Electron. Commun. Probab. **19** (2014), no. 12, 1–16.
DOI: 10.1214/ECP.v19-2681
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Large deviation principles for words drawn from correlated letter sequences

Frank den Hollander* Julien 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.

Keywords: Letters and words ; renewal times ; empirical process ; annealed vs. quenched large deviation principle ; rate function ; specific relative entropy ; mixing.
AMS MSC 2010: 60F10 ; 60G10.
Submitted to ECP on March 18, 2013, final version accepted on February 23, 2014.

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. Both E and $\widetilde{E}$ are Polish spaces under the discrete topology. Write $E^{\mathbb{Z}}$ and $\widetilde{E}^{\mathbb{Z}}$ for the sets of two-sided sequences of letters and words, endowed with the product topology, and let θ 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. θ 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 ν on $E^{\mathbb{Z}}$. Let $\tau = (\tau_i)_{i\in\mathbb{Z}}$ be a two-sided i.i.d.

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

sequence of *renewal times* drawn from a common probability law ϱ 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 τ, 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 τ is denoted by P. Write $|Y_i|$ to denote the length of word i.

The reverse of cutting is glueing. The *concatenation operator* $\kappa\colon \widetilde{E}^{\mathbb{Z}} \to E^{\mathbb{Z}}$ 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 $m_Q = E_Q(|Y_1|) < \infty$, 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 ν-a.a. X (= quenched law).

1.2 Large deviation principles

If ν 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 ϱ has an exponentially bounded tail, and in Birkner, Greven and den Hollander [3] for the case where ϱ 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 ϱ 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. ν, 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 ν 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], Section 6.5 for the definition and key properties).

Theorem 1.1 (Annealed LDP). *If ν 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^{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 ν has summable variation, then for ν-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^{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* $\lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu)$, where tr is the operator that truncates all the words to length $\leq \mathrm{tr}$. Moreover, for all $Q \in \mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{Z}}) = \{\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})\colon\ m_Q < \infty\}$,

$$\lim_{\mathrm{tr}\to\infty} H([Q]_{\mathrm{tr}} \mid \mathrm{P}) = H(Q \mid \mathrm{P}), \qquad \lim_{\mathrm{tr}\to\infty} m_{[Q]_{\mathrm{tr}}} H(\Psi_{[Q]_{\mathrm{tr}}} \mid \nu) = m_Q H(\Psi_Q \mid \nu). \tag{1.8}$$

For details, see Birkner, Greven and den Hollander [3] and the end of Section 4.4 below.

Remark: Both rate functions are the same as for the i.i.d. case, even though the reference laws P and ν 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

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

The associated sigma-algebra's are written as

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

Write $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. For $x \in E^{\mathbb{Z}}$ and $y \in \widetilde{E}^{\mathbb{Z}}$ we use the short-hand notation $x^- = x_{-\mathbb{N}_0}$ and $y^- = y_{-\mathbb{N}_0}$. Let $(\nu_{x^-}(\cdot); x^- \in E^{-\mathbb{N}_0})$ be a regular version of $\nu(\cdot \mid X_{(-\infty,0]})$ (see Parthasarathy [15, Theorem 8.1]), i.e.,

$$\nu(A) = \int_{x^- \in E^{-\mathbb{N}_0}} \nu_{x^-}(A)\, \mathrm{d}\nu(x^-), \qquad A \in \mathcal{F}_{(0,\infty)}. \tag{2.3}$$

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. This allows us to obtain a regular version of the conditional probabilities of P as follows.

Lemma 2.1. *The collection* $(\mathrm{P}_{y^-}(\cdot), y^- \in \widetilde{E}^{-\mathbb{N}_0})$ *of probability laws on* $\widetilde{E}^{\mathbb{N}}$ *defined by*

$$\mathrm{P}_{y^-}(A) = \int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\, \mathrm{d}\nu_{\kappa(y^-)} \qquad \forall\, A \in \mathcal{G}_{(0,\infty)}, \tag{2.4}$$

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 (2.4) is a probability measure. We must show that

$$\int_{\widetilde{E}^{-\mathbb{N}_0}} \mathrm{P}_{y^-}(\cdot)\, \mathrm{dP}(y^-) = \mathrm{P}(\cdot). \tag{2.5}$$

By the monotone class theorem, it is enough to prove the claim for finite cylinder sets. Fix $r \in \mathbb{N}$, $(y_i)_{1\le i\le r} \in \widetilde{E}^r$ and pick $A = \bigcap_{1\le i\le r}\{Y_i = y_i\}$. Then

$$\int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\, \mathrm{d}\nu_{\kappa(y^-)} = \int_{E^{\mathbb{Z}}} \mathrm{d}\nu_{\kappa(y^-)}\, 1_{\{X\in\kappa(A)\}} \prod_{i=1}^{r} \varrho(|y_i|) = \nu_{\kappa(y^-)}(X \in \kappa(A)) \prod_{i=1}^{r} \varrho(|y_i|), \tag{2.6}$$

where $\kappa(A)$ is the concatenation of A. 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

$$\int_{\widetilde{E}^{-\mathbb{N}_0}} \mathrm{dP}(y^-) \int_{E^{\mathbb{Z}}} \mathrm{P}(A \mid \mathcal{F}_{\mathbb{Z}})\, \mathrm{d}\nu_{\kappa(y^-)} = \nu(X \in \kappa(A)) \prod_{i=1}^{r} \varrho(|y_i|) = \mathrm{P}(A), \tag{2.7}$$

which proves the claim. □

2.2 Mixing coefficients

We need the following mixing coefficients for letters and words:

Definition 2.2. *(a) For* $\Lambda_1 \subset -\mathbb{N}_0$ *and* $\Lambda_2 \subset \mathbb{N}$*, let*

$$\varphi(\Lambda_1, \Lambda_2) = \sup_{\substack{x^-,\hat{x}^- \in E^{-\mathbb{N}_0} \\ (x^-)_{\Lambda_1} = (\hat{x}^-)_{\Lambda_1}}} \ \sup_{\substack{A\in\mathcal{F}_{\Lambda_2}: \\ \nu_{x^-}(A)>0}} \left|\log \nu_{x^-}(A) - \log \nu_{\hat{x}^-}(A)\right|. \tag{2.8}$$

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

$$\psi(\Lambda) = \sup_{y^-,\hat{y}^- \in \widetilde{E}^{-\mathbb{N}_0}} \ \sup_{\substack{A\in\mathcal{G}_\Lambda \\ P_{y^-}(A)>0}} \left|\log \mathrm{P}_{y^-}(A) - \log \mathrm{P}_{\hat{y}^-}(A)\right|. \tag{2.9}$$

The restrictions $\nu_{x^-}(A) > 0$ and $\mathrm{P}_{\hat{y}^-}(A) > 0$ are put in to avoid $\infty-\infty$. Nonetheless, (2.8) and (2.9) may be infinite. Note that if $\Lambda_1 = \emptyset$, then the supremum in Definition 2.2(a) is taken over all $x^-, \hat{x}^- \in E^{-\mathbb{N}_0}$ without any restriction ($(x^-)_\Lambda$ denotes the restriction of x^- to Λ). We will use the following abbreviations:

$$\varphi(k,\cdot) = \varphi((-k,0],\cdot), \quad k \in \mathbb{N}, \qquad \varphi(0,\cdot) = \varphi(\emptyset,\cdot), \qquad \varphi(\cdot,\ell) = \varphi(\cdot,(0,\ell]), \quad \ell \in \mathbb{N}. \tag{2.10}$$

Lemma 2.3. *Let* $0 \le m < n$, $y_{(m,n]} \in \widetilde{E}^{n-m}$ *and* $A = \{Y_{(m,n]} = y_{(m,n]}\}$*. For all* $y^-, \hat{y}^- \in \widetilde{E}^{-\mathbb{N}_0}$*,*

$$\mathrm{P}_{y^-}(A) \le \mathrm{E}\left[\exp\left\{\varphi\Big(0, \Big(T_m, T_m + \sum_{k=m+1}^{n} |y_k|\Big]\Big)\right\} \mathrm{P}_{\hat{y}^-}(A \mid T_m)\right]. \tag{2.11}$$

Proof. Using Definition 2.2(a), we have

$$\begin{aligned}
\mathrm{P}_{y^-}(A) &= \mathrm{E}\left[\nu_{\kappa(y^-)}\Big(X_{\big(T_m, T_m+\sum_{k=m+1}^n |y_k|\big]} = \kappa(y_{(m,n]})\Big) \prod_{k=m+1}^{n} \varrho(|y_k|)\right] \\
&\le \mathrm{E}\Bigg[\exp\left\{\varphi\left(0, \Big(T_m, T_m + \sum_{k=m+1}^{n} |y_k|\Big]\right)\right\} \nu_{\kappa(\hat{y}^-)}\Big(X_{\big(T_m, T_m+\sum_{k=m+1}^n |y_k|\big]} = \kappa(y_{(m,n]})\Big) \\
&\qquad\qquad \times \prod_{k=m+1}^{n} \varrho(|y_k|)\Bigg] \\
&= \mathrm{E}\left[\exp\left\{\varphi\left(0, \Big(T_m, T_m + \sum_{k=m+1}^{n} |y_k|\Big]\right)\right\} \mathrm{P}_{\hat{y}^-}(A \mid T_m)\right].
\end{aligned} \tag{2.12}$$

□

Lemma 2.4. *For all* $k \in \mathbb{N}_0, \ell \in \mathbb{N}$,

$$\varphi(k, \ell) \le \sum_{m=0}^{\ell-1} \varphi(k+m), \tag{2.13}$$

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

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

$$\varphi(m, k+\ell) \le \varphi(m, k) + \varphi(m+k, \ell), \tag{2.14}$$

which yields the claim via iteration. To prove (2.14), 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

$$A_{(0,k+\ell]} = \{X_{(0,k+\ell]} = x_{(0,k+\ell]}\}, \quad A_{(0,k]} = \{X_{(0,k]} = x_{(0,k]}\}, \quad A_{(k,k+\ell]} = \{X_{(k,k+\ell]} = x_{(k,k+\ell]}\}. \tag{2.15}$$

Estimate

$$\begin{aligned}
\nu_{x^-}(A_{(0,k+\ell]}) &= \nu_{x^-}(A_{(0,k]})\, \nu_{x^- x_{(0,k]}}(A_{(k,k+\ell]}) \\
&\le e^{\varphi(m,k)}\, \nu_{\hat{x}^-}(A_{(0,k]})\, e^{\varphi(m+k,\ell)}\, \nu_{\hat{x}^- x_{(0,k]}}(A_{(k,k+\ell]}) \\
&= e^{\varphi(m,k)+\varphi(m+k,\ell)}\, \nu_{\hat{x}^-}(A_{(0,k+\ell]}),
\end{aligned} \tag{2.16}$$

where $\hat{x}^- x_{(0,k]}$ is the concatenation of $\hat{x}^-$ and $x_{(0,k]}$. Insert this estimate into (2.2) and take the supremum over $x_{(0,k+\ell]}$ and $x^-, \hat{x}^-$ to get (2.14). □

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*:

$$\text{(SV)} \qquad \sum_{n \in \mathbb{N}_0} \varphi(n) < \infty. \tag{2.17}$$

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 $n \geq m$), 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 (see Berbee [1]).

(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 renewal process $(A_k)_{k\in\mathbb{Z}}$ on $\mathbb{N}_0$ with the following transition probabilities:

$$\mathrm{P}(A_1 = n+1 \mid A_0 = n) = \frac{\sum_{\ell>n+1} p(\ell)}{\sum_{\ell>n} p(\ell)}, \quad \mathrm{P}(A_1 = 0 \mid A_0 = n) = \frac{p(n+1)}{\sum_{\ell>n} p(\ell)}, \quad n \in \mathbb{N}_0. \tag{2.18}$$

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

$$\varphi(1) \geq \log \nu_x(X_1 = 1) - \log \nu_{x[n]}(X_1 = 1) = \log p(1) - \log\left(\frac{p(n+1)}{\sum_{\ell>n} p(\ell)}\right). \tag{2.19}$$

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 and continuous-dependence properties:*

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

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

(CD) *For all bounded continuous functions* $f \colon \widetilde{E}^{\mathbb{Z}} \mapsto \mathbb{R}$,

$$y^- \mapsto \int_{y_\mathbb{Z}\in\widetilde{E}^\mathbb{Z}} f(y_\mathbb{Z})\, \mathrm{dP}_{y^-}(y_\mathbb{Z}) \text{ is continuous.} \tag{3.1}$$

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*

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

where $Q_{y^-}|_1$ *and* $\mathrm{P}_{y^-}|_1$ *are the one-word marginals of* Q_{y^-} *and* P_{y^-} *(i.e., of* Q *and* P *conditional on* y^-*).*

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

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

where $h(\,\cdot\,|\,\cdot\,)$ denotes relative entropy. 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. Theorem 1.1 follows by an application of Proposition 3.1, which is a rewriting of Theorem 2.1 in Orey and Pelikan [14]. The state space in Orey and Pelikan [14] is assumed to be compact, which is not the case for $\widetilde{E}$ under the discrete topology. The non-compact case is treated by Orey [13, Theorem 5.11]. Conditions 5.8 and Eq. (3.3) in Orey [13] correspond respectively to Conditions (RM) and (CD) in this paper. The condition in Eq. (3.2) of Orey [13] is implied by Orey [13, Theorem 3.4], which holds by choosing the sequence of truncated state spaces $\widetilde{E}_{(\ell_n)} = \bigcup_{1\le k\le \ell_n} E^k$, where ℓ_n is any strictly increasing sequence of integers satisfying $\mathrm{P}(T_1 > \ell_n) \le 2^{-n}$. First we check that P satisfies (RM). From Lemma 2.3 and the fact that $\ell \mapsto \varphi(0,\ell)$ is non-decreasing, we get $\mathrm{P}_{y^-}(A) \le e^{\varphi(0,\infty)}\mathrm{P}_{\hat{y}^-}(A)$. Hence Definition 2.2(b) gives $\psi((m,n]) \le \varphi(0,\infty)$ for all $0 \le m < n$. From Lemma 2.4 we get

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

Hence, if (SV) holds, then (RM) holds for $m(n) = 0$. Next we check that P satisfies (CD). Note that it is enough to consider bounded and continuous $f\colon\ \widetilde{E}^{\mathbb{N}} \mapsto \mathbb{R}$, because

$$\mathrm{P}_{y^-}(Y_{\mathbb{Z}} \in d\bar{y}_{\mathbb{Z}}) = 1_{\{\bar{y}^- = y^-\}}\mathrm{P}_{y^-}(Y_{\mathbb{N}} \in d\bar{y}_{\mathbb{N}}). \tag{3.5}$$

Choose y^- and $\hat{y}^-$ such that $y^-_{(-k,0]} = \hat{y}^-_{(-k,0]}$ for some $k \in \mathbb{N}$. From Lemmas 2.3–2.4 we obtain

$$\int_{y_{\mathbb{N}}\in\widetilde{E}^{\mathbb{N}}} f(y_{\mathbb{N}})\mathrm{P}_{y^-}(Y_{\mathbb{N}} \in dy_{\mathbb{N}}) \le e^{\sum_{\ell\ge k+1}\varphi(\ell)} \int_{y_{\mathbb{N}}\in\widetilde{E}^{\mathbb{N}}} f(y_{\mathbb{N}})\mathrm{P}_{\hat{y}^-}(Y_{\mathbb{N}} \in dy_{\mathbb{N}}). \tag{3.6}$$

The same holds with y^- and $\hat{y}^-$ interchanged. Under (SV), $\lim_{k\to\infty}\sum_{\ell\ge k+1}\varphi(\ell) = 0$, which proves (CD). □

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. ν. We focus on those ingredients where the lack of independence of ν 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 $\varphi(0,k) < \infty$ and there is nothing to prove, so we can assume $\nu_{\hat{x}^-}(A) > 0$. Then, by the definition of $\varphi(0,k)$ and Lemma 2.4,

$$C(\varphi)^{-1} \le e^{-\varphi(0,k)} \le \frac{\nu_{x^-}(A)}{\nu_{\hat{x}^-}(A)} \le e^{\varphi(0,k)} \le C(\varphi). \tag{4.4}$$

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 $\widetilde{x}^- x^-_{(-n,0]}$ is the concatenation of $\widetilde{x}^-$ and $x^-_{(-n,0]}$, and the inequality uses (4.2). The reverse inequality is obtained in a similar manner. □

Lemma 4.2. *Let $m \in \mathbb{N}$, and let $(i_1, \ldots, i_m)$, $(j_1, \ldots, j_m)$ be two collections of integers satisfying $i_1 < j_1 \le i_2 < j_2 \le \ldots < i_{m-1} < j_{m-1} \le i_m < j_m$. For $1 \le k \le m$, let $A_k \in \mathcal{F}_{(i_k,j_k]}$ and $p_k = \nu(A_k)$. Suppose that ν 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 $x^- \in \mathrm{E}^{-\mathbb{N}_0}$,

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

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

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

□

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*

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

If ν satisfies condition (SV)*, then ν-a.s.,*

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

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 of times $(\sigma_n)_{n\in\mathbb{Z}}$ cuts a sequence of blocks $B = (B_n)_{n\in\mathbb{Z}}$ out of the letter sequence X given by

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

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

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

Define the process $B^\star = (B^\star_n)_{n\in\mathbb{Z}}$ 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

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

where xy is the concatenation of x and y. For the collection $(\mathrm{P}^\star_A(\cdot|x), x \in E^{-\mathbb{N}_0})$ to be a proper transition kernel, σ_1 must be ν_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 σ_1 is ν-a.s. finite. But since ν and $(\nu_x)_{x\in E^{-\mathbb{N}_0}}$ are equivalent under condition (SV) (note that $C(\varphi)^{-1}\nu(\cdot)\leq\nu_x(\cdot)\leq C(\varphi)\nu(\cdot)$ as a consequence of (4.2) in Lemma 4.1), σ_1 indeed is ν_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

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

where π 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 $\mathrm{P}_A^\star$. It is easy to check that $\mathrm{P}_A^\star$ 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. $\sigma(B_{(-\infty,0]})$ and $\sigma(B_{(0,\infty)})$, respectively,

$$C(\varphi)^{-1}\mathrm{P}_A(\mathcal{A})\mathrm{P}_A(\mathcal{B})\leq\mathrm{P}_A(\mathcal{A}\cap\mathcal{B})\leq C(\varphi)\mathrm{P}_A(\mathcal{A})\mathrm{P}_A(\mathcal{B}), \tag{4.15}$$

where P_A is the law of B induced by $\mathrm{P}_A^\star$. Therefore P_A is Weak Bernoulli (Ledrappier [12]), and hence is ergodic. Thus, we have

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

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

$$E_{\mathrm{P}_A}(\sigma_1-\sigma_0)=\int E_{\nu_{x^-}}(\sigma_1-\sigma_0)\mathrm{dP}_A(x^-)\leq C(\varphi)E_{\nu_{\hat{x}^-}}(\sigma_1-\sigma_0), \tag{4.17}$$

which gives $E_{\mathrm{P}_A}(\sigma_1-\sigma_0)\leq C(\varphi)E_\nu(\sigma_1-\sigma_0)$ and completes the proof. □

4.3 Decomposition of relative entropy

Write $H(Q)$ to denote the specific entropy of Q. Let

$$\begin{aligned}\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{Z}})&=\{\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})\colon\ m_Q<\infty\},\\ \mathcal{P}^{\mathrm{inv,erg,fin}}(\widetilde{E}^{\mathbb{Z}})&=\{\mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}})\colon\ \text{Q is ergodic},\ m_Q<\infty\}.\end{aligned} \tag{4.18}$$

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

$$\begin{aligned}H(Q\mid\mathrm{P})&=-H(Q)-E_Q[\log\varrho(\tau_1)]-m_QE_{\Psi_Q}[\log\nu_{X_{(-\infty,0]}}(X_1)],\\ H(\Psi_Q\mid\nu)&=-H(\Psi_Q)-E_{\Psi_Q}[\log\nu_{X_{(-\infty,0]}}(X_1)].\end{aligned} \tag{4.19}$$

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

$$E_Q[\log\mathrm{P}_{Y_{(-\infty,0]}}(Y_1)]=E_Q[\log\varrho(\tau_1)]+E_Q[\log\nu_{X_{(-\infty,0]}}(X_{(0,\tau_1]})] \tag{4.20}$$

and (recall (1.3))

$$E_Q[\log\nu_{X_{(-\infty,0]}}(X_{(0,\tau_1]})]=E_Q\left[\sum_{k=0}^{\tau_1-1}\log\nu_{X_{(-\infty,k]}}(X_{k+1})\right]=m_QE_{\psi_Q}[\log\nu_{X_{(-\infty,0]}}(X_1)], \tag{4.21}$$

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 (4.19) are finite: (i) $|E|<\infty$ and $0\leq H(Q)\leq m_Q\log|E|-E_Q(\log Q(\tau_1=n))<\infty$ for all $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{Z}})$; (ii) $-E_Q[\log\varrho(\tau_1)]<\infty$ for all $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{Z}})$ because ϱ satisfies (1.5); (iii) $\varphi(0)<\infty$.

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

$$\lim_{n\to\infty} \frac{1}{n} \log \nu(X_{(0,T_n]}) = m_Q E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)] \quad Q-a.s. \tag{4.22}$$

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

$$C(\varphi)^{-1}\nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) \le \nu(X_{(0,T_n]}) \le C(\varphi)\nu_{X_{(-\infty,0]}}(X_{(0,T_n]}). \tag{4.23}$$

Next write

$$\log \nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) = \sum_{k=0}^{T_n-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1}) = \sum_{i=0}^{n-1}\sum_{k=T_i}^{T_{i+1}-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1}). \tag{4.24}$$

Use (4.24) and the ergodicity of Q to obtain, for Q-a.s. Y,

$$\lim_{n\to\infty} \frac{1}{n} \log \nu_{X_{(-\infty,0]}}(X_{(0,T_n]}) = E_Q\left[\sum_{k=0}^{\tau_1-1} \log \nu_{X_{(-\infty,k]}}(X_{k+1})\right] = m_Q E_{\Psi_Q}[\log \nu_{X_{(-\infty,0]}}(X_1)]. \tag{4.25}$$

Combine (4.23–4.25) 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. ν. 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 [3, Eqs.(1.25–1.26)] by the relations in Lemma 4.4. These relations allow us to decompose I^{que} as a sum of three terms that appear in the proofs of the lower bound and the upper bound of the LDP as given in [3, Section 1.3].

Upper bound. The upper bound in [3, Proposition 3.1] is proved by first restricting to $Q \in \mathcal{P}^{\mathrm{inv,erg,fin}}(\widetilde{E}^{\mathbb{Z}})$. The event in [3, Eq. (3.4)] is used to define a suitable neighbourhood of Q. In that equation only the fourth line has to be replaced by

$$\left\{\frac{1}{M} \log \nu(X_{(0,T_M]}) \in m_Q E_{\Psi_Q}\big[\log \nu_{X_{(-\infty,0]}}(X_1)\big] + [-\varepsilon_1, \varepsilon_1]\right\}. \tag{4.26}$$

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

$$\mathrm{P}\left(\cap_{1\le k\le n}\{A_k = a_k\}\right) \le [C(\varphi)p]^{\sum_{1\le k\le n} a_k}, \tag{4.27}$$

where A_k is the indicator defined in [3, Eqs.(3.36–3.37)], and $a_k \in \{0,1\}$ labels whether or not at some specific location of the letter sequence X there is a string of letters arising from the concatenation of *Q-typical* words (see [3, Eq (3.5–3.6)]). The inequality in (4.27) is proved via Lemma 4.2 and allows us to use [3, Lemma 2.1], which controls the occurrence of certain patterns in X. We are then able to complete the argument in [3, Section 3.4].

A further step consists in removing the ergodicity assumption on Q, as was done in [3, Section 3.5]. The arguments in [3, Sections 3.5.1-3.5.4] carry over verbatim, since they do not require that the reference measure is product. However, the proof of [3, Eq. (3.140)] (see [3, Section 3.5.5 B]) uses the i.i.d. structure of the reference measure, but the proof can be adapted as follows: (i) construct the $(A_{i,r})_{i\in\mathbb{N},1\le r\le R}$ as in [3, Section 3.5.5 B]; (ii) use Lemma 4.2 to derive an extended version of (4.27), namely,

$$\mathrm{P}(A_{i_1,r_1}=1,\dots,A_{i_n,r_n}=1)\le C(\varphi)^{n-1}\prod_{j=1}^{n}p_{r_j}, \tag{4.28}$$

where the p_r's are analogous to [3, Eq. (3.144)]. We may then use [3, Lemma 3.3] to complete the argument. Note that the finite pre-factor $C(\varphi)$ is harmless in the limit.

Lower Bound. The lower bound in [3, Proposition 4.1] is proved by bounding from below the probability that R_n lies in a neighbourhood of some $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\widetilde{E}^{\mathbb{Z}})$. When Q is ergodic we can use the same strategy as in [3] (namely, by jumping to Q*-typical* substrings of letters), but a modification is needed to go from [3, Eq.(4.7)] to [3, Eq.(4.8)], since the increments of the $\sigma_\ell^{(M)}$, $\ell\in\mathbb{N}$, defined in [3, Eq.(4.6)] are no longer i.i.d. This can again be handled with the help of Lemma 4.3. Note that the extra constant $\log C(\varphi)$ is killed when letting $M\to\infty$ in [3, Eq. (4.8)]. Using ergodic decomposition, we get rid of the ergodicity assumption on Q, exactly as in [3, Eqs. (4.9–4.11)]. □

Truncation limits. The argument in [3, Section 3] also uses [3, Lemma A.1], which in our case is (1.8). The proof in [3, Appendix A] carries over almost verbatim: in [3, Eqs. (A.3–A.4) and (A.13–A.14)], now use

$$\mathrm{E}_{\psi_{[Q]_{\mathrm{tr}}}}[\log\nu_{X_{(-\infty,0]}}(X_1)]\overset{\mathrm{tr}\to\infty}{\longrightarrow}\mathrm{E}_{\psi_{[Q]}}[\log\nu_{X_{(-\infty,0]}}(X_1)], \tag{4.29}$$

for which it is enough to prove

$$\left|\mathrm{E}_Q\left[\sum_{i=1}^{\tau_1\wedge\mathrm{tr}}\log\nu_{X_{(-\infty,i)}}(X_i)\right]-\mathrm{E}_{[Q]_{\mathrm{tr}}}\left[\sum_{i=1}^{\tau_1\wedge\mathrm{tr}}\log\nu_{X_{(-\infty,i)}}(X_i)\right]\right|\le\sum_{\ell\ge\mathrm{tr}}\varphi(\ell). \tag{4.30}$$

Note that the last sum tends to zero as $\mathrm{tr}\to\infty$ because of (SV). Equation (4.30) can be proven via a coupling argument. More precisely, we may rewrite

$$\mathrm{E}_{[Q]_{\mathrm{tr}}}\left[\sum_{i=1}^{\tau_1\wedge\mathrm{tr}}\log\nu_{X_{(-\infty,i)}}(X_i)\right]=\mathrm{E}_Q\left[\sum_{i=1}^{\tau_1\wedge\mathrm{tr}}\log\nu_{X^{(\mathrm{tr})}_{(-\infty,i)}}(X_i^{(\mathrm{tr})})\right], \tag{4.31}$$

where $X^{(\mathrm{tr})}$ is defined as the concatenation of the sequence of truncated words. Thus, (4.30) comes from the fact that $X_{(-\mathrm{tr},\mathrm{tr}]}=X^{(\mathrm{tr})}_{(-\mathrm{tr},\mathrm{tr}]}$.

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 (2.8) 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.

5.1 Preparatory lemmas

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

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

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

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

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, as can be easily checked by hand.

The following lemma generalizes (4.2) in Lemma 4.1.

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

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

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

$$\begin{aligned} \frac{\mathrm{d}\nu_{x^-}|_n}{\mathrm{d}\nu_{\hat{x}^-}|_n}(x_1,\dots,x_n) &= \frac{\mathrm{d}\nu_{x^-}|_1}{\mathrm{d}\nu_{\hat{x}^-}|_1}(x_1) \times \frac{\mathrm{d}\nu_{x^-x_1}|_1}{\mathrm{d}\nu_{\hat{x}^-x_1}|_1}(x_2) \times \cdots \times \frac{\mathrm{d}\nu_{x^-x_1\cdots x_{n-1}}|_1}{\mathrm{d}\nu_{\hat{x}^-x_1\cdots x_{n-1}}|_1}(x_n) \\ &\leq \exp[\varphi(0)+\varphi(1)+\cdots+\varphi(n-1)] \leq C(\varphi), \end{aligned} \tag{5.7}$$

where $\nu_{x^-}|_n$ denotes the n-letter marginal conditional on x^-. This 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

$$X_n^{(c)} = \sum_{i=1}^{c} i\, 1_{\{X_n\in E_i\}}. \tag{5.8}$$

The following lemma is another consequence of (SV). Let $(\mathrm{P}^{(c)}, Q^{(c)})$ and $(\nu^{(c)}, \Psi_Q^{(c)})$ denote the coarse-grained versions of (P, Q) and (ν, Ψ_Q).

Lemma 5.2. *Under condition* (SV)*,*

$$H(Q \mid \mathrm{P}) = \sup_{n\in\mathbb{N}} \frac{1}{n}\Big\{h\big(Q(Y_{(0,n]}\in\cdot) \mid \mathrm{P}(Y_{(0,n]}\in\cdot)\big) - \log C(\varphi)\Big\}, \tag{5.9}$$

and the supremum is also a limit. The same result holds when (P, Q) *is replaced by* $(\mathrm{P}^{(c)}, Q^{(c)})$ *or* (ν, Ψ_Q) *or* $(\nu^{(c)}, \Psi_Q^{(c)})$*.*

Proof. We prove the result for (P, Q). The other cases are similar. For $n\in\mathbb{N}$, let $\mathcal{B}(\widetilde{E}^n)$ be the set of bounded measurable functions on $\widetilde{E}^n$. From the variational characterization of relative entropy (see Dembo and Zeitouni [4, Lemma 6.2.13]), we get that for all $n, m\in\mathbb{N}$,

$$\begin{aligned}
&h\big(Q(Y_{(0,n+m]}\in\cdot) \mid \mathrm{P}(Y_{(0,n+m]}\in\cdot)\big) \\
&= \sup_{f\in\mathcal{B}(\widetilde{E}^{n+m})} \Big\{E_Q[f(Y_{(0,n+m]})] - \log E_{\mathrm{P}}\Big[e^{f(Y_{(0,n+m]})}\Big]\Big\} \\
&\geq \sup_{\substack{f_1\in\mathcal{B}(\widetilde{E}^{n})\\ f_2\in\mathcal{B}(\widetilde{E}^{m})}} \Big\{E_Q\big[f_1(Y_{(0,n]}) + f_2(Y_{(n,n+m]})\big] - \log E_{\mathrm{P}}\Big[e^{f_1(Y_{(0,n]})}e^{f_2(Y_{(n,n+m]})}\Big]\Big\}.
\end{aligned} \tag{5.10}$$

Using the decoupling inequality of Lemma 5.1 and the stationarity of P and Q, we may bound the right-hand side from below by

$$\begin{aligned}
&\sup_{f\in\mathcal{B}(\widetilde{E}^{n})} \Big\{E_Q[f(Y_{(0,n]})] - \log E_{\mathrm{P}}\Big[e^{f(Y_{(0,n]})}\Big]\Big\} \\
&\quad + \sup_{f\in\mathcal{B}(\widetilde{E}^{m})} \Big\{E_Q[f(Y_{(0,m]})] - \log E_{\mathrm{P}}\Big[e^{f(Y_{(0,m]})}\Big]\Big\} - \log C(\varphi) \\
&= h\big(Q(Y_{(0,n]}\in\cdot) \mid \mathrm{P}(Y_{(0,n]}\in\cdot)\big) + h\big(Q(Y_{(0,m]}\in\cdot) \mid \mathrm{P}(Y_{(0,m]}\in\cdot)\big) - \log C(\varphi).
\end{aligned} \tag{5.11}$$

The claim now follows from the superadditivity of the sequence $\{h(Q(Y_{(0,n]}\in\cdot) \mid \mathrm{P}(Y_{(0,n]}\in\cdot)) - \log C(\varphi)\}_{n\in\mathbb{N}}$. □

5.2 Proof

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

Annealed LDP. Lemma 5.1 shows that under condition (SV)Lemmas 2.3–2.4 carry over from finite letters to Polish letters. Condition (CD) in Proposition 3.1 is also implied by condition (SV), by an argument similar to the case of finite letters. Therefore the ratio-mixing and continuous dependence properties of Orey and Pelikan [14] again yields the annealed LDP.

Quenched LDP. The proof comes in 4 steps.

1. We first show that Lemmas 4.2–4.5 carry over to the coarse-grained process $X^{(c)}$ defined in (5.8) for every $c\in\mathbb{N}$. However, for $Q\in\mathcal{P}^{\mathrm{inv,fin}}(\mathcal{E}_c^{\mathbb{Z}})$ we cannot directly replace X by $X^{(c)}$ or ν by $\nu^{(c)}$ in the expectation

$$E_{\Psi_Q}\left[\log \nu^{(c)}_{X^{(c)}_{(-\infty,0]}}(X_1^{(c)})\right], \tag{5.12}$$

which is present in both Lemma 4.4 and Lemma 4.5. The reason is that the probability distribution under the logarithm is only a version of the conditional probability $\nu^{(c)}(\cdot|X^{(c)}_{(-\infty,0]})$, and if $\Psi_Q\neq\nu^{(c)}$, then this conditional probability can take any value on a set of Ψ_Q-probability 1. One way out is to consider a modified version of these lemmas, where the expectation in (5.12) is replaced by

$$\lim_{N\to\infty} \frac{1}{M} E_{\Psi_Q}\big[\log \nu(X^{(c)}_{(0,M]})\big] \Big(= \lim_{M\to\infty} E_{\Psi_Q}\Big[\log \nu(X^{(c)}_1 | X^{(c)}_{(-M,0]})\Big]\Big). \tag{5.13}$$

Note that the limit exists, since

$$\frac{1}{M} E_{\Psi_Q}\big[\log \nu(X^{(c)}_{(0,M]})\big] = -\frac{1}{M} h\big(\Psi_Q(X^{(c)}_{(0,M]} \in \cdot)\big) - \frac{1}{M} h\big(\Psi_Q(X^{(c)}_{(0,M]} \in \cdot) | \nu(X^{(c)}_{(0,M]} \in \cdot)\big), \tag{5.14}$$

and both terms in the right-hand side converge (by standard arguments for the first term and by Lemma 5.2 for the second term). The limit equals $-H(\Psi_Q) - H(\Psi_Q|\nu^{(c)})$, which is precisely what we stated as the modified version of Lemma 4.4 (second line).

2. To prove the modified version of Lemma 4.5 and of the first line in Lemma 4.4, we show that, Q-a.s. and in $L^1(Q)$, $\frac{1}{n}\log\nu(X^{(c)}_{(0,T_n]})$ converges to

$$m_Q \lim_{M\to\infty} \frac{1}{M} E_{\Psi_Q}[\log \nu(X^{(c)}_{(0,M]})].$$

Note that the limit exists because of condition (SV) and super-additivity. Choose $n = kM$ with $k, M \in \mathbb{N}$. Using Lemma 5.1 k times, we may write

$$\left| \frac{1}{n} \log \nu(X^{(c)}_{(0,T_n]}) - \frac{1}{kM} \sum_{i=1}^{k} \log \nu(X^{(c)}_{(T_{(i-1)M}, T_{iM}]}) \right| \leq \frac{\log C(\varphi)}{M}. \tag{5.15}$$

Letting $k \to \infty$ and using Birkhoff's ergodic theorem, we obtain

$$\left| \lim_{n\to\infty} \frac{1}{n} \log \nu(X^{(c)}_{(0,T_n]}) - \frac{1}{M} E_Q[\log \nu(X^{(c)}_{(0,T_M]})] \right| \leq \frac{\log C(\varphi)}{M}. \tag{5.16}$$

It remains to use the law of large numbers for T_M/M under $Q \in \mathcal{P}^{\mathrm{inv,erg,fin}}(\widetilde{E}^{\mathbb{Z}})$ (recall (1.1)), and the fact that $\frac{1}{n} E_Q(\log \nu(X^{(c)}_{(0,n]}))$ has the same limit has $\frac{1}{n} E_{\Psi_Q}(\log \nu(X^{(c)}_{(0,n]}))$.

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 $\mathcal{E}_c = \{E_1, \ldots, E_c\}$, $c \in \mathbb{N}$, 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

$$I^{\mathrm{que}}(Q) = \sup_{c\in\mathbb{N}} I^{\mathrm{que}}_c(Q^{(c)}), \quad Q \in \mathcal{P}^{\mathrm{inv}}(\widetilde{E}^{\mathbb{Z}}), \tag{5.17}$$

where $Q^{(c)}$ is the coarse-graining of Q, and I^{que}_c is the coarse-grained rate function. The argument in [3, Section 8] can be adapted to show, with the help of Lemma 5.2, 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.) □

References

[1] H. Berbee, *Chains with infinite connections: uniqueness and Markov representation*, Probab. Th. Rel. Fields 76 (1987) 243–253. MR-0906777

[2] M. Birkner, *Conditional large deviations for a sequence of words*, Stoch. Proc. Appl. 118 (2008) 703–729. MR-2411517

[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. MR-2678894

[4] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications* (2nd Ed.), Springer, 1998. MR-1619036

[5] J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Academic, Boston (1989). MR-0997938

[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. MR-0690656

[7] M.D. Donsker and S.R.S. Varadhan, *Large deviations for stationary Gaussian processes*, Comm. Math. Phys. 97 (1985) 187–210. MR-0782966

[8] P.R. Halmos, *Lectures on ergodic theory*, Chelsea Publishing Co., New York, 1960. MR-0111817

[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. MR-1482994

[10] A. Johansson and A. Öberg, *Square summability of variations of g-functions and uniqueness of g-measures*, Math. Res. Lett. 10 (2003) 587–601. MR-2024717

[11] A. Johansson, A. Öberg and M. Pollicott, *Countable state shifts and uniqueness of g-measures*, Amer. J. Math. 129 (2007) 1501–1511. MR-2369887

[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. MR-0507548

[13] S. Orey, *Large deviations in ergodic theory*, in: *Seminar on Stochastic Processes, Evanston, Illinois, 1984*, Progr. Probab. Statist. 9 (1984) 195—249. MR-0896730

[14] S. Orey and S. Pelikan, *Large deviation principles for stationary processes*, Ann. Probab. 16 (1988) 1481–1495. MR-0958198

[15] K.S. Parthasarathy, *Probability Measures on Metric Spaces*, Vol. 352, American Mathematical Society, 1967.

Acknowledgments. The authors are supported by ERC Advanced Grant 267356 VARIS, and are grateful to an anonymous referee for careful reading and detailed remarks.