}$ with $\{1, \dots, c\}^{\mathbb{Z}}$. Let $X^{(c)} = (X_k^{(c)})_{k\in\mathbb{Z}}$ on $\mathcal{E}_c^{\mathbb{Z}}$ be the coarse-graining of $X$ on $E^{\mathbb{Z}}$ defined by

$$X_n^{(c)} = \sum_{i=1}^{c} i\, 1_{\{X_n\in E_i\}}. \tag{5.9}$$

Write $\mathcal{F}^{(c)}_{(0,\infty)} = \sigma(X^{(c)}_{(0,\infty)})$. The following lemma generalizes (4.3) in Lemma 4.1.

**Lemma 5.2.** *For all $x^- \in E^{-\mathbb{N}_0}$, $c \in \mathbb{N}$, $i^-, j^- \in \{1,\dots,c\}^{-\mathbb{N}_0}$, $A \in \mathcal{F}^{(c)}_{(0,\infty)}$ and $m,n \in \mathbb{N}$,*

$$C(\varphi)^{-1}\nu_{x^-}(A) \le \nu\left(A \mid X^{(c)}_{(-n,0]} = i^-_{(-n,0]}\right) \le C(\varphi)\nu_{x^-}(A), \tag{5.10}$$

$$\begin{aligned} C(\varphi)^{-2}\nu\left(A \mid X^{(c)}_{(-m,0]} = j^-_{(-m,0]}\right) &\le \nu\big(A \mid X^{(c)}_{(-n,0]} = i^-_{(-n,0]}\big) \\ &\le C(\varphi)^2\nu\left(A \mid X^{(c)}_{(-m,0]} = j^-_{(-m,0]}\right), \end{aligned} \tag{5.11}$$

$$C(\varphi)^{-1}\nu\left(A \mid X^{(c)}_{(-n,0]} = i^-_{(-n,0]}\right) \le \nu(A) \le C(\varphi)\nu\left(A \mid X^{(c)}_{(-n,0]} = i^-_{(-n,0]}\right), \tag{5.12}$$

*provided that the events on which we condition have positive probability.*

*Proof.* Note that (5.11) follows by applying (5.10) twice, while (5.12) follows by integrating $x^-$ w.r.t. $\nu$ in (5.10). Therefore it suffices to prove (5.10). To that end write

$$\nu\left(A \mid X^{(c)}_{(-n,0]} = i^-_{(-n,0]}\right) = \frac{\int_{E^{-\mathbb{N}_0}} \nu_{\widetilde{x}^-}(\{X^{(c)}_{(0,n]} = i^-_{(-n,0]}\} \cap \theta^{-n}A)\,\mathrm{d}\nu(\widetilde{x}^-)}{\nu(X^{(c)}_{(-n,0]} = i^-_{(-n,0]})}. \tag{5.13}$$

The integral in the numerator equals

$$\int_{E^{-\mathbb{N}_0}} \Big[\int_{E^n} \mathrm{d}\nu_{\widetilde{x}^-}(x_{(0,n]})\, 1_{\{x^{(c)}_{(0,n]} = i^-_{(-n,0]}\}}\, \nu_{x^-,x_{(0,n]}}(A)\Big]\,\mathrm{d}\nu(\widetilde{x}^-), \tag{5.14}$$

from which the claim follows via Lemma 5.1. □

In what follows we need the notion of *conditional local absolute continuity* (which is weaker than absolute continuity).

**Definition 5.3.** *Let $F$ be a Polish space equipped with its Borel $\sigma$-algebra, and let $\lambda, \mu$ be two stationary probability measures on $F^{\mathbb{Z}}$ with respective regular conditional probabilities $(\lambda_{x^-},\, x^- \in F^{-\mathbb{N}_0})$ and $(\mu_{x^-},\, x^- \in F^{-\mathbb{N}_0})$. The law $\lambda$ is said to be conditionally locally absolutely continuous w.r.t. to the law $\mu$* (written as $\lambda \ll_{\mathrm{cond}} \mu$) when, for $\lambda$-a.a. $x^-$ and all $n \in \mathbb{N}$, $\lambda_{x^-}|_n$ is absolutely continuous w.r.t. to $\mu_{x^-}|_n$ (written as $\lambda_{x^-}|_n \ll \mu_{x^-}|_n$), where $\lambda_{x^-}|_n$ and $\mu_{x^-}|_n$ are the marginal laws on the first $n$ coordinates.

Note that because $F$ is Polish the set $\{x^- \in F^{-\mathbb{N}_0} \colon \lambda_{x^-}|_n \ll \mu_{x^-}|_n\}$ is measurable. We are now ready to prove Theorem 1.3.

*Proof.* We need to prove both the annealed LDP and the quenched LDP.

**Annealed LDP.** Lemma 5.1 shows that under condition (SV) Lemmas 2.4–2.5 carry over from finite letters to Polish letters. Therefore the ratio-mixing property of Orey and Pelikan [14] again yields the annealed LDP.

**Quenched LDP.** The proof comes in 4 steps.

**1.** We first use Lemmas 5.1–5.2 to show that Lemmas 4.2–4.5 carry over to the coarse-grained process $X^{(c)}$ defined in (5.9) for every $c \in \mathbb{N}$. This is straightforward, except that Lemma 4.5 carries over to $Q \in \mathcal{P}^{\mathrm{inv,erg}}((\widetilde{\mathcal{E}}_c)^{\mathbb{Z}})$ only when $\Psi_Q \ll_{\mathrm{cond}} \nu^{(c)}$, where $\nu^{(c)}$ denotes the law of $X^{(c)}$. We will see in Step 4 below that, because $H(\Psi_Q \mid \nu^{(c)}) = \infty$ when $\Psi_Q \ll_{\mathrm{cond}} \nu^{(c)}$ fails, this restriction does not affect the LDP.

**2.** To prove the restricted version of Lemma 4.5, let $Q \in \mathcal{P}^{\mathrm{inv,erg}}((\widetilde{\mathcal{E}}_c)^{\mathbb{Z}})$ be such that $\Psi_Q \ll_{\mathrm{cond}} \nu^{(c)}$. Using the notation introduced below (4.20), we know from Lemma 5.2 (by letting $n \to \infty$ in Eq. (5.12) and using the Martingale Convergence Theorem) that, for $\nu^{(c)}$-a.a. $X^{(c)}_{(-\infty,0]}$,

$$\nu\Big(C(\varphi)^{-1}\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big) \le \nu\big(X^{(c)}_{(0,n)}\big) \le C(\varphi)\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big) \;\Big|\; X^{(c)}_{(-\infty,0]}\Big) = 1. \quad (5.15)$$

By conditional local absolute continuity we have, for $\Psi_Q$-a.a. $X^{(c)}_{(-\infty,0]}$,

$$\Psi_Q\Big(C(\varphi)^{-1}\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big) \le \nu\big(X^{(c)}_{(0,n)}\big) \le C(\varphi)\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big) \;\Big|\; X^{(c)}_{(-\infty,0]}\Big) = 1. \quad (5.16)$$

This implies that, for $\Psi_Q$-a.a. $X^{(c)}_{(-\infty,0]}$,

$$C(\varphi)^{-1}\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big) \le \nu\big(X^{(c)}_{(0,n)}\big) \le C(\varphi)\nu_{X^{(c)}(-\infty,0]}\big(X^{(c)}_{(0,n)}\big), \quad (5.17)$$

which settles the restricted version of Lemma 4.5.

**3.** By the same argument as in Section 4.4, we now know that the quenched LDP holds for $X^{(c)}$ for all $c \in \mathbb{N}$ (see Step 4 below for comments). Picking for $\mathcal{E}_c = \{E_1, \dots, E_c\}$, $c \in \mathbb{N}$, a *nested sequence* of finite partitions of $E$ as in [3, Section 8], we conclude from the Dawson-Gärtner projective limit LDP that the quenched LDP also holds for $X$, with rate function

$$I^{\mathrm{que}}(Q) = \sup_{c\in\mathbb{N}} I^{\mathrm{que}}_c(Q^{(c)}), \quad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}}), \quad (5.18)$$

where $Q^{(c)}$ is the coarse-graining of $Q$, and $I^{\mathrm{que}}_c$ is the coarse-grained rate function. The argument in [3, Section 8] shows that the supremum equals the rate function given in (1.7), i.e., the coarse-grained relative entropies converge to the full relative entropies as $c \to \infty$. (Deuschel and Stroock [5, Lemma 4.4.15] implies that the coarse-grained relative entropies are monotone in $c$.)

**4.** To obtain the quenched LDP, we must prove Eq.(3.1) and Eq.(4.1) in [3] for the coarse-grained process. In Steps 1–3 this has already been achieved for $Q \in \mathcal{P}^{\mathrm{inv,fin}}((\widetilde{\mathcal{E}}_c)^{\mathbb{Z}})$ with $\Psi_Q \ll_{\mathrm{cond}} \nu^{(c)}$. Eq.(4.1) in [3] trivially carries over when the latter restriction fails, but for Eq.(3.1) an additional argument is needed. We must show that there exists a sequence $(\mathcal{O}_k(Q))_{k\in\mathbb{N}}$ of shrinking open neighborhoods of $Q$ such that

$$\lim_{k\to\infty} \limsup_{N\to\infty} \frac{1}{N} \log \mathrm{P}^{(c)}\big(R^{(c)}_N \in \mathcal{O}_k(Q) \mid X^{(c)}\big) = -\infty, \quad (5.19)$$

where $\mathrm{P}^{(c)}$ denotes the coarse-graining of P. This can be done via an annealed estimate. Indeed, for $\nu^{(c)}$-a.a. $X^{(c)}$,

$$\begin{aligned}\limsup_{N\to\infty} \frac{1}{N} \log \mathrm{P}^{(c)}\big(R^{(c)}_N \in \mathcal{O}_k(Q) \mid X^{(c)}\big) &\le \limsup_{N\to\infty} \frac{1}{N} \log \mathrm{P}^{(c)}\big(R^{(c)}_N \in \mathcal{O}_k(Q)\big)\\ &\le -\inf_{Q'\in\overline{\mathcal{O}_k(Q)}} H(Q' \mid \mathrm{P}^{(c)}),\end{aligned} \quad (5.20)$$

where the last inequality follows from the annealed LDP. (This needs justification, since the annealed LDP was proved under condition (SV), which is not necessarily satisfied for $\nu^{(c)}$. However, by Lemma 5.2, a decoupling inequality holds for a.a. pairs of coarse-grained pasts. Therefore there must be a regular conditional probability of $\nu^{(c)}$ satisfying Orey and Pelikan's ratio-mixing condition.) A sequence $(\mathcal{O}_k(Q))_{k\in\mathbb{N}}$ satisfying (5.19) is easily obtained by letting $k \to \infty$ and using the lower semi-continuity of $Q' \mapsto H(Q' \mid \mathrm{P}^{(c)})$ together with the fact that $H(Q \mid P^{(c)}) \ge m_Q H(\Psi_Q \mid \nu^{(c)}) = \infty$ (see [3, Eqs.(1.30–1.32)]). $\square$

## References

[1] H. Berbee, *Chains with infinite connections: uniqueness and Markov representation*, Probab. Th. Rel. Fields 76 (1987) 243–253.
[2] M. Birkner, *Conditional large deviations for a sequence of words*, Stoch. Proc. Appl. 118 (2008) 703–729.
[3] M. Birkner, A. Greven and F. den Hollander, *Quenched large deviation principle for words in a letter sequence*, Probab. Theory Related Fields 148 (2010) 403–456.
[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (2nd Ed.), Springer, 1998.
[5] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Academic, Boston (1989).
[6] M.D. Donsker and S.R.S. Varadhan, *Asymptotic evaluation of certain Markov process expectations for large time. IV*, Comm. Pure Appl. Math. 36 (1983) 183–212.
[7] M.D. Donsker and S.R.S. Varadhan, *Large deviations for stationary Gaussian processes*, Comm. Math. Phys. 97 (1985), no. 1-2, 187–210.
[8] P.R. Halmos, *Lectures on ergodic theory*, Chelsea Publishing Co., New York, 1960.
[9] F. den Hollander and J.E. Steif, *Mixing properties of the generalized $T,T^{-1}$-process*, J. d' Analyse Math. 72 (1997) 165–202.
[10] A. Johansson and A. Öberg, *Square summability of variations of g-functions and uniqueness of g-measures*, Math. Res. Lett. 10 (2003), nos. 5–6, 587–601.
[11] A. Johansson, A. Öberg and M. Pollicott, *Countable state shifts and uniqueness of g-measures*, Amer. J. Math. 129 (2007), no. 6, 1501–1511.
[12] F. Ledrappier, *Sur la condition de Bernoulli faible et ses applications*, in: *Théorie Ergodique* (Actes Journées Ergodiques, Rennes, 1973/1974), Lecture Notes in Math. 532, Springer, Berlin, 1976, pp. 152–159.
[13] S. Orey, *Large deviations in ergodic theory*, in: *Seminar on Stochastic Processes, Evanston, Illinois, 1984*, Progr. Probab. Statist. 9 (1984) 195—249.
[14] S. Orey and S. Pelikan, *Large deviation principles for stationary processes*, Ann. Probab. 16 (1988) 1481–1495.

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands.
*E-mail address*: `denholla@math.leidenuniv.nl`

Mathematical Institute, Leiden University, P.O. Box 9512, 2300 RA Leiden, The Netherlands.
*E-mail address*: `poisatj@math.leidenuniv.nl`