Conditional large deviations for a sequence of words

Matthias Birkner

Weierstraß-Institut für Angewandte Analysis und Stochastik
Mohrenstraße 39
10117 Berlin
Germany

Abstract

Cut an i.i.d. sequence (X_i) of 'letters' into 'words' according to an independent renewal process. Then one obtains an i.i.d. sequence of words, and thus the level three large deviation behaviour of this sequence of words is governed by the specific relative entropy. We consider the corresponding problem for the *conditional* empirical process of words, where one conditions on a typical underlying (X_i). We find that if the tails of the word lengths decay exponentially, the large deviations under the conditional distribution are almost surely again governed by the specific relative entropy, but the set of attainable limits is restricted.

We indicate potential applications of such a conditional LDP to the computation of the quenched free energy for directed polymer models with random disorder.

Key words: Conditional process level large deviations, quenched free energy
2000 MSC: 60F10, 60G10, 82D60

1 Scenario and main result

Let E be a countable set ('letters' or 'symbols'), $\nu \in \mathcal{P}(E)$ a probability measure on E with $\nu(x) > 0$ for all $x \in E$. Let $(X_i)_{i\in\mathbb{N}}$ be an i.i.d.-ν sequence, $(\tau_j)_{j\in\mathbb{N}}$ an independent i.i.d.-ρ sequence with values in $\mathbb{N}$. We assume that ρ has exponentially bounded tails

$$\exists\, C, \lambda \,:\, \forall\, n \,:\, \rho_n \le C \exp(-\lambda\, n) \tag{1}$$

Email address: birkner@wias-berlin.de (Matthias Birkner).

15th May 2007

and that the τs generate an aperiodic renewal process, i.e. $\gcd\{i : \rho_i > 0\} = 1$.

Cut out the X-sequence according to τ: Put $T_0 := 0$, $T_i := T_{i-1} + \tau_i$ for $i \geq 1$,

$$Y^i = \Big(X_{T_{i-1}+1}, X_{T_{i-1}+2}, \dots, X_{T_i}\Big), \quad i \in \mathbb{N}, \tag{2}$$

with values in $\tilde{E} = \cup_{k=1}^{\infty} E^k$ ('words'). We write $|y| = k$ for the 'length' of $y = (y_1, \dots, y_k) \in \tilde{E}$. By the independence properties of the ingredients, $Y = (Y^i)_{i=1,2,\dots}$ is then an i.i.d. sequence with marginal distribution

$$q^0\Big((x_1, \dots, x_k)\Big) := \mathbb{P}(Y^1 = (x_1, \dots, x_k)) = \rho_k \prod_{i=1}^{k} \nu(x_i). \tag{3}$$

For a sequence (Y^i) with values in $\tilde{E}^{\mathbb{N}}$ we write $L_i = |Y^i|$ for the length of the i-th word (in the present scenario, we have $L_i = \tau_i$, but it will be convenient to have a variable for word lengths also if Y does not arise from a construction with a τ-sequence). Note that we have a (left) shift $\theta : E^{\mathbb{N}} \to E^{\mathbb{N}}$ on letter sequences and a (left) shift $\tilde{\theta} : \tilde{E}^{\mathbb{N}} \to \tilde{E}^{\mathbb{N}}$ on word sequences. Let

$$R_N := \frac{1}{N} \sum_{i=0}^{N-1} \delta_{\tilde{\theta}^i\left((Y^1, \dots, Y^N)^{\mathrm{per}}\right)} \tag{4}$$

be the empirical distribution process of the words with values in $\mathcal{P}(\tilde{E}^{\mathbb{N}})$, the probability measures on sequences of words. Here, $\Big(y^1, \dots, y^m\Big)^{\mathrm{per}}$ denotes the periodic extension of $(y^1, \dots, y^m) \in \tilde{E}^m$ to an element of $\tilde{E}^{\mathbb{N}}$.

The sets E and $\tilde{E}$ are countable, so they are Polish spaces with the discrete metric. Then $E^{\mathbb{N}}$ and $\tilde{E}^{\mathbb{N}}$ are again metric spaces e.g. via

$$d_{A^{\mathbb{N}}}\Big((z_1, z_2, \dots), (z_1', z_2', \dots)\Big) := \sum_{n=1}^{\infty} 2^{-|n|}\Big(d_A(z_n, z_n') \wedge 1\Big)$$

for $A = E$ or $A = \tilde{E}$. This metric induces the product topology on $E^{\mathbb{N}}$ resp. $\tilde{E}^{\mathbb{N}}$. We equip $\mathcal{P}(\tilde{E}^{\mathbb{N}})$ with the topology of weak convergence. Write $\mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ for the shift invariant probability measures on $\tilde{E}^{\mathbb{N}}$, and $\mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ for the set of ($\tilde{\theta}$-shift) ergodic probability measures on $\tilde{E}^{\mathbb{N}}$. Note that $\mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ is a closed subset of $\mathcal{P}(\tilde{E}^{\mathbb{N}})$.

It is well known that the family of distributions $\mathscr{L}(R_N)$ satisfies a large deviation principle, the 'good' rate function is given by

$$H(Q|Q^0) = \lim_{N\to\infty} \frac{1}{N} h\Big(Q|_{\mathscr{F}_N} \,\Big|\, Q^0|_{\mathscr{F}_N}\Big), \tag{5}$$

the specific relative entropy with respect to $Q^0 := \mathscr{L}(Y) = (q^0)^{\otimes\mathbb{N}}$, see e.g. [6], [5], Chap. IX or [3], Chap. 6.5. Here $\mathscr{F}_N = \sigma(Y_1, \dots, Y_N)$, $Q|_{\mathscr{F}_N}$ is Q restricted to the first N words, and for probability measures μ, μ' on some measurable space,

$$h(\mu|\mu') = \begin{cases} \int \log \frac{d\mu}{d\mu'}\, d\mu & \text{if } \mu \text{ is absolutely continuous w.r.t. } \mu', \\ \infty & \text{otherwise,} \end{cases}$$

denotes the relative entropy of μ with respect to μ'. Our aim is to understand the almost sure large deviation behaviour of the family of random probability distributions

$$\mathscr{L}(R_N \,|\, X).$$

As $\mathcal{P}(E^{\mathbb{N}})$ and $\mathcal{P}(\tilde{E}^{\mathbb{N}})$ are Polish, we can and shall think in the following of a family of regular conditional distributions $\mathbb{P}(R_N \in \cdot\,|X)$. In fact, it can be given explicitly as follows

$$\begin{aligned} &\mathscr{L}(R_N \,|\, X) \qquad (6) \\ &\quad = \sum_{j_1<\cdots<j_N} \prod_{i=1}^{N} \rho(j_i - j_{i-1}) \sum_{k=0}^{N-1} \frac{1}{N} \delta_{\tilde{\theta}^k \left(X|_{[1\dots j_1]}, X|_{[j_1+1\dots j_2]}, X|_{[j_{N-1}+1\dots j_N]}\right)^{\mathrm{per}}}, \end{aligned}$$

where for $x = (x_i) \in E^{\mathbb{N}}$, $k < \ell$

$$x|_{[k\dots\ell]} := (x_k, x_{k+1}, \dots, x_\ell) \in \tilde{E}. \qquad (7)$$

Quantities involving the conditional expectation of exponential functionals of R_N appear naturally in the computation of the quenched free energy for polymer models in disordered media. In particular, the asymptotic evaluation of the free energy can be formulated as a conditional large deviation problem, and variational formulas as in Corollary 1 make the energy-entropy trade-off explicit. This potential application motivated our original interest in the question studied in this note, see Section 2 for more details.

It is natural to invert the cutting by concatenation: Let the concatenation operator $\kappa : \tilde{E}^{\mathbb{N}} \to E^{\mathbb{N}}$ be defined in the obvious way by

$$\kappa\Big((y^1, y^2, y^3, \dots)\Big) = \Big(y^1_1, y^1_2, \dots, y^1_{\ell_1}, y^2_1, y^2_2, \dots, y^2_{\ell_2}, y^3_1, \dots\Big)$$

for $y^i = (y^i_1, \dots, y^i_{\ell_i}) \in \tilde{E}$. For finite sequences of words, $\kappa(y^1, \dots, y^n) \in E^{|y^1|+\cdots+|y^n|}$ is defined analogously.

One can imagine that because of the conditioning, which fixes a *typical* realisation of the X-sequence, the conditional law $\mathscr{L}(R_N \,|\, X)$ feels restrictions, and that some deviations, which are simply exponentially unlikely under the unconditional law, become actually impossible once a typical X is fixed. Let

$$\mathscr{R} := \Big\{ Q \in \mathcal{P}(\tilde{E}^{\mathbb{N}}) : w\!-\!\lim_{L\to\infty} \frac{1}{L} \sum_{j=0}^{L-1} \delta_{\theta^{j}\kappa(Y)} = \nu^{\otimes\mathbb{N}} \quad Q-\text{a.s.} \Big\}, \tag{8}$$

where $w-\lim$ denotes the limit with respect to the weak topology on $\mathcal{P}(E^{\mathbb{N}})$. $Q \in \mathscr{R}$ means that under Q, the concatenation of words has almost surely the same asymptotic statistics as a typical realisation of (X_i). Obviously $Q^0 \in \mathscr{R}$.

Our main result is a full LDP for the (random) family $\mathscr{L}(R_N|X)$, $N \in \mathbb{N}$, it roughly states that under $\mathbb{P}(R_N \in \cdot\,|X)$, only such deviations can be realised which respect the restriction set $\mathscr{R}$.

Theorem 1 *Under Assumption (1), the following events occur with probability one:*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\Big(R_N \in F \,\Big|\, X\Big) \le - \inf_{Q\in F\cap\mathscr{R}\cap\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})} H(Q|Q^0) \qquad \text{for all closed } F \subset \mathcal{P}(\tilde{E}^{\mathbb{N}}), \tag{9}$$

$$\liminf_{N} \frac{1}{N} \log \mathbb{P}\Big(R_N \in G \,\Big|\, X\Big) \ge - \inf_{Q\in G\cap\mathscr{R}\cap\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})} H(Q|Q^0) \qquad \text{for all open } G \subset \mathcal{P}(\tilde{E}^{\mathbb{N}}). \tag{10}$$

A standard application of Varadhan's Lemma yields

Corollary 1 *For any bounded continuous function $\Phi : \tilde{E}^{\mathbb{N}} \to \mathbb{R}$ we have*

$$\begin{aligned} &\lim_{N} \frac{1}{N} \log \mathbb{E}\Big[\exp\Big(N \int \Phi(y) R_N(dy)\Big)\Big| X\Big] \\ &= \sup_{Q\in\mathscr{R}\cap\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})} \Big\{ \int \Phi(y) Q(dy) - H(Q|Q^0) \Big\} \quad \text{a.s.} \end{aligned} \tag{11}$$

Remark 1 The same results hold for the 'non-periodic' flavour of the empirical process,

$$R_N^{\text{non-per}} := \frac{1}{N} \sum_{i=0}^{N-1} \delta_{\tilde{\theta}^{i}Y} \,.$$

Furthermore, the restriction to aperiodic ρ is not severe. If ρ has period $d > 1$, simply consider $E' := E^d$ as a new alphabet.

Remark 2 Theorem 1 does not hold in this form without assumptions on the tails of ρ. In fact, in a situation where ρ_n decays only algebraically, one can probe exponentially (in N, the number of pieces one wants to cut) far ahead into the X-sequence in order to find regions where X looks atypical.

For a concrete example, consider the following scenario: Let (X_i) be i.i.d. Ber(1/2), $\rho_n = C/n^a$, $a > 2$, so $m_\rho := \sum_n n\rho_n < \infty$. Put

$$\sigma_N := \min\{k \in \mathbb{N} : X_k = X_{k+1} = X_{k+[N(m_\rho+\epsilon)]} = 1\}.$$

Let $q^1(x_1, \dots, x_m) := \rho_m \mathbf{1}(x_1 = \cdots = x_m = 1)$, and let $O \subset \mathcal{P}(\tilde{E}^{\mathbb{N}})$ be a (small) neighbourhood of $(q^1)^{\otimes\mathbb{N}}$. Under $(q^1)^{\otimes\mathbb{N}}$, all words consist entirely of 1s. Note that $\log\sigma_N \sim N(m_\rho + \epsilon)\log 2$ by the Erdős-Rényi law and $\mathbb{P}(R_N \in O \,|\, X) \geq e^{-\epsilon N}\rho_{\sigma_N}$ by Lemma 9 below (note that for $Q = (q^1)^{\otimes\mathbb{N}}$, we have $H^c_L(Q) = -\mathbb{E}_Q \log\rho_{L_1}$ in this case, cf Lemma 3) for large enough N, so

$$\liminf_{N\to\infty} \frac{1}{N}\log\mathbb{P}(R_N \in O|X) \geq \liminf_{N\to\infty} \frac{1}{N}\log\rho_{\sigma_N} > -\infty.$$

On the other hand, if (9) held true in this scenario, the answer would have to be $-\infty$, because $(q^1)^{\otimes\mathbb{N}} \notin \mathscr{R}$.

By Lemma 8, (9) will hold with $\mathscr{R}$ replaced by $\overline{\mathscr{R}}$, but in view of Remark 8 in Section 3, this amounts essentially only to the unconditional upper bound, which we expect not to be sharp. The intuitive argument advocated on page 3, that any limiting Q must be built 'on top' of a typical X-sequence, is not valid in general. In fact, when ρ has algebraic tails, there will be a trade-off on the exponential scale between how deep one probes into the fixed X-sequence, which allows to find more atypical regions, and the price for those long jumps. In view of the potential application to the computation of quenched free energies for polymer models in random media considered in Section 2, it appears a very interesting problem to find a quantitative description of this phenomenon. This question will be pursued in future work.

Remark 3 In many applications, see e.g. Section 2 below, one is actually interested in a level-2 large deviation problem, i.e. the behaviour of the empirical distribution $N^{-1}\sum_{i=1}^N \delta_{Y^i}$. This can be obtained from Theorem 1 via a contraction principle. It appears that there is no 'intrinsic' formulation of the conditional large deviation behaviour on level 2, as the restriction set $\mathscr{R}$ can only be expressed in terms of the empirical process (i.e. a level 3 object).

Remark 4 It is conceivable that the results continue to hold if the discrete set E is replaced by a Polish space. A technical difficulty one will encounter when transferring the arguments to a general context is to give a suitably generalised definition of the (conditional) specific entropy appearing in Lemmas 3 and 4. We have not pursued this issue further.

The rest of this paper is organised as follows: In Section 2 we indicate how Corollary 1, or rather, its analogue in a scenario where in contrast to Assumption (1), ρ has algebraic tails, could be used to represent the quenched free energy of directed polymer models with random disorder via a variational formula. We illustrate the use of Corollary 1 by expressing the quenched free energy of a modified polymer model. Coming back to the main plot, we give in Section 3 a useful characterisation of the property $Q \in \mathscr{R}$ under the additional constraint that Q has finite mean word lengths. This characterisation allows to make a connection between Q and an 'underlying' i.i.d.-ν sequence, and to decompose the relative entropy into a part derived from the concatenated letter sequence plus a part related to the word lengths, given the concatenation. In Section 4, we prove the upper bound (9), Section 5 treats the lower bound (10).

2 Relation to quenched free energy computations

Computations involving conditional expectations of exponential functionals of R_N appear in studies of directed polymer models in random environments. As an example let us consider the (modified) quenched specific free energy for the random heteropolymer model (see [1] and references there), defined as

$$f^{\mathrm{que}}(\lambda, h) := \lim \frac{1}{N} \log Z^*_{N,X},$$

where

$$Z^*_{N,X} = \mathbf{E}\Big[\exp\Big(\lambda \sum_{n=1}^{N} (X_n + h)\mathrm{sign}(S_n)\Big); S_N = 0\Big].$$

Here, $\lambda, h \geq 0$ are parameters, (S_n) is a symmetric simple random walk on $\mathbb{Z}$ starting at $S_0 = 0$, (X_n) are i.i.d. random variables, independent of S, taking the values ± 1 with probability $1/2$ each, and $\mathbf{E}$ refers to expectation with respect to (S_n). In this context, if $S_n = 0$, 'sign(S_n)' is defined as $\mathrm{sign}(S_{n-1})$ – one thinks of the 'bonds' between the steps of the random walk being above or below the axis. We implicitly assume that N is even, otherwise $Z^*_{N,X} = 0$. This is a model for a polymer with a random composition of hydrophilic and hydrophobic monomers near an oil-water interface. The 'letter' X_i models the affinity of monomer i towards different parts of the solvent. h models differences in the affinity of the two types of monomers, and λ is an inverse temperature parameter. The free energy itself uses the same expression without the restriction on $\{S_N = 0\}$, this difference is irrelevant in the limit (see [1], Lemma 2).

Note that for the computation of the free energy, the details of the a priori measure on paths (S_n) are not important. All that matters is the fact that excursions from 0 are independent and symmetric, the only datum that is required to compute $Z^*_{N,X}$ is the distribution (ρ_n) of the excursion lengths: By decomposing the path $S_0, S_1, \dots, S_N$ into excursions away from 0 and assigning independent random signs to the excursions, we can rewrite

$$Z^*_{N,X} = \sum_k \sum_{j_1<\cdots<j_k=N} \prod_{i=1}^k \rho_{j_i-j_{i-1}} \times \prod_{\ell=1}^k \cosh\Big(\lambda \sum_{i=j_{\ell-1}+1}^{j_\ell} (X_i+h)\Big), \tag{12}$$

where $\rho_n = P_0(S_1, \dots, S_{n-1} \neq 0, S_n = 0)$ are the return probabilities for the random walk. Thus for $z \geq 0$ the (random) generating function of $Z^*_{N,X}$ is given by

$$\begin{aligned}
\theta(z) &= \sum_N z^N Z^*_{N,X} \\
&= \sum_N \sum_k \sum_{j_1<\cdots<j_k=N} \prod_{i=1}^k \rho_{j_i-j_{i-1}} \times \prod_{\ell=1}^k \left\{ z^{j_i-j_{i-1}} \cosh\Big(\lambda \sum_{i=j_{\ell-1}+1}^{j_\ell} (X_i+h)\Big)\right\} \\
&= \sum_{k=1}^\infty F_k(X;z),
\end{aligned}$$

where

$$F_k(X;z) := \sum_{j_1<\cdots<j_k} \prod_{i=1}^k \rho_{j_i-j_{i-1}} \exp\Big(\sum_{\ell=1}^k f_z\big((X_{j_{\ell-1}+1}, \dots, X_{j_\ell})\big)\Big) \tag{13}$$

with

$$f_z\big((x_1, \dots, x_\ell)\big) := \ell \log z + \log\cosh\Big(\lambda \sum_{i=1}^\ell (x_i+h)\Big). \tag{14}$$

By introducing an auxiliary i.i.d.-ρ sequence (τ_i) as in Section 1 and defining (Y^i) as in (2), this can be expressed as

$$F_k(X;z) = \mathbb{E}\Big[\exp\Big(k \int f_z(y)\, \pi_1 R_k(dy)\Big)\Big| X\Big], \tag{15}$$

where $\pi_1 : \tilde{E}^{\mathbb{N}} \to \tilde{E}$ is the projection to the first coordinate (and hence $\pi_1 R_k := R_k \circ (\pi_1)^{-1}$ the empirical distribution of the first k words).

Thus if we could (at least in principle) compute the almost sure asymptotic growth rate

$$\varphi(z) := \lim_{k\to\infty} \frac{1}{k} \log F_k(X;z)$$

via an analogue of Corollary 1, we obtained that the radius of convergence of $\theta(z)$ is given by $r_\theta := \sup\{z \geq 0 : \varphi(z) < 0\}$, and hence the quenched specific free energy

$$f^{\text{que}}(\lambda, h) = -\log\sup\{z \geq 0 : \varphi(z) < 0\} = -\log r_\theta.$$

Note that the tails of ρ_n, the return probability of a 1-dimensional random walk, decay only algebraically in this scenario. In particular, ρ does not satisfy Assumption (1), so that the application of Corollary 1 to the computation of $\varphi(z)$ is not justified (and would, in view of Remark 2, almost certainly yield an incorrect result). We reiterate our statement from the end of Remark 2 that in view of the above considerations, it would be very interesting to extend Theorem 1 to the general case.

In order to illustrate the application of the conditional large deviation principle stated in Section 1, let us consider a modified model, where

$$\begin{gathered}\text{the partition function } Z^*_{N,X} \text{ is given by (12) with} \\ \rho \text{ satisfying } \limsup_{n\to\infty}\big(\log\rho_n\big)/n < 0.\end{gathered} \tag{16}$$

This is a model for a situation where the polymer has a strong attraction towards the interface, as under the a priori measure excursions have short tails. We do not advertise this model as particularly physically relevant, we would rather view it as an illustration of the use of the techniques developed in this paper under the restriction of Assumption 1. There can never be a de-pinning transition (as is the case for the original model, see [1]), but still for fixed realisation of (X_i), the polymer can try to optimise its configuration by grouping excursions according to stretches of X_is with the same sign, and there will be an energy-entropy trade-off. In this situation, the application of Corollary 1 will be justified.

Let us briefly discuss the corresponding annealed scenario, where one also averages over the sequence X describing the polymer composition. Let

$$f^{\mathrm{ann}}(\lambda,h) := \lim \frac{1}{N}\log\mathbb{E}\Big[Z^*_{N,X}\Big]$$

be the annealed specific free energy and $\theta^{\mathrm{ann}}(z)$ be the generating function of the sequence $\mathbb{E}[Z^*_{N,X}]$. Arguing as above we have $\theta^{\mathrm{ann}}(z) = \sum_{k=1}^\infty F_k^{\mathrm{ann}}(z)$ where $F_k^{\mathrm{ann}}(z) := \mathbb{E}[F_k(X;z)]$. As under the annealed measure the 'marked excursions' $(Y^i)_{i=1,2,\dots}$ are i.i.d., we see from (15) that $F_k^{\mathrm{ann}}(z) = \big(F_1^{\mathrm{ann}}(z)\big)^k$, hence

$$\varphi^{\mathrm{ann}}(z) := \lim_{k\to\infty}\frac{1}{k}\log F_k^{\mathrm{ann}}(X;z) = \log F_1^{\mathrm{ann}}(z).$$

Note that

$$\begin{aligned} F_1^{\mathrm{ann}}(z) &= \sum_{j=1}^\infty z^j\rho_j\mathbb{E}\Big[\cosh\Big(\lambda\sum_{i=1}^j(X_i+h)\Big)\Big] \\ &= \sum_{j=1}^\infty z^j\rho_j\sum_{m=0}^j 2^{-j}\binom{j}{m}\cosh\Big(\lambda(j-2m+jh)\Big). \end{aligned}$$

This can be viewed as a power series in z with positive coefficients, let R_1^{ann} be its radius of convergence (note that $R_1^{\text{ann}} > 0$ as $\cosh(\lambda(1+h)j)$ grows only exponentially in j). Let z_*^{ann} be the (unique) solution of $F_1^{\text{ann}}(z_*^{\text{ann}}) = 1$ (which exists because $F_1^{\text{ann}}(0) = 0, F_1^{\text{ann}}(z) \to \infty$ as $z \nearrow R_1^{\text{ann}}$), hence

$$f^{\text{ann}}(\lambda, h) = -\log\Big(\sup\{z \geq 0 : \varphi^{\text{ann}}(z) < 0\}\Big) = -\log(z_*^{\text{ann}}).$$

An application of Corollary 1 yields

Lemma 1 *For the modified model (16) we have for any* $0 \leq z < R_1^{\text{ann}}$

$$\varphi(z) = \sup_{Q \in \mathscr{R} \cap \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})} \Big\{ \int f_z(y)(\pi_1 Q)(dy) - H(Q|Q^0) \Big\} \quad a.s., \tag{17}$$

where in the notation of Section 1, $E = \{\pm 1\}$, $\nu(\pm 1) = 1/2$, $q^0\Big((x_1, \dots, x_\ell)\Big) = 2^{-\ell}\rho_\ell$ *for* $(x_1, \dots, x_\ell) \in \{\pm 1\}^\ell$, $Q^0 = (q^0)^{\otimes \mathbb{N}}$, f_z *is defined in (14) and* $\mathscr{R}$ *in (8).*

Note that (15) actually requires only a level-2 large deviation analysis, but it seems that in order to express the restriction set $\mathscr{R}$, one is forced to use a level-3 formulation – the empirical distribution of words alone seems too weak to capture the restrictions coming from conditioning on a typical X sequence.

An explicit evaluation of the variational problem in (17) appears extremely difficult in general. Still, we can obtain from Lemma 1 that the 'quenched to annealed bound' is always strict in this model, i.e.

$$f^{\text{que}}(\lambda, h) < f^{\text{ann}}(\lambda, h) \quad \forall \lambda > 0, h \geq 0 \tag{18}$$

so there is no so-called weak disorder regime. This is not very surprising, we will see below that in the unconditional problem, the sequence X and the excursions both behave atypically in order to maximise the free energy, while in the quenched case, X is forced to be typical.

Lemma 1 is basically Corollary 1 applied to the asymptotic evaluation of (15). There is a slight complication because f_z is not bounded, but (at least) for $z < R_1^{\text{ann}}$ we can find $\epsilon > 0$ such that

$$\limsup_{k \to \infty} \frac{1}{k} \log \mathbb{E}\bigg[\exp\Big((1+\epsilon)k \int f_z(y)\,(\pi_1 R_k)(dy)\Big)\bigg| X\bigg] < \infty \quad \text{a.s.}, \tag{19}$$

which suffices for an application of Varadhan's Lemma, see e.g. Condition 4.3.3 in [3]. In order to check (19) note that $f_z(y) \leq C'|y|$, thus for $z < R_1^{\text{ann}}$ we can find $\epsilon > 0$ and $z' \in (z, R_1^{\text{ann}})$ such that $(1+\epsilon) f_z(y) \le f_{z'}(y)$ for all $y \in \tilde{E}$. As $F_k^{\text{ann}}(z')$ grows only exponentially, the same will hold true for the sequence of conditional expectations inside the log in (19), e.g. by a simple combination of Markov's Inequality and the Borel-Cantelli Lemma as in the proof of Lemma 8.

In order to prove (18), it suffices to check that $\varphi(z) < \varphi^{\text{ann}}(z)$ for all $z \in (0, R_1^{\text{ann}})$. For this it is instructive to apply Varadhan's Lemma to the unconditional distribution and represent

$$
\begin{aligned}
\varphi^{\text{ann}}(z) = \log F_1^{\text{ann}}(z) &= \sup_{Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})} \Big\{ \int f_z(y)(\pi_1 Q)(dy) - H(Q|Q^0) \Big\} \\
&= \sup_{q \in \mathcal{P}(\tilde{E})} \Big\{ \int f_z(y) q(dy) - h(q|q^0) \Big\} \\
&= \log F_1^{\text{ann}}(z) - \inf_{q \in \mathcal{P}(\tilde{E})} h(q|q^{*,\text{ann}}),
\end{aligned}
\tag{20}
$$

where $q^{*,\text{ann}}\big((x_1 \dots, x_\ell)\big) = \frac{1}{F_1^{\text{ann}}(z)} \rho_\ell \prod_{i=1}^{\ell} \nu(x_i) \times \exp f\big((x_1, \dots, x_\ell)\big)$ is (the marginal of) the unconstrained maximiser, which depends implicitly on z. Equality between the two sup-terms above stems from the fact that among all Q with given marginal $\pi_1 Q = q$, the specific relative entropy $H(Q|Q^0)$ is minimised by the product measure $Q = q^{\otimes \mathbb{N}}$.

Fix $z \in (0, R_1^{\text{ann}})$, note that $Q^{*,\text{ann}} := (q^{*,\text{ann}})^{\otimes \mathbb{N}} \notin \mathscr{R}$. A quick way to check this is as follows: In case $h > 0$, we see easily that $\sum_y y_1 q^{*,\text{ann}}(y) > 0$, so $\lim_{L\to\infty} L^{-1} \sum_{j=1}^{L} \kappa(Y)_j > 0$ almost surely under $Q^{*,\text{ann}}$, and hence $Q^{*,\text{ann}} \notin \mathscr{R}$. On the other hand, if $h = 0$ we can observe that $\sum_{|y|=\ell} y_i y_j q^{*,\text{ann}}(y) > 0$ for any $\ell \ge 2$, $1 \le i, j \le \ell$, i.e. letters are positively correlated under $q^{*,\text{ann}}$, so $\lim_{L\to\infty} L^{-1} \sum_{j=1}^{L} \kappa(Y)_j \kappa(Y)_{j+1} > 0$ almost surely under $Q^{*,\text{ann}}$, and hence again $Q^{*,\text{ann}} \notin \mathscr{R}$.

As $\mathscr{R} \cap \mathscr{A}_M$ is compact (see Remark 8), where $\mathscr{A}_M = \{Q : H(Q|Q^0) \le M\}$ is the M-level set of the rate function, and $Q^{*,\text{ann}} \notin \mathscr{R}$, we can find for any $M > 0$ a $\delta > 0$ such that $B_\delta(Q^{*,\text{ann}}) \cap \mathscr{A}_M \subset \mathscr{R}^c$, and so by Lemma 1

$$
\begin{aligned}
\varphi(z) &\le \sup_{Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) \cap \left((B_\delta(Q^{*,\text{ann}}))^c \cup \mathscr{A}_M^c \right)} \Big\{ \int f(y)(\pi_1 Q)(dy) - H(Q|Q^0) \Big\} \\
&< \varphi^{\text{ann}}(z)
\end{aligned}
$$

for a suitable choice of M and δ in view of (20).

3 A characterisation of the restriction set

Imagine cutting the sequence X into pieces and then looking at the empirical process of these pieces. Then obviously the concatenation $\kappa(Y)$ under a limiting $Q \in \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ need not be shift invariant. For example, if we arrange the τs in such a way that the cut-points tend to occur before a certain pattern, then under R_N, the law of the concatenated sequence will have a (possibly atypical under $\nu^{\otimes\mathbb{N}}$) inclination to begin with this pattern.

A way to reinstate shift-invariance (and in some way 'get back the underlying i.i.d. sequence') which works when Q has finite mean word lengths is to size-bias Q according to $L_1 := |Y^1|$ and then 'randomise out the origin' – this is familiar from the theory of stationary renewal processes. Using this idea we obtain in this section a characterisation of the set $\mathscr{R}$ defined in (8).

For $Q \in \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ with $m_Q := \mathbb{E}_Q\, L_1 < \infty$ let $\hat{Q} \in \mathcal{P}(\tilde{E}^{\mathbb{N}})$ be defined by

$$\hat{Q}\Big((Y^i,\dots,Y^k) \in B_k\Big) = \frac{1}{m_Q}\mathbb{E}_Q\Big[L_1 \mathbf{1}_{B_k}\Big((Y^i,\dots,Y^k)\Big)\Big] \tag{21}$$

(for any $k \in \mathbb{N}$, and measurable $B_k \subset \tilde{E}^k$). Let $(\hat{Y}^i)_{i\in\mathbb{N}}$ have law $\hat{Q}$, given $\hat{Y}$, V uniform on $\{0,1,\dots,L_1-1\}$, put

$$Z := \theta^V\Big(\kappa(\hat{Y})\Big). \tag{22}$$

We denote the distribution of Z obtained in this way by $\Psi_Q \in \mathcal{P}(E^{\mathbb{N}})$ to stress that it depends on Q. Explicitly, for measurable $\mathcal{A} \subset E^{\mathbb{N}}$

$$\Psi_Q(\mathcal{A}) = \frac{1}{m_Q}\mathbb{E}_Q\Big[\sum_{i=0}^{L_1-1} \mathbf{1}_{\mathcal{A}}\Big(\theta^i(\kappa(Y))\Big)\Big]. \tag{23}$$

We check that Ψ_Q is shift-invariant: Fix $m \in \mathbb{N}$, $B_m \subset E^m$ measurable. We have

$$\mathbb{P}\Big((Z_1,\dots,Z_m) \in B_m \,\Big|\, \hat{Y}\Big) = \frac{1}{|\hat{Y}^1|}\sum_{i=1}^{|\hat{Y}^1|} \mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big),$$

hence (with a slight abuse of notation)

$$\begin{aligned}\Psi_Q\Big((Z_1,\dots,Z_m) \in B_m\Big) &= \frac{1}{m_Q}\mathbb{E}_Q\Big[L_1 \frac{1}{L_1}\sum_{i=1}^{L_1} \mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big)\Big] \\ &= \frac{1}{m_Q}\mathbb{E}_Q\Big[\sum_{i=1}^{L_1} \mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big)\Big].\end{aligned}$$

As Q is $\tilde{\theta}$-shift invariant,

$$\begin{aligned}&\mathbb{E}_Q\Big[\sum_{i=1}^{L_1}\mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big)\Big]\\&=\mathbb{E}_Q\Big[\sum_{i=L_1+\cdots+L_{k-1}+1}^{L_1+\cdots+L_k}\mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big)\Big]\end{aligned}$$

for any $k\in\mathbb{N}$, hence

$$\begin{aligned}&\Psi_Q\Big((Z_1,\dots,Z_m)\in B_m\Big)\\&=\frac{1}{Mm_Q}\mathbb{E}_Q\Big[\sum_{i=1}^{L_1+\cdots+L_M}\mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_i,\dots,\kappa(\hat{Y})_{i+m-1})\Big)\Big]\end{aligned}$$

for all $M\in\mathbb{N}$. Similarly, we have

$$\begin{aligned}&\Psi_Q\Big((Z_2,\dots,Z_{m+1})\in B_m\Big)\\&=\frac{1}{Mm_Q}\mathbb{E}_Q\Big[\sum_{i=1}^{L_1+\cdots+L_M}\mathbf{1}_{B_m}\Big((\kappa(\hat{Y})_{i+1},\dots,\kappa(\hat{Y})_{i+m})\Big)\Big],\end{aligned}$$

consequently

$$\Big|\Psi_Q\Big((Z_1,\dots,Z_m)\in B_m\Big)-\Psi_Q\Big((Z_2,\dots,Z_{m+1})\in B_m\Big)\Big|\le\frac{2}{Mm_Q}.$$

Taking $M\to\infty$ we see that Ψ_Q is shift invariant.

Remark 5 *If Q is $\tilde{\theta}$-shift ergodic and has finite mean word lengths $\mathbb{E}_Q|Y^1|<\infty$, then Ψ_Q is θ-shift ergodic.*

Proof. Let $\mathcal{A}\subset E^{\mathbb{N}}$ be θ-shift invariant. Then for $y=(y^1,y^2,\dots)\in\tilde{E}$, $\kappa(y)\in\mathcal{A}$ implies $\theta^i(\kappa(y))\in\mathcal{A}$ for any i, so in particular $\kappa(\tilde{\theta}(y))=\theta^{|y^1|}(\kappa(y))\in\mathcal{A}$. Thus, the event $\{\kappa(Y)\in\mathcal{A}\}$ is $\tilde{\theta}$-shift invariant, so $Q(\kappa(Y)\in\mathcal{A})\in\{0,1\}$ by assumption. On the other hand, we see from (23) and the discussion above that

$$\Psi_Q(\mathcal{A})=\frac{1}{m_Q}\mathbb{E}_Q\Big[\sum_{i=0}^{|Y^1|-1}\mathbf{1}_{\mathcal{A}}\Big(\theta^i(\kappa(Y))\Big)\Big]=\frac{1}{m_Q}\mathbb{E}_Q\Big[|Y^1|\,\mathbf{1}_{\mathcal{A}}\Big(\kappa(Y)\Big)\Big]\in\{0,1\}.$$

□

Lemma 2 *Assume that $Q\in\mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ satisfies $\mathbb{E}_Q|Y^1|<\infty$. Then we have $Q\in\mathscr{R}$ if and only if $\Psi_Q=\nu^{\otimes\mathbb{N}}$. In this case, $\mathscr{L}_Q(\kappa(Y))\ll\nu^{\otimes\mathbb{N}}$.*

Proof. Let $\Psi_Q = \nu^{\otimes\mathbb{N}}$. Then under $\hat{Q}$, the sequence $\kappa(Y)$ almost surely has the 'right' asymptotic pattern frequencies (i.e.

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{B_k}\Big((\kappa(Y)_i,\dots,\kappa(Y)_{i+k-1})\Big) = \nu^{\otimes k}(B_k) \quad \text{a.s.}$$

for any measurable $B_k \subset E^k$, $k \in \mathbb{N}$). As $Q \ll \hat{Q}$ (in fact, the density $(\mathbb{E}_Q L_1)/L_1$ is strictly positive), the same holds true for Q, i.e. $Q \in \mathscr{R}$.

Now assume that $Q \in \mathscr{R}$. As $\hat{Q} \ll Q$, the sequence Z_i, $i \in \mathbb{N}$ under Ψ_Q also has the 'right' asymptotic pattern frequencies, i.e.

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{B_k}\Big((Z_i,\dots,Z_{i+k-1})\Big) = \nu^{\otimes k}(B_k) \quad \text{a.s.} \tag{24}$$

for any $k \in \mathbb{N}$, $B_k \subset E^k$ measurable. It suffices to verify that any shift invariant sequence (Z_i) satisfying (24) is in fact an i.i.d.-ν sequence. The limit on the left-hand side of (24) is equal to

$$\mathbb{P}\Big((Z_1,\dots,Z_k) \in B_k \,\Big|\, \mathscr{I}\Big)$$

where $\mathscr{I}$ is the shift-invariant σ-field. Thus

$$\mathbb{P}\Big((Z_1,\dots,Z_k) \in B_k\Big) = \mathbb{E}\Big[\mathbb{P}\Big((Z_1,\dots,Z_k) \in B_k \,\Big|\, \mathscr{I}\Big)\Big] = \nu^{\otimes k}(B_k)$$

so that indeed $\mathscr{L}(Z) = \nu^{\otimes\mathbb{N}}$.

Now assume that $\Psi_Q = \nu^{\otimes\mathbb{N}}$ and let $\mathcal{A} \subset E^{\mathbb{N}}$ be a (measurable) $\nu^{\otimes\mathbb{N}}$-null set. Then we have

$$0 = \nu^{\otimes\mathbb{N}}(\mathcal{A}) = \Psi_Q(\mathcal{A}) = \frac{1}{\mathbb{E}_Q L_1}\mathbb{E}_Q\Big[\sum_{i=0}^{L_1-1} \mathbf{1}_{\mathcal{A}}(\theta^i\kappa(Y))\Big] \geq \frac{1}{\mathbb{E}_Q L_1}\, Q\Big(\kappa(Y) \in \mathcal{A}\Big).$$

This proves that $\mathscr{L}_Q(\kappa(Y)) \ll \nu^{\otimes\mathbb{N}}$. □

Remark 6 If $Q \in \mathscr{R}$ and $\mathbb{E}_Q L_1 < \infty$, by the above there is a random (Y, V) such that $Y \sim \hat{Q}$ and $\theta^V\kappa(Y)$ is distributed like an i.i.d.-ν sequence. We can 'invert' this relation, at least in the two-sided scenario: There is (on some probability space) a random pair (Δ, Z) with values in $\mathbb{Z} \times E^{\mathbb{Z}}$ such that $\mathscr{L}(Z) = \nu^{\otimes\mathbb{Z}}$ and $\mathscr{L}(\theta^\Delta Z) = \mathscr{L}_{\hat{Q}}(\kappa(Y))$. For example, one can take (Y, V) as above then define $Z := \theta^V\kappa(Y)$, $\Delta := -V$.

Remark 7 Note that the mappings $Q \mapsto \hat{Q}$, $Q \mapsto \Psi_Q$ are not continuous with respect to the weak topology on $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ (as $\tilde{E}^{\mathbb{N}} \ni (y^i)_i \mapsto |y^1|$ is not bounded, weak convergence need not imply convergence of the first moment of piece lengths). On the other hand, assume that $Q_N \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ converge weakly to Q_∞ and that additionally $\mathbb{E}_{Q_N}\big[L_1\big] \to \mathbb{E}_{Q_\infty}\big[L_1\big]$ as $N \to \infty$. Then

$$\hat{Q}_N \to \hat{Q}_\infty \text{ weakly on } \mathcal{P}(\tilde{E}^{\mathbb{N}}) \quad \text{and} \quad \Psi_{Q_N} \to \Psi_{Q_\infty} \text{ weakly on } \mathcal{P}(E^{\mathbb{N}}).$$

Proof. Note that by the assumptions, the family $\{\mathscr{L}_{Q_N}(L_1), N \in \mathbb{N}\}$ is uniformly integrable. Hence also for any $k \in \mathbb{N}$, $y^i \in \tilde{E}$, the family $\{\mathscr{L}_{Q_N}(L_1\mathbf{1}(Y^i = y^i, i = 1, \dots, k)), N \in \mathbb{N}\}$ is uniformly integrable. This implies

$$\hat{Q}_N(Y^i = y^i, i = 1, \dots, k) \to \hat{Q}_\infty(Y^i = y^i, i = 1, \dots, k).$$

Similarly, because $0 \le \sum_{i=1}^{L_1} \mathbf{1}(\kappa(Y)_i = z_1, \dots, \kappa(Y)_{i+m} = z_{m+1}) \le L_1$ (for any $m \in \mathbb{N}$, $z_j \in E$), we conclude that

$$\Psi_{Q_N}(Z_1 = z_1, \dots, Z_{m+1} = z_{m+1}) \to \Psi_{Q_\infty}(Z_1 = z_1, \dots, Z_{m+1} = z_{m+1}).$$

□

Remark 8 Much of the difficulty in the proofs below stems from the fact that the set $\mathscr{R}$ is not closed in the weak topology. In fact, $\overline{\mathscr{R}} \cap \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) = \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$. On the other hand, let

$$\mathscr{A}_M := \Big\{Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) : H(Q|Q^0) \le M\Big\}, \quad M \ge 0$$

be the level sets of the rate function $Q \mapsto H(Q|Q^0)$. One can see from the considerations in Lemma 5 and Proposition 2 that

$$\text{for any } M, \text{ the set } \mathscr{R} \cap \mathscr{A}_M \text{ is closed (in the weak topology on } \mathcal{P}(\tilde{E}^{\mathbb{N}})). \tag{25}$$

Proof. For the first claim it suffices to show $\overline{\mathscr{R}} \supset \Big\{Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) : \mathbb{E}_Q\Big[|Y^1|\Big] < \infty\Big\}$, as this set is dense in $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$. Fix an arbitrary Q in $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ satisfying $\mathbb{E}_Q\,|Y^1| < \infty$. Let $\tilde{q} \in \mathcal{P}(\tilde{E})$ be given by

$$\tilde{q}\Big((x_1, \dots, x_n)\Big) = \frac{C}{n^{-3/2}} \prod_{i=1}^{n} \nu(x_i),$$

i.e. the length of the word has heavy tails, given the length is n, it looks like n independent draws from ν. Define Q_N as follows: under $\tilde{Q}_N$, the blocks $(Y^{kN+1}, Y^{kN+2}, \dots, Y^{(k+1)N-1})$, $k \in \mathbb{N}_+$, are i.i.d, $\mathscr{L}_{\tilde{Q}_N}((Y^1, \dots, Y^N)) = \tilde{q} \otimes Q|_{\sigma(Y^1,\dots,Y^{N-1})}$. Q_N is defined as $\tilde{Q}_N$ with randomised origin, formally $Q_N = N^{-1}\sum_{i=0}^{N-1} \tilde{Q}_N \circ \tilde{\theta}^i$. Then we have $Q_N \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ (in fact even $Q_N \in \mathcal{P}^{\text{erg}}(\tilde{E}^{\mathbb{N}})$), $Q_N \to Q$ weakly. Finally, each $Q_N \in \mathscr{R}$ because the word length under $\tilde{q}$ has no mean: imagine pointing at position U in $\kappa(Y)$ under Q_N, where

$U \sim \mathrm{Unif}(\{1,\ldots,L\})$. As $L \to \infty$, the probability tends to one that one actually looks inside a '$\tilde{q}$-word' of the concatenation, where the pattern frequencies are what they ought to be in a $\nu^{\otimes\mathbb{N}}$-sequence.

In order to verify (25), note that $\mathscr{A}_M^c$ is open because $H(\cdot|Q^0)$ is lower semicontinuous. By combining Lemmas 7 and 5 we can choose for any $Q \in \mathscr{A}_M \setminus \mathscr{R}$ an open neighbourhood $\mathcal{U}_Q \ni Q$ such that $\limsup \frac{1}{N} \log \mathbb{P}\big(R_N \in \mathcal{U}_Q \big| X\big) \le -2M$. By Proposition 2, we must have $\mathcal{U}_Q \cap \mathscr{R} \subset \mathscr{A}_M^c$. Hence $(\mathscr{R} \cap \mathscr{A}_M)^c$ is open. □

3.1 A decomposition of the specific relative entropy

In this section we study how the specific entropy (and the specific relative entropy w.r.t. Q^0) of a Q can be expressed in terms involving Ψ_Q, which will be useful later on. Here and in the following, for a probability measure P and a discrete random variable U we will be writing $P(U)$ for the random variable $f(U)$, where $f(u) = P(U = u)$. Similarly, $P(U|V)$ means $g(U,V)$, where $g(u,v) = P(U = u|V = v)$.

Lemma 3 *Let $Y = (Y^i)_{i\in\mathbb{N}}$ have distribution Q, write $L_i := |Y^i|$, $K^N := \kappa(Y^1,\ldots,Y^N)$. Assume $Q \in \mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ satisfies $m_Q := \mathbb{E}_Q\, L_1 < \infty$. Then we have*

$$\lim_{N\to\infty} -\frac{1}{N} \log Q(K^N) = m_Q H(\Psi_Q) \qquad Q\text{-a.s.,} \tag{26}$$

$$\lim_{N\to\infty} -\frac{1}{N} \log Q\Big(L_1,\ldots,L_N \Big| K^N\Big) =: H_L^c(Q) \tag{27}$$

exists Q-almost surely, the limit $H_L^c(Q)$ is a constant. In particular, the specific entropy of Q can be represented as

$$H(Q) = \lim_{N\to\infty} -\frac{1}{N} \log Q\Big(Y^1,\ldots,Y^N\Big) = m_Q H(\Psi_Q) + H_L^c(Q). \tag{28}$$

We call $H_L^c(Q)$ the conditional specific entropy of word lengths under Q, given the concatenation. Intuitively, a 'Ψ_Q-typical' word $x \in \tilde{E}$ of length $|x| \approx N m_Q$ can be decomposed in $\approx \exp(N H_L^c(Q))$ different ways into '$Q|_{\mathscr{F}_N}$-typical' N-vectors of words $(y^1,\ldots,y^N)$ satisfying $\kappa(y^1,\ldots,y^N) = x$. See the proof of Lemma 9 for a rigorous implementation of this notion.

Proof. Write $S_N := L_1 + \cdots + L_N$ $(= |K^N|)$, fix $\epsilon > 0$. Note that on the event

$$A_N := \Big\{N(m_Q - \epsilon) \le S_N \le N(m_Q + \epsilon)\Big\}$$

we have

$$Q(\kappa(Y)|_{[1\ldots N(m_Q+\epsilon)]}, S_N) \le Q(K^N) \le Q(\kappa(Y)|_{[1\ldots N(m_Q-\epsilon)]}).$$

The second inequality together with the facts that $\liminf_{N\to\infty} \mathbf{1}_{A_N} = 1$ almost surely by ergodicity of Q and $\mathscr{L}_Q(\kappa(Y)) \ll \Psi_Q$ by Lemma 2 shows that

$$\limsup_{N\to\infty} \frac{1}{N} \log Q(K^N) \leq -(m_Q - \epsilon) H(\Psi_Q) \quad \text{a.s.} \tag{29}$$

because

$$\lim_{n\to\infty} \frac{1}{n} \log \Psi_Q|_{[1\dots n]}\Big((Z_1, \dots, Z_n)\Big) = -H(\Psi_Q) \quad \text{for } \Psi_Q\text{–a.a. } Z = (Z_1, Z_2, \dots),$$

where $H(\Psi_Q)$ is the specific entropy of Ψ_Q (recall that Ψ_Q is θ-shift ergodic by Remark 5). On the other hand, writing

$$Q(\kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}, S_N) = Q(\kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}) Q\Big(S_N \Big| \kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}\Big)$$

and noting that

$$\lim_{N\to\infty} \frac{1}{N} \log Q\Big(S_N \Big| \kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}\Big) = 0 \qquad \text{a.s.} \tag{30}$$

we obtain

$$\liminf \frac{1}{N} \log Q(K^{(N)}) \geq -(m_Q + \epsilon) H(\Psi_Q) \tag{31}$$

almost surely as above. Taking $\epsilon \to 0$ in (29) and (31), we obtain (26). Intuitively, (30) holds true because the conditional distribution concentrates on a set of size $\approx$ const. $\times$ N, a formal argument might be as follows: For any $x \in E^{[N(m_Q+\epsilon)]}$, $\delta > 0$ we have

$$\sum_{\substack{k=[N(m_Q-\epsilon)],\\ Q(S_N=k \,|\, \kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}=x) \leq \exp(-\delta N)}}^{[N(m_Q+\epsilon)]} Q\Big(S_N = k \,\Big|\, \kappa(Y)|_{[1\dots N(m_Q+\epsilon)]} = x\Big) \leq 2N\epsilon \exp(-\delta N)$$

which is summable in N. Thus the Borel-Cantelli Lemma together with $\liminf \mathbf{1}_{A_N} = 1$ a.s. shows that

$$\limsup_{N\to\infty} -\frac{1}{N} \log Q\Big(S_N \Big| \kappa(Y)|_{[1\dots N(m_Q+\epsilon)]}\Big) \leq \delta \quad \text{a.s.}$$

for any $\delta > 0$.

Finally, we know by ergodicity of Q that

$$\lim_{N\to\infty} -\frac{1}{N} \log Q\Big(Y^1, \dots, Y^N\Big)$$

exists almost surely and equals $H(Q)$, the specific entropy of Q. Writing

$$Q\Big(Y^1, \dots, Y^N\Big) = Q(K^N) Q\Big(L_1, \dots, L_N \Big| K^N\Big),$$

this gives (27) and (28). □

The following result decomposes the specific entropy of Q with respect to Q^0 into a part which comes from the concatenated letters and a part describing the different word length distributions.

Lemma 4 *Assume $Q \in \mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ satisfies $m_Q := \mathbb{E}_Q\, L_1 < \infty$. Then we have*

$$H(Q|Q^0) = m_Q H(\Psi_Q|\nu^{\otimes\mathbb{N}}) - \mathbb{E}_Q\, \log \rho_{L_1} - H^c_L(Q). \tag{32}$$

Note that the term $-\mathbb{E}_Q\, \log \rho_{L_1} - H^c_L(Q)$ can be interpreted as the conditional specific relative entropy of word lengths under Q with respect to $\rho^{\otimes\mathbb{N}}$, given the concatenation.

Proof. We have Q-a.s. by ergodicity of Q

$$\begin{aligned} H(Q|Q^0) &= \lim_{N\to\infty} \frac{1}{N} \log \frac{Q\Big(Y^1,\dots,Y^N\Big)}{Q^0\Big(Y^1,\dots,Y^N\Big)} \\ &= -H(Q) - \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \log \rho_{L_i} - \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{L_1+\cdots+L_N} \log \nu(\kappa(Y)_j) \\ &= -H^c_L(Q) - \mathbb{E}_Q\, \log \rho_{L_1} - m_Q H(\Psi_Q) - \lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{L_1+\cdots+L_N} \log \nu(\kappa(Y)_j) \end{aligned}$$

by Lemma 3. Furthermore note that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{L_1+\cdots+L_N} \log \nu(\kappa(Y)_j) = m_Q \int_{E^{\mathbb{N}}} \log \rho(z_1)\, \Psi_Q(dz) \quad Q-\text{a.s.} \tag{33}$$

because $(L_1 + \cdots + L_N)/N \to m_Q$, $\mathscr{L}_Q(\kappa(Y)) \ll \Psi_Q$ by Lemma 2 and Ψ_Q is θ-shift ergodic by Remark 5.

Finally note that (because $\nu^{\otimes\mathbb{N}}$ is a product measure)

$$-m_Q H(\Psi_Q) - m_Q \int_{E^{\mathbb{N}}} \log \nu(z_1)\, \Psi_Q(dz) = m_Q H(\Psi_Q|\nu^{\otimes\mathbb{N}})$$

to complete the proof. □

4 Conditioning and the restriction set, upper bound

In this section we prove the upper bound in Theorem 1. First we show that $\mathbb{P}(R_N \approx Q|X)$ is super-exponentially expensive for any typical X and $Q \notin \mathscr{R}$.

Intuitively, this is so because then $R_N \approx Q$ requires to include substantial (i.e. with length of order N) atypical pieces of the X-sequence in the sum (6), which requires that at least some of the j-increments appearing in (6) are exponentially long in N. Because of (1), all such terms will be extremely small.

Lemma 5 *Let $Q \in \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}}) \setminus \mathscr{R}$ satisfy $m_Q = \mathbb{E}_Q\,|Y^{(1)}| < \infty$. Then for any $B \geq 0$ there is an open neighbourhood $\mathcal{U} \subset \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ of Q such that*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\Big(R_N \in \mathcal{U} \Big| X\Big) \leq -B \quad a.s. \tag{34}$$

Proof. Step 1. First we claim that there exist $\varepsilon_1, \varepsilon_2 > 0$ such that for any large enough $M \in \mathbb{N}$ there is a subset $\mathcal{B}_M \subset \tilde{E}^M$ of 'X-unlikely sentences' with the following properties:

$$\sum_{(w^1,\dots,w^M)\in\mathcal{B}_M} Q\Big((Y^1,\dots,Y^M) = (w^1,\dots,w^M)\Big) \geq \varepsilon_1 \quad \text{and} \tag{35}$$

$$\mathbb{P}\Big(\text{the sequence } X \text{ begins with an element of } \kappa(\mathcal{B}_M)\Big) \leq \exp(-\varepsilon_2 M). \tag{36}$$

In order to see this note that if Q is also $\tilde{\theta}$-ergodic, by combining (26) and (33) we have (recall $K^N = \kappa(Y^1,\dots,Y^N)$)

$$\frac{1}{N} \log \frac{Q(K^N)}{\prod_{i=1}^{|K^N|} \nu(K_i^N)} \underset{N\to\infty}{\longrightarrow} m_Q\, H(\Psi_Q|\nu^{\otimes\mathbb{N}}) \qquad Q\text{-a.s.} \tag{37}$$

and the righthand side is strictly positive for $Q \notin \mathscr{R}$. In the general case we see by decomposing Q into its ergodic components (cf e.g. [4], Thm. 5.2.16) that there is a random variable $\tilde{Z} \geq 0$, adapted to the shift-invariant sigma-field, such that

$$\frac{1}{N} \log \frac{Q(K^N)}{\prod_{i=1}^{|K^N|} \nu(K_i^N)} \underset{N\to\infty}{\longrightarrow} \tilde{Z} \qquad Q\text{-a.s.}, \tag{38}$$

and the event $\{\tilde{Z} > 0\}$ has strictly positive probability under Q if and only if $Q \notin \mathscr{R}$. Thus by assumption we can find $\varepsilon_1, \varepsilon_2 > 0$ such that $Q(\tilde{Z} > \varepsilon_2) > 2\varepsilon_1$, hence

$$Q\left(\frac{1}{N} \log \frac{Q(K^N)}{\prod_{i=1}^{|K^N|} \nu(K_i^N)} > \varepsilon_2\right) > 2\varepsilon_1 \tag{39}$$

for N large enough. As $\tilde{E}$ is countable, for any large enough N we can find (pairwise different) words $w^1, \dots, w^L \in \tilde{E}$ (the w^i and L will depend on N, but we suppress this dependency in the notation) such that

$$\sum_{j=1}^{L} Q\Big(K^N = w^j\Big) > 2\varepsilon_1 \quad \text{and} \tag{40}$$

$$\log Q\Big(K^N = w^j\Big) \geq \varepsilon_2 N + \sum_{i=1}^{|w^j|} \log \nu(w_i^j), \quad j = 1, \dots, L. \tag{41}$$

Note that (41) implies

$$\sum_{j=1}^{L} \prod_{i=1}^{|w^j|} \nu(w_i^j) \leq \exp(-\varepsilon_2 N) \sum_{j=1}^{L} Q\Big(K^N = w^j\Big) \leq \exp(-\varepsilon_2 N). \tag{42}$$

Finally, for each of the words w^j $(j = 1, \dots, L)$ choose M_j (pairwise different) ordered decompositions into N subwords $w^{j,k,1}, \dots, w^{j,k,N}$ $(k = 1, \dots, M_j)$ such that $w^j = \kappa(w^{j,k,1}, \dots, w^{j,k,N})$ for each j, k and

$$\sum_{k=1}^{M_j} Q\Big((Y^1, \dots, Y^N) = (w^{j,k,1}, \dots, w^{j,k,N})\Big) \geq \frac{1}{2} Q\Big(K^N = w^j\Big), \quad j = 1, \dots, L.$$

This yields (35) and (36) with

$$\mathcal{B}_N := \Big\{(w^{j,k,1}, \dots, w^{j,k,N}) : 1 \leq k \leq M_j,\ 1 \leq j \leq L\Big\}.$$

Step 2. Let $A \subset \mathbb{N}$ have asymptotic density $p \in (0,1)$, i.e.

$$\lim_{n \to \infty} \frac{|A \cap \{1, \dots, n\}|}{n} = p, \tag{43}$$

and $\varepsilon \in (0,1)$. We claim that for any $p' > p$ we have for all N large enough

$$\sum_{\substack{1 \leq s_1 < s_2 < \cdots < s_N \\ |\{s_1, \dots, s_N\} \cap A| \geq \varepsilon N}} \prod_{j=1}^{N} \rho(s_j - s_{j-1}) \\ \leq (\tilde{C})^N \exp\left(N\Big(\varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon}\Big)\right) \left(\frac{\exp(-\tilde{\lambda}/p')}{1 - \exp(-\tilde{\lambda}/p')}\right)^{\varepsilon N}, \tag{44}$$

where $\tilde{C}$, $\tilde{\lambda}$ are given by Lemma 6. The left hand side of (44) is not more than

$$
\begin{aligned}
&\sum_{1\le i_1<\cdots<i_{\lceil\varepsilon N\rceil}\le N}\ \sum_{\substack{j_1<\cdots<j_{\lceil\varepsilon N\rceil}\\ j_1,\dots,j_{\lceil\varepsilon N\rceil}\in A}}\ \prod_{\ell=1}^{\lceil\varepsilon N\rceil}\rho^{*(i_\ell-i_{\ell-1})}(j_\ell-j_{\ell-1})\\
&\le \sum_{1\le i_1<\cdots<i_{\lceil\varepsilon N\rceil}\le N}\ \sum_{\substack{j_1<\cdots<j_{\lceil\varepsilon N\rceil}\\ j_1,\dots,j_{\lceil\varepsilon N\rceil}\in A}}\ \prod_{\ell=1}^{\lceil\varepsilon N\rceil}\tilde C^{i_\ell-i_{\ell-1}}\exp\Big(-\tilde\lambda(j_\ell-j_{\ell-1})\Big)\\
&\le \tilde C^N\binom{N}{\lceil\varepsilon N\rceil}\sum_{\substack{j_1<\cdots<j_{\lceil\varepsilon N\rceil}\\ j_1,\dots,j_{\lceil\varepsilon N\rceil}\in A}}\exp\Big(-\tilde\lambda j_{\lceil\varepsilon N\rceil}\Big)\\
&=\tilde C^N\binom{N}{\lceil\varepsilon N\rceil}\sum_{r=\lceil\varepsilon N\rceil}^{\infty}\exp(-\tilde\lambda t_r(A))\binom{r-1}{\lceil\varepsilon N\rceil-1},
\end{aligned}
\tag{45}
$$

where we used Lemma 6 in the first inequality and

$$
t_r(A):=\min\Big\{k : A\cap\{1,\dots,k\}=r\Big\}
\tag{46}
$$

is the position of the r-th element of A. Note that for large N

$$
\binom{N}{\lceil\varepsilon N\rceil}\le \text{const.}\times\exp\Big(N\Big(\varepsilon\log\frac{1}{\varepsilon}+(1-\varepsilon)\log\frac{1}{1-\varepsilon}\Big)\Big)
\tag{47}
$$

by Stirling's Formula, that for $s\in[0,1)$

$$
\sum_{r=n}^{\infty}s^r\binom{r-1}{n-1}=\sum_{r_1,\dots,r_n=1}^{\infty}s^{r_1+\cdots+r_n}=\Big(\frac{s}{1-s}\Big)^n
\tag{48}
$$

and that by (43)

$$
\lim_{r\to\infty}\frac{t_r(A)}{r}=\frac{1}{p}.
\tag{49}
$$

Finally, combine (45) and (47)–(49) to obtain the claim (44).

Step 3. Consider (a large) $M\in\mathbb{N}$, let $\varepsilon_1,\varepsilon_2$ and $\mathcal{B}_M$ be as chosen in Step 1, put

$$
\mathcal{U}:=\Big\{Q'\in\mathcal{P}^{\text{shift}}(\tilde E^{\mathbb{N}})\ :\ Q'\Big((Y^1,\dots,Y^M)\in\mathcal{B}_M\Big)>\varepsilon_1/2\Big\}.
\tag{50}
$$

Note that by (35), $\mathcal{U}$ is an open neighbourhood of Q. Let

$$
A:=\Big\{i\in\mathbb{N}\ :\ (X_i,X_{i+1},\dots,X_{i+k})\in\kappa(\mathcal{B}_M)\text{ for some }k\Big\}
\tag{51}
$$

be the (random) set of positions where some element of $\kappa(\mathcal{B}_M)$ starts on the given X. As X is i.i.d. and $|\kappa(\mathcal{B}_M)|<\infty$, A has a non-random asymptotic density p, and $p\le\exp(-\varepsilon_2 M)$ by (36). Furthermore we see from (6) that for large enough N (so that boundary terms coming from the periodisation become negligible) only such summands $(j_1, \dots, j_N)$ will contribute to $\mathbb{P}(R_N \in \mathcal{U}|X)$ which have the property that

$$\#\Big\{1 \le i \le N \;:\; j_i \in A\Big\} \ge \frac{\varepsilon_1}{4M} =: \varepsilon. \tag{52}$$

(We divide by M to account for possible overlaps of the concatenations of different elements of $\mathcal{B}_M$.) Now (44) yields

$$\begin{aligned}
&\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{U}|X) \\
&\le \log \tilde{C} + \Big(\varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon}\Big) + \varepsilon \log \Big(2 \exp\big(-\tilde{\lambda}/p\big)\Big) \\
&\le \log \tilde{C} + \Big(\varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon}\Big) + \varepsilon \log 2 - \frac{\varepsilon_1}{4M} \times \tilde{\lambda} \exp\big(\varepsilon_2 M\big).
\end{aligned}$$

The expression in the last line can be made arbitrarily negative by picking a large M (note that the terms involving ε are uniformly bounded for $\varepsilon \in (0,1)$). □

Lemma 6 *Let ρ satisfy (1). There are $\tilde{C}$ and $\tilde{\lambda} > 0$ such that*

$$\forall\, k, n \in \mathbb{N}:\; \rho^{*k}(n) \le \tilde{C}^k \exp(-\tilde{\lambda} n). \tag{53}$$

Proof. We have

$$\rho^{*k}(n) \le \sum_{\substack{n_1,\dots,n_k=1 \\ n_1+\cdots+n_k=n}}^{\infty} \prod_{i=1}^{k} C e^{-\lambda n_i} = C^k e^{-\lambda n} \binom{n-1}{k-1}. \tag{54}$$

Fix $\varepsilon \in (0, 1/2)$. As $k \mapsto \binom{n-1}{k-1}$ is increasing for $k < (n-1)/2$, the right hand side of (54) for $k \le \varepsilon n$ is not more than

$$C^k e^{-\lambda n} \binom{n-1}{\lceil \varepsilon n \rceil} \le \text{const.} \times C^k \exp\Big(-\lambda n + n\Big(\varepsilon \log \frac{1}{\varepsilon} + (1-\varepsilon) \log \frac{1}{1-\varepsilon}\Big)\Big)$$

by Stirling's Formula, while for $k > \varepsilon n$ the observation $\binom{n-1}{k-1} \le 2^{n-1}$ yields the bound

$$C^k e^{-\lambda n} 2^{k/\varepsilon} = (2^{1/\varepsilon} C)^k e^{-\lambda n}.$$

Put $\tilde{\lambda} := \lambda + \varepsilon \log \varepsilon + (1-\varepsilon) \log(1-\varepsilon)$, $\tilde{C} := 2^{1/\varepsilon} C$. Note that $\tilde{\lambda} < \lambda$ can be chosen arbitrarily close to λ, at the expense of enlarging $\tilde{C}$. □

Lemma 7 *Let ρ satisfy (1). Then any $Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ with $H(Q|Q^0) < \infty$ has $m_Q = \mathbb{E}_Q\,|Y^1| < \infty$.*

Proof. Let $\mu := \mathscr{L}_Q(|Y^1|) \in \mathcal{P}(\mathbb{N})$ be the marginal distribution of word lengths under Q. As $N^{-1}h(Q|_{\mathscr{F}_N} \,|\, Q^0|_{\mathscr{F}_N}) \nearrow H(Q|Q^0) < \infty$ and $h(\mu|\rho) \leq h(Q|_{\mathscr{F}_1} \,|\, Q^0|_{\mathscr{F}_1})$ it suffices to check that

$$h(\mu|\rho) = \sum_{n=1}^{\infty} \rho_n \frac{\mu_n}{\rho_n} \log \frac{\mu_n}{\rho_n} < \infty \tag{55}$$

implies $\sum_n n\mu_n < \infty$. This must be well known, for completeness and lack of reference, here is a short argument: Split the sum in (55) into

$$\sum_{\substack{n=1 \\ \mu_n \geq \rho_n}}^{\infty} \rho_n \frac{\mu_n}{\rho_n} \log \frac{\mu_n}{\rho_n} + \sum_{\substack{n=1 \\ \mu_n < \rho_n}}^{\infty} \rho_n \frac{\mu_n}{\rho_n} \log \frac{\mu_n}{\rho_n}.$$

As $x \mapsto x \log x$ is continuous on $[0,1]$, the second sum has some finite value $\in (-\infty, 0]$, so the assumption implies

$$\begin{aligned} \infty > \sum_{\substack{n=1 \\ \mu_n \geq \rho_n}}^{\infty} \mu_n \log \frac{\mu_n}{\rho_n} &= \sum_{\substack{n=1 \\ \mu_n \geq \rho_n}}^{\infty} \mu_n \Big(\underbrace{\log \mu_n - \log \rho_n}_{\geq 0} \Big) \\ \geq \sum_{\substack{n=1 \\ \mu_n \geq C \exp(-\lambda n/2)}}^{\infty} \mu_n \Big(\log \mu_n - \log \rho_n \Big) &\geq \sum_{\substack{n=1 \\ \mu_n \geq C \exp(-\lambda n/2)}}^{\infty} \mu_n \frac{\lambda n}{2} \end{aligned}$$

by (1). On the other hand,

$$\sum_{\substack{n=1 \\ \mu_n < C \exp(-\lambda n/2)}}^{\infty} n\mu_n < \infty$$

holds automatically. Combining these two estimates yields the claim. □

Next we observe that an unconditional upper bound is automatically also an upper bound for the conditional distributions:

Lemma 8 *For any closed $F \subset \mathcal{P}(\tilde{E}^{\mathbb{N}})$ we have*

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in F \,|\, X) \leq - \inf_{Q \in F \cap \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})} H(Q|Q^0) \qquad a.s. \tag{56}$$

This is well known, here is a short proof for the sake of completeness.

Proof. Write $I(F) := \inf_{Q \in F \cap \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})} H(Q|Q^0)$. For $\epsilon > 0$ we have by Markov's Inequality and the unconditional LDP

$$
\begin{aligned}
&\mathbb{P}\Big(\mathbb{P}(R_N \in F \,|\, X) \geq \exp(-N(I(F) - 2\epsilon)\Big) \\
&\leq e^{N(I(F)-2\epsilon)} \mathbb{E}\Big[\mathbb{P}(R_N \in F \,|\, X)\Big] = e^{N(I(F)-2\epsilon)} \mathbb{P}(R_N \in F) \\
&\leq e^{N(I(F)-2\epsilon)} e^{-N(I(F)-\epsilon)} = e^{-\epsilon N}
\end{aligned}
$$

for N large enough, and hence

$$
\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in F \,|\, X) \leq -I(F) - 2\epsilon \qquad \text{a.s.}
$$

by the Borel-Cantelli Lemma. Take $\epsilon \to 0$ to conclude. □

The following is the main result of this section:

Proposition 1 *For any closed $F \subset \mathcal{P}(\tilde{E}^{\mathbb{N}})$ we have a.s.*

$$
\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in F \,|\, X) \leq - \inf_{Q \in F \cap \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) \cap \mathscr{R}} H(Q|Q^0). \tag{57}
$$

In particular, for $F \cap \mathscr{R} = \emptyset$ the conditional probability $\mathbb{P}(R_N \in F \,|\, X)$ decays almost surely super-exponentially.

Remark 9 *As the weak topology on $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ is separable, it is standard to strengthen (57) to hold with probability one simultaneously for all closed sets F, see e.g. [2], proof of Prop. III.2.*

Proof of Proposition 1. First note that even though R_N is not exactly shift-invariant because of boundary terms, it is nearly so: for any weak neighbourhood O of $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$, there is n_0 such that $R_N \in O$ for $N \geq n_0$. As $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ is closed in the weak topology, we can restrict to $F \cap \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ on the right-hand side of (57).

Fix $B > 0$ and $\varepsilon > 0$ for the moment, let $\mathscr{A}_B = \{Q : H(Q|Q^0) \leq B\}$ be the B-level set of $H(\cdot\,|Q^0)$. Recall that $\mathscr{A}_B$ is compact with respect to the weak topology on $\mathcal{P}(\tilde{E}^{\mathbb{N}})$. For any $Q \in F \cap \mathscr{A}_B$ choose an open neighbourhood $\mathcal{U}_Q$ of Q as follows:

(1) If $Q \notin \mathscr{R}$, take $\mathcal{U}_Q = \mathcal{U}$ as guaranteed by Lemma 5, so that (34) is satisfied.
(2) If $Q \in \mathscr{R}$, choose $\mathcal{U}_Q$ such that $\inf_{Q' \in \overline{\mathcal{U}_Q}} H(Q'|Q^0) \geq H(Q|Q^0) - \varepsilon$. This is possible by lower semicontinuity of $H(\cdot\,|Q^0)$.

As $F \cap \mathscr{A}_B$ is compact, we can pick a finite sub-cover $\mathcal{U}_{Q_1}, \dots, \mathcal{U}_{Q_m}$. Note that $F \cap \Big(\cup_{i=1}^m \mathcal{U}_{Q_i} \Big)^c$ is closed and contained in $\mathscr{A}_B^c$, so

$$\inf_{Q \in F \cap \left(\cup_{i=1}^m \mathcal{U}_{Q_i} \right)^c} H(Q|Q^0) \geq B,$$

and hence

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\Big(R_N \in F \cap \Big(\cup_{i=1}^m \mathcal{U}_{Q_i} \Big)^c \,\Big|\, X \Big) \leq -B \quad \text{a.s.}$$

by Lemma 8. On the other hand, for $i = 1, \dots, m$ we have by construction (employing Lemma 5 if $Q_i \notin \mathscr{R}$ and Lemma 8 if $Q_i \in \mathscr{R}$)

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\Big(R_N \in \mathcal{U}_{Q_i} \,\Big|\, X \Big) \leq \Big(-B \Big) \vee \Big(- \inf_{Q \in F \cap \mathscr{R}} H(Q|Q^0) + \varepsilon \Big) \quad \text{a.s.},$$

consequently

$$\limsup_{N\to\infty} \frac{1}{N} \log \mathbb{P}\Big(R_N \in F \,\Big|\, X \Big) \leq \Big(-B \Big) \vee \Big(- \inf_{Q \in F \cap \mathscr{R}} H(Q|Q^0) + \varepsilon \Big) \quad \text{a.s.}$$

Take $B \to \infty$, $\varepsilon \to 0$ to conclude. □

5 Lower bound

Proposition 2 *Let $Q \in \mathscr{R} \cap \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$, and let $O \subset \mathcal{P}(\tilde{E}^{\mathbb{N}})$ be an open neighbourhood of Q. Then we have*

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in O | X) \geq -H(Q|Q^0) \qquad a.s. \tag{58}$$

Remark 10 *Again it is standard to strengthen (58) to hold with probability one simultaneously for all open sets O, see e.g. [2], proof of Prop. III.3.*

We will have occasion to consider open neighbourhoods of $Q \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$ of the following form

$$\tilde{\mathcal{O}}_Q := \Big\{ Q' \in \mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}}) \,:\, \Big| \int g_i \, dQ' - \int g_i \, dQ \Big| < \tilde{\varepsilon}_i, i = 1, \dots, A_{\tilde{\mathcal{O}}_Q} \Big\} \tag{59}$$

where $g_i : \tilde{E}^{\mathbb{N}} \to \mathbb{R}$ $(i = 1, \dots, A_{\tilde{\mathcal{O}}_Q})$ satisfy $||g_i||_\infty \leq 1$ and depend only on $y^1, \dots, y^{B_{\tilde{\mathcal{O}}_Q}}$ for some $B_{\tilde{\mathcal{O}}_Q} \in \mathbb{N}$. Note that such sets generate the weak topology on $\mathcal{P}^{\text{shift}}(\tilde{E}^{\mathbb{N}})$.

For $x \in E^m$, $m \in \{n, n+1, \dots\} \cup \{\infty\}$ let

$$R_n(x) := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\theta^i\left((x|_{[1\dots n]})^{\mathrm{per}}\right)} \in \mathcal{P}(E^{\mathbb{N}}) \tag{60}$$

be the corresponding n-th empirical letter process measure. Furthermore, for $1 \le j_1 < \dots < j_n$ let

$$\tilde{R}^n_{j_1,\dots,j_n}(x) := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{\tilde{\theta}^i\left((x|_{[1\dots j_1]}, x|_{[j_1+1\dots j_2]}, \dots, x|_{[j_{n-1}+1\dots j_n]})^{\mathrm{per}}\right)} \in \mathcal{P}(\tilde{E}^{\mathbb{N}}) \tag{61}$$

the n-th empirical word process measure obtained by cutting x at the cut-points j_i. Note that in this notation (see equations (2) and (4)),

$$R_N = \tilde{R}^N_{T_1,\dots,T_N}(X).$$

Proof of Proposition 2. We can assume that $H(Q|Q^0) < \infty$, and hence in view of Lemma 7 we may also assume that $m_Q := \mathbb{E}_Q |Y^1| < \infty$. Let us first consider a shift-ergodic Q. Note that $Q \in \mathscr{R} \cap \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ with $\mathbb{E}_Q L_1 < \infty$ implies $\Psi_Q = \nu^{\otimes\mathbb{N}}$, and hence $H(Q|Q^0) = -\mathbb{E}_Q \log \rho_{L_1} - H^c_L(Q)$ by Lemma 4. We can find a neighbourhood $\tilde{\mathcal{O}}_Q \subset O$ of the type defined in (59), and it suffices to restrict to $\tilde{\mathcal{O}}_Q$. For given $\varepsilon > 0$, take the open neighbourhood $\mathcal{U} \subset \mathcal{P}^{\mathrm{shift}}(E^{\mathbb{N}})$ of $\nu^{\otimes\mathbb{N}}$ guaranteed by Lemma 9. By the strong law, the event

$$\Big\{R_n(X) \in \mathcal{U} \text{ for all sufficiently large } n\Big\}$$

has probability one. As

$$\mathbb{P}(R_N \in \tilde{\mathcal{O}}_Q | X) \ge \sum_{\substack{0<j_1<\dots<j_N=[m_Q N],\\ \tilde{R}^N_{j_1,\dots,j_N}(X)\in\tilde{\mathcal{O}}_Q}} \prod_{i=1}^N \rho_{j_i - j_{i-1}},$$

we obtain

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{O}|X) \ge -H(Q|Q^0) - \varepsilon$$

by Lemma 9. Take $\varepsilon \to 0$ to conclude the proof in the ergodic case.

Now consider a general $Q \in \mathscr{R} \cap \mathcal{P}^{\mathrm{shift}}(\tilde{E}^{\mathbb{N}})$ with $\mathbb{E}_Q |Y^1| < \infty$. By the Ergodic Decomposition Theorem (cf e.g. [4], Thm. 5.2.16), we can represent

$$Q = \int_{\mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})} R \, \rho_Q(dR), \tag{62}$$

where ρ_Q is a probability measure on $\mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$. The event

$$\Big\{w-\lim_{L\to\infty}\frac{1}{L}\sum_{j=0}^{L-1}\delta_{\theta^j\kappa(Y)}=\nu^{\otimes\mathbb{N}}\Big\}$$

is invariant under the (word-level) shift $\tilde{\theta}$ and has Q-probability one, thus by (62), we have $\rho_Q(\mathscr{R})=1$. Furthermore, as $\mathbb{E}_Q\,|Y^1|=\int\mathbb{E}_R\,|Y^1|\,\rho_Q(dR)$, ρ_Q must be concentrated on $\{Q':\mathbb{E}_Q\,|Y^1|<\infty\}$. As $R\mapsto H(R|Q^0)$ is lower semicontinuous and affine, (62) implies $H(Q|Q^0)=\int H(R|Q^0)\,\rho_Q(dR)$, and Q can be approximated by finite convex combinations of $Q_i\in\mathscr{R}\cap\mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ in such a way that the corresponding specific relative entropies converge as well (see e.g. [4], Lemma 5.4.24 and its proof). More precisely, for any $\delta>0$, we can find $n\in\mathbb{N}$, $\lambda_1,\dots,\lambda_n\in(0,1)$ with $\sum_{i=1}^n\lambda_i=1$ and $Q_i\in\mathscr{R}\cap\mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ with $m_{Q_i}=\mathbb{E}_{Q_i}\,|Y^1|<\infty$ (so in particular $\Psi_{Q_i}=\nu^{\otimes\mathbb{N}}$) such that

$$\begin{aligned}\tilde{Q}:=\lambda_1Q_1+\cdots+\lambda_nQ_n\in\mathcal{O}\quad\text{and}\\ H(Q|Q^0)\geq H(\tilde{Q}|Q^0)-\delta=\lambda_1H(Q_1|Q^0)+\cdots+\lambda_nH(Q_n|Q^0)-\delta.\end{aligned}\tag{63}$$

Let $\tilde{\mathcal{O}}\subset\mathcal{O}$ be an open neighbourhood of $\tilde{Q}$ of the type defined in (59), and let $\tilde{\mathcal{O}}_m$ be corresponding open neighbourhoods of Q_m, $m=1,\dots,n$, but with $\tilde{\varepsilon}_i$ replaced by $\tilde{\varepsilon}_i/(2n)$. For (large) $N\in\mathbb{N}$ put $N_m:=[\lambda_mN]$, $\tilde{N}_m:=N_1+\cdots+N_m$ $(m=1,\dots,n)$ and $\bar{N}_m:=[N_1m_{Q_1}]+\cdots[N_mm_{Q_m}]$. Note that by construction for any $x\in E^{\bar{N}_n}$ and $j_1<\cdots<j_{\tilde{N}_n}$,

$$\begin{aligned}&\tilde{R}^{N_m}_{j_{\tilde{N}_{m-1}+1},j_{\tilde{N}_{m-1}+2},\dots,j_{\tilde{N}_m}}\Big(x\big|_{\left[(\bar{N}_{m-1}+1)\dots\bar{N}_m\right]}\Big)\in\tilde{\mathcal{O}}_m\ \text{ for } m=1,\dots,n\\ &\qquad\Longrightarrow\ \tilde{R}^{\tilde{N}_n}_{j_1,\dots,j_{\tilde{N}_n}}(x)\in\tilde{\mathcal{O}}.\end{aligned}$$

Let $\mathcal{U}_m$ be a neighbourhood of Ψ_{Q_m} $(=\nu^{\otimes\mathbb{N}})$ as constructed in Lemma 9 corresponding to $Q=Q_m$ and $\varepsilon=\delta$. Applying Lemma 9 separately on the stretches $X|_{[(\bar{N}_{m-1}+1)\dots\bar{N}_m]}$, $m=1,\dots,n$ and 'glueing together' the corresponding vectors of cut-points, we obtain from the discussion above that on the event

$$G_N:=\Big\{R_{[N_mm_{Q_m}]}(X|_{[(\bar{N}_{m-1}+1)\dots\bar{N}_m]})\in\mathcal{U}_m,\ m=1,\dots,n\Big\}$$

we have

$$\begin{aligned}\mathbb{P}(R_N\in\tilde{\mathcal{O}}|X)&\geq\prod_{m=1}^n\exp\Big(-N_m\Big(H(Q_m|Q^0)+\delta\Big)\Big)\\&\geq\exp\Big(-N\Big(\lambda_1H(Q_1|Q^0)+\cdots+\lambda_nH(Q_n|Q^0)+2\delta\Big)\Big)\\&\geq\exp\Big(-N\Big(H(Q|Q^0)+3\delta\Big)\Big)\end{aligned}$$

when N is sufficiently large. Now $\cup_M \cap_{N\geq M} G_N$ occurs almost surely (one can e.g. use large deviation results for the empirical distribution of X to see that $\mathbb{P}\Big((G_N)^c\Big)$ decays exponentially in N), hence

$$\liminf_{N\to\infty} \frac{1}{N} \log \mathbb{P}(R_N \in \mathcal{O}|X) \geq -H(Q|Q^0) - 3\delta.$$

Now take $\delta \to 0$. □

The following lemma is the combinatorial core of the lower bound, its intuitive content is that for a word x of length $\approx N m_Q$ which looks 'Ψ_Q-typical', there are $\approx \exp(N H_L^c(Q))$ ways of cutting it into N subwords in such a way that a 'Q-typical' sequence arises. The 'price' for any such pattern of cut points will then be $\approx \exp(N \mathbb{E}_Q \log \rho(|Y^1|))$.

Lemma 9 *Let $Q \in \mathcal{P}^{\mathrm{erg}}(\tilde{E}^{\mathbb{N}})$ with $m_Q := \mathbb{E}_Q[L_1] < \infty$ be given, and let $\tilde{\mathcal{O}}_Q$ be a neighbourhood of Q as defined in (59). For any $\varepsilon > 0$ there exists an open neighbourhood $\mathcal{U} \subset \mathcal{P}^{\mathrm{shift}}(E^{\mathbb{N}})$ of Ψ_Q and $N_0 \in \mathbb{N}$ such that*

$$N \geq N_0, \ x \in E^{[m_Q N]} \ \text{ with } \ R_{[m_Q N]}(x) \in \mathcal{U}$$

implies

$$\sum_{\substack{0<j_1<\cdots<j_N=[m_Q N], \\ \tilde{R}^N_{j_1,\dots,j_N}(x)\in\tilde{\mathcal{O}}_Q}} \prod_{i=1}^{N} \rho(j_i - j_{i-1}) \geq \exp\Big(N\Big(\mathbb{E}_Q \log \rho_{L_1} + H_L^c(Q) - \varepsilon\Big)\Big). \quad (64)$$

Proof. Step 1. Let $\tilde{\mathcal{O}}'_Q$ be defined as in (59) with $\tilde{\varepsilon}_i$ replaced by $\tilde{\varepsilon}_i/2$ ($i = 1, \dots, A_{\mathcal{O}_Q}$), and similarly $\tilde{\mathcal{O}}''_Q$ with $\tilde{\varepsilon}_i$ replaced by $\tilde{\varepsilon}_i/4$. For $M \in \mathbb{N}$, $\varepsilon_1 > 0$, $x \in E^{[M m_Q]}$ let

$$\mathscr{J}_{M,\varepsilon_1}(x) := \left\{ (j_1,\dots,j_M) : \begin{array}{c} 0 \leq j_1 < \cdots < j_M = [M m_Q], \tilde{R}^M_{j_1,\dots,j_M}(x) \in \tilde{\mathcal{O}}'_Q, \\ \frac{1}{M}\sum_{i=1}^M \log \rho_{j_i - j_{i-1}} \in [\mathbb{E}_Q \log \rho_{L_1} - \varepsilon_1, \mathbb{E}_Q \log \rho_{L_1} + \varepsilon_1] \end{array} \right\}$$

This is the set of all cut-vectors which are 'suitable' for the given word x. We claim that for given $\varepsilon_2 > 0$ we can choose M sufficiently large and pairwise different words $\xi^1, \dots, \xi^L \in E^{[M m_Q]}$ such that

$$\sum_{i=1}^{L} \Psi_Q|_{[1\dots M m_Q]}(\xi^i) \geq 1 - \varepsilon_2 \qquad \text{and} \quad (65)$$

$$\Big|\mathscr{J}_{M,\varepsilon_1}(\xi^i)\Big| \geq \exp\Big(M(H_L^c(Q) - \varepsilon_2)\Big), \quad i = 1, \dots, L. \quad (66)$$

In order to check this let $\hat{Q}$ be defined as in (21), recall $(d\hat{Q}/dQ)(Y) = |Y^1|/m_Q$. By Lemma 3 and the fact that $\hat{Q} \ll Q$ we have $\hat{Q}$-a.s.

$$\begin{aligned}
\lim_{N\to\infty} \frac{1}{N}|\kappa(Y^1,\dots,Y^N)| &= m_Q \\
\lim_{N\to\infty} \frac{1}{N} \log \hat{Q}(\kappa(Y^1,\dots,Y^N)) &= -m_Q H(\Psi_Q), \\
\lim_{N\to\infty} \frac{1}{N} \log \hat{Q}(Y^1,\dots,Y^N) &= -H(Q), \\
\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^N \log \rho(|Y^i|) &= \mathbb{E}_Q \log \rho(|Y^1|), \\
\lim_{N\to\infty} R_N &= Q \in \tilde{\mathcal{O}}''_Q.
\end{aligned}$$

Thus, for large enough N we can find A pairwise different $z^i \in \tilde{E}$ and for each z^k we can choose B_k different decompositions $(y^{k,j,1},\dots,y^{k,j,N}) \in \tilde{E}^N$, $j = 1,\dots,B_k$, where

$$\kappa(y^{k,j,1},\dots,y^{k,j,N}) = z^k \quad \text{for each } j,k,$$

such that each $|z^k| \in [N(m_Q - \varepsilon_1), N(m_Q - \varepsilon_1)]$,

$$\sum_{k=1}^A \hat{Q}(K^N = z^k) \geq 1 - \varepsilon_1 \tag{67}$$

and the following holds for $k = 1,\dots,A$ and $j = 1,\dots,B_k$ (unless otherwise quantified):

$$\hat{Q}(K^N = z^k) \geq \exp(-N(m_Q H(\Psi_Q) + \varepsilon_1)), \tag{68}$$

$$\hat{Q}\Big((Y^1,\dots,Y^N) = (y^{k,j,1},\dots,y^{k,j,N})\Big) \leq \exp(-N(H(Q) - \varepsilon_1)), \tag{69}$$

$$\sum_{j=1}^{B_k} \hat{Q}\Big((Y^1,\dots,Y^N) = (y^{k,j,1},\dots,y^{k,j,N})\Big) \geq (1-\varepsilon_1)\hat{Q}(K^N = z^k), \tag{70}$$

$$\frac{1}{N}\sum_{i=1}^N \log \rho(|y^{k,j,i}|) \;\in\; [\mathbb{E}_Q \log \rho(|Y^1|) - \frac{\varepsilon_1}{2}, \mathbb{E}_Q \log \rho(|Y^1|) + \frac{\varepsilon_1}{2}], \tag{71}$$

$$\frac{1}{N}\sum_{i=1}^{N-1} \delta_{\tilde{\theta}^i}\Big((y^{k,j,1},\dots,y^{k,j,N})^{\mathrm{per}}\Big) \in \tilde{\mathcal{O}}''_Q. \tag{72}$$

Note that (69), (70) and (68) imply for each k that

$$
\begin{aligned}
B_k e^{-N(H(Q)-\varepsilon_1)} &\geq \sum_{j=1}^{B_k} \hat{Q}\Big((Y^1,\dots,Y^N) = (y^{k,j,1},\dots,y^{k,j,N})\Big) \\
&\geq (1-\varepsilon_1)\hat{Q}(K^N = z^k) \geq (1-\varepsilon_1) e^{-N(m_Q H(\Psi_Q)+\varepsilon_1)} \\
&\geq \exp\Big(-N(m_Q H(\Psi_Q)+2\varepsilon_1)\Big)
\end{aligned}
$$

for N large enough, hence

$$
B_k \geq \exp\Big(N(H(Q) - m_Q H(\Psi_Q) - 3\varepsilon_1)\Big) = \exp\Big(N(H_L^c(Q) - 3\varepsilon_1)\Big)
$$

for $k = 1,\dots,A$ by Lemma 3. Note that this together with (71) and (72) shows that

$$
\Big|\mathscr{I}_{N,\varepsilon_1}(z^k)\Big| \geq \exp\Big(N(H_L^c(Q) - 3\varepsilon_1)\Big) \quad \text{for } k = 1,\dots,A \tag{73}
$$

(with a notational grain of salt because $|z^k|$ is not exactly $[Nm_Q]$). This is almost what we need to prove (65) and (66), except for the slight nuisance that the z^k have not exactly length $[Nm_Q]$ and the fact that (67) guarantees that one of the z^k is very likely to occur as the concatenation of the first N words under $\hat{Q}$, whereas (65) speaks about the first $[Nm_Q]$ letters under Ψ_Q. Remembering the definition (22) of Ψ_Q involving $\hat{Q}$, this can e.g. be remedied as follows: Pick M so large that (67)–(73) are satisfied for $N = M$. Consider the set of words

$$
\begin{aligned}
&\Big\{\tilde{\xi}^r : r = 1,\dots,R\Big\} \\
&\quad := \Big\{\theta^i\Big(\kappa(y^{k,j,1},\dots,y^{k,j,N})\Big) \;:\; 0 \leq i < |y^{k,j,1}|,\ j = 1,\dots,B_k,\ k = 1,\dots,A\Big\},
\end{aligned}
$$

truncate each of them at $[M(m_Q - 2\varepsilon_1]$ letters. Thus in view of (22) and (67),

$$
\sum_{r=1}^{R} \Psi_Q\big|_{[1\dots M(m_Q-2\varepsilon_1)]}(\tilde{\xi}^r|_{[1\dots M(m_Q-2\varepsilon_1)]}) \geq 1 - \varepsilon_1.
$$

Now generate a set $\{\xi^i : i = 1,\dots,L\}$ from $\{\tilde{\xi}^r : r = 1,\dots,R\}$ by attaching various suffixes of length $[Mm_Q] - [M(m_Q - 2\varepsilon_1)]$ to each $\tilde{\xi}^r$ in such a way that $\sum_{i=1}^{L} \Psi_Q|_{[1\dots Mm_Q]}(\xi^i) \geq 1 - 2\varepsilon_1$. As each ξ^i agrees with some z^k except for a very short initial piece and a short final piece, (73) implies $\Big|\mathscr{I}_{M,\varepsilon_1}(z^k)\Big| \geq \exp\Big(M(H_L^c(Q) - 4\varepsilon_1)\Big)$ for each i when M is large enough. By choosing ε_1 small enough, this proves (65) and (66).

Step 2. Let $\mathscr{A} := \{\xi^i : i = 1,\dots,L\}$ denote the set of words of length $[Mm_Q]$ constructed in Step 1. For $K \geq [Mm_Q]$ (we think of $K \gg [Mm_Q]$) and $\varepsilon_3 > 0$

denote the set of all $x \in E^K$ such that

$$\left| \frac{1}{N} \#\{1 \le j \le K - [Mm_Q] : (x_j, \ldots, x_{j+[Mm_Q]-1}) = \xi^i\} - \Psi_Q|_{[1\ldots Mm_Q]}(\xi^i) \right| < \varepsilon_3/L$$

for all $\xi^i \in \mathscr{A}$ by $\mathscr{D}_{K,\varepsilon_3}$. Note that $x \in \mathscr{D}_{K,\varepsilon_3}$ means that the letter sequence x is typical for Ψ_Q in the sense that the frequency of all the patterns ξ^i $(i = 1, \ldots, L)$ chosen above is close to the theoretical value.

We claim that for any $\varepsilon > 0$ we can choose above $\varepsilon_1, \varepsilon_2, \varepsilon_3$ sufficiently small and L, M sufficiently large and $N_0 \in \mathbb{N}$ such that

$$\begin{aligned} &N \ge N_0,\ x \in \mathscr{D}_{[m_Q N],\varepsilon_3} \\ &\Longrightarrow \sum_{\substack{0<j_1<\cdots<j_N=[m_Q N], \\ \tilde{R}^N_{j_1,\ldots,j_N}(x) \in \tilde{\mathcal{O}}_Q}} \prod_{i=1}^N \rho_{j_i - j_{i-1}} \ge \exp\Big(N \Big(\mathbb{E}_Q \log \rho_{L_1} + H^c_L(Q) - \varepsilon \Big) \Big). \end{aligned} \tag{74}$$

Note that (74) implies the claim of the lemma by choosing $\mathcal{U}$ as

$$\Big\{ \Psi \in \mathcal{P}^{\text{shift}}(E^{\mathbb{N}}) : \Big| \Psi|_{[1\ldots m_Q M]}(\xi^i) - \Psi_Q|_{[1\ldots m_Q M]}(\xi^i) \Big| < \varepsilon_3/(2L),\ i = 1, \ldots, L \Big\}.$$

Step 3. It remains to prove (74), the idea is as follows: $x \in \mathscr{D}_{[m_Q N],\varepsilon_3}$ implies that we can cover x with $\approx N/M$ non-overlapping patterns from $\mathscr{A} := \{\xi^i, i = 1, \ldots, L\}$, up to a small fraction of remaining 'gaps'. On each of the patterns from the 'almost covering', we have by construction sufficiently many choices of 'good cut-points', and the probability that the jumps bridge exactly the given 'gaps' is controlled on the exponential scale because the total gap length is only a small fraction of N. Here are the details:

$x \in \mathscr{D}_{[m_Q N],\varepsilon_3}$ implies

$$\#\Big\{ j \le [m_Q N] - [m_Q M] : \text{one of the words from } \mathscr{A} \text{ starts at } j \Big\} \ge [N m_Q](1 - \varepsilon_2 - \varepsilon_3)$$

(as $\sum_{i=1}^L (\Psi_Q(\xi^i) - \varepsilon_3/L) \ge 1 - \varepsilon_2 - \varepsilon_3$). Let $n_1 := n_0 \vee (2m_Q/\delta_0)$, where n_0, δ_0 are as given by Lemma 10. When N is large enough, we can find

$$\tilde{\varepsilon} \in [\varepsilon/2, 2\varepsilon] \tag{75}$$

($\tilde{\varepsilon}$ will implicitly depend on N because we require certain expressions below to be integers, but (75) will be satisfied independently of N) such that

$$k = (1 - \tilde{\varepsilon})\frac{N}{M} \in \mathbb{N}$$

and we can find k positions

$$n_1 \leq r_1 < \cdots < r_k \leq [Nm_Q] - n_1$$

where one of the patterns from $\mathscr{A}$ is written on x, i.e.

$$x|_{[r_j \dots r_j + [Mm_Q] - 1]} = \xi^i \quad \text{for some } i \in \{1, \dots, L\}, \qquad j = 1, \dots, k\,,$$

and

$$r_j - r_{j-1} \geq [m_Q M] + n_1, \quad j = 1, \dots, k.$$

Note that between the end of the $(j-1)$-th and the beginning of the j-th subword from $\mathscr{A}$ on x, there is a 'gap' of length

$$s_j := r_j - r_{j-1} - [m_Q M] \ (\geq n_1).$$

The total length of these gaps is

$$s_1 + \cdots + s_{k+1} = [Nm_Q] - (1 - \tilde{\varepsilon})\frac{N}{M}[Mm_Q] = \varepsilon' N m_Q.$$

The display above implicitly defines ε', when N (and M) are large enough, it will satisfy

$$\varepsilon' \in \Big[\tilde{\varepsilon}(1 - \delta_0/2), \tilde{\varepsilon}(1 + \delta_0/2)\Big], \tag{76}$$

where δ_0 is as given by Lemma 10.

The sum appearing in (74) has N summation variables, and on each of the k 'good subwords of x' fixed above, we will use M of them. Thus there remain

$$N - kM = \tilde{\varepsilon} N$$

summation variables which we can use to 'fill the gaps'. We can find $m_1, \dots, m_{k+1} \in \mathbb{N}$ such that

$$m_1 + \cdots + m_{k+1} = \tilde{\varepsilon} N$$

and

$$(1 - \delta_0)\frac{s_j}{m_Q} \leq m_j \leq (1 + \delta_0)\frac{s_j}{m_Q}, \quad j = 1, \dots, k+1.$$

To see this consider first $\tilde{m}_i := (s_i/m_Q)(\tilde{\varepsilon}/\varepsilon')$. Then we have $\sum \tilde{m}_i = \tilde{\varepsilon} N$, but the $\tilde{m}_i$ need not be integers. On the other hand we have

$$\frac{s_i}{m_Q}(1 - \delta_0) \leq \tilde{m}_i - 1 \leq \lceil \tilde{m}_i \rceil \leq \tilde{m}_i + 1 \leq \frac{s_i}{m_Q}(1 + \delta_0)$$

and $S := \sum_{i=1}^{k+1}(\tilde{m}_i - \lceil \tilde{m}_i \rceil) \in \{0, 1, \dots, k+1\}$. Then put e.g. $m_i := \lceil \tilde{m}_i \rceil + 1$ if $i \leq S$, $m_i := \lceil \tilde{m}_i \rceil$ otherwise.

In order to generate vectors $(j_1, \dots, j_N)$ suitable for the righthand side of (74), we can proceed as follows: On the 'good subwords', choose any M-vector of cut-points from the corresponding $\mathscr{J}_{M,\varepsilon_1}$, and use m_i summation variables to generate the 'jump' over the i-th gap. Using Lemma 10, we can choose for each gap m_i cut-lengths whose total probability under ρ is at least $\exp(-Cs_i)$. By the definition of $\mathscr{J}_{M,\varepsilon_1}$, any such vector $(j_1, \dots, j_N)$ will have the property that $R^N_{j_1,\dots,j_N}(x) \in \tilde{\mathcal{O}}_Q$ because the contribution to $R^N_{j_1,\dots,j_N}(x)$ from the 'gaps' is negligible. Furthermore, again by the definition of $\mathscr{J}_{M,\varepsilon_1}$ and the choice of the cut-points on the gaps, we have

$$\begin{aligned}\prod_{i=1}^{N} \rho_{j_i - j_{i-1}} &\geq \exp\Big(kM\Big(\mathbb{E}_Q \log \rho_{L_1} - \varepsilon_1\Big)\Big) \times \exp\Big(-C\varepsilon' m_Q N\Big) \\ &= \exp\Big(N\Big((1-\tilde{\varepsilon})(\mathbb{E}_Q \log \rho_{L_1} - \varepsilon_1) - \varepsilon' C m_Q\Big)\Big)\end{aligned}$$

for each such choice. Finally, by (66) there are at least

$$\Big(\exp\Big(M(H^c_L(Q) - \varepsilon_2)\Big)\Big)^k = \exp\Big(N(1-\tilde{\varepsilon})(H^c_L(Q) - \varepsilon_2)\Big)$$

admissible choices of cut-points on the good subwords. Combining, we obtain that the righthand side of (74) is at least

$$\exp\Big(N\Big((1-\tilde{\varepsilon})(\mathbb{E}_Q \log \rho_{L_1} + H^c_L(Q) - \varepsilon_1 - \varepsilon_2) - \varepsilon' C m_Q\Big)\Big)$$

whenever N is sufficiently large. □

The following lemma is a standard result about aperiodic renewal processes:

Lemma 10 *Let $m_Q \in [1, \infty)$. There exist $n_0 \in \mathbb{N}$, $C > 0$ and $\delta_0 > 0$ such that for any pair $(m, n) \in \mathbb{N}^2$ satisfying*

$$n \geq n_0, \ \frac{n}{m_Q}(1 - \delta_0) \leq m \leq \frac{n}{m_Q}(1 + \delta_0)$$

there are $\ell_1, \dots, \ell_m$ with

$$\ell_1 + \dots + \ell_m = n \quad \textit{and} \quad \prod_{i=1}^{m} \rho_{\ell_i} \geq \exp(-Cn).$$

Remark 11 Note that the proof of Proposition 2 via Lemma 9 is rather combinatoric. At least in the case of a neighbourhood of an *ergodic* $Q \in \mathscr{R}$, one can use the coupling between X and a shift of $\kappa(Y)$ under $\hat{Q}$ given by Remark 6 to employ a 'conditional tilting' argument which is more in the probabilistic spirit of classical proofs of lower large deviation bounds.

Acknowledgement

The author would like to thank Andreas Greven and Frank den Hollander for very stimulating discussions on this subject.

References

[1] Erwin Bolthausen, Frank den Hollander, Localization transition for a polymer near an interface. *Ann. Probab.* **25** (1997), no. 3, 1334–1366.

[2] Francis Comets, Large deviation estimates for a conditional probability distribution. Applications to random interaction Gibbs measures. *Probab. Theory Related Fields* **80** (1989), no. 3, 407–432.

[3] Amir Dembo, Ofer Zeitouni, *Large deviations techniques and applications.* Jones and Bartlett Publishers, Boston, 1993.

[4] Jean-Dominique Deuschel, Daniel W. Stroock, *Large deviations.* Academic Press, Boston, 1989.

[5] Richard S. Ellis, *Entropy, large deviations, and statistical mechanics.* Springer-Verlag, New York, 1985.

[6] Hans Föllmer, Steven Orey, Large deviations for the empirical field of a Gibbs measure. *Ann. Probab.* **16** (1988), no. 3, 961–977.