

Longest common substring for random subshifts of finite type

Jérôme Rousseau

Departamento de Matemática, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal.

E-mail: jerome.rousseau@ufba.br; url: www.sd.mat.ufba.br/~jerome.rousseau

Departamento de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, 40170-110 Salvador, Brazil

Received 24 October 2019; revised 11 September 2020; accepted 17 November 2020

Abstract. In this paper, we study the behaviour of the longest common substring for random subshifts of finite type (for dynamicists) or of the longest common substring for random sequences in random environments (for probabilists). We prove that, under some exponential mixing assumptions, this behaviour is linked to the Rényi entropy of the stationary measure. We emphasize that what we establish is a quenched result.

Résumé. Dans cet article, nous étudions le comportement de la plus longue sous-chaîne commune pour des sous-shifts aléatoires de type fini (pour les dynamiciens) ou de la plus longue sous-chaîne commune pour des suites aléatoires en milieux aléatoires (pour les probabilistes). Nous prouvons que, sous des hypothèses de mélange exponentiel, ce comportement est lié à l'entropie de Rényi de la mesure stationnaire. Nous soulignons que ce que nous établissons est un résultat de type «quenched».

MSC2020 subject classifications: 60F15; 60K37; 37A50; 37A25; 37Hxx; 94A17; 92D20

Keywords: Longest common substring; Rényi entropy; Random dynamical systems; Random sequences in random environments; String matching

Introduction

To try and measure the similarity between sequences, one has to develop computational tools to compare the sequences (and to optimize the algorithm) and probabilistic tools to discern the significance of the relationship. Thus sequences comparison (and in particular sequences alignment and sequences matching) takes its roots in computer science and probability and has applications in areas as diverse as bioinformatics, geology, linguistics or social sciences. We refer the reader to [31,39] for a broad introduction to sequences comparison (with a particular attention to biology).

One particularly relevant object in DNA comparison is the longest common substring, i.e. the longest string of DNA which appears in two (or more) strands. For example, for the following two strands

ACAATGAGAGGATGACCTTG

TGACTGTAAGTACACAAGC

a longest common substring is ACAA (TGAC is also a longest common substring) and is of length 4 when the total length of the strands is 20. A way to distinguish if this behaviour is common or rare is to obtain probabilistic results which allow us to understand the statistical significance of our comparison.

In this paper, we will concentrate on the behaviour of the length of the longest common substring when the length of the strings grows, more precisely, for two sequences x and y , the behaviour, when n goes to infinity, of

$$M_n(x, y) = \max\{m : x_{i+k} = y_{j+k} \text{ for } k = 1, \dots, m \text{ and for some } 0 \leq i, j \leq n - m\}.$$

For sequences drawn randomly from the same alphabet, this problem was studied by Arratia and Waterman in [5]. More precisely, if each term of the sequences is drawn independently within some alphabet \mathcal{A} with respect to some probability \mathcal{P} , then they proved that for $\mathcal{P}^{\mathbb{N}} \otimes \mathcal{P}^{\mathbb{N}}$ -almost every $(x, y) \in \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$

$$\lim_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} = \frac{2}{-\log p},$$

where $p = \sum_{a \in \mathcal{A}} \mathcal{P}(a)^2$.

They also proved the same result for independent irreducible and aperiodic Markov chains on a finite alphabet, and in this case p is the largest eigenvalue of the matrix $[(p_{ij})^2]$ (where $[p_{ij}]$ is the transition matrix).

In fact, one can observe that in both case, $-\log p$ corresponds to the Rényi entropy of μ defined (provided that it exists) by

$$H_2(\mu) = \lim_{k \rightarrow \infty} \frac{\log \sum \mu(C_k)^2}{-k},$$

where the sums are taken over all k -cylinders. Even if the existence of the Rényi entropy is not known in general, it was computed in some particular cases: Bernoulli shift, finite state Markov chains, Gibbs measure of a Hölder-continuous potential [20] and infinite state Markov chains [11]. The existence was also proved for ϕ -mixing measures [27], for weakly ψ -mixing processes [20] and for ψ_g -regular processes [1,2].

Generalizations of the work [5] to sequences of different lengths, different distributions, more than two sequences, extreme value theory for sequence matching and distributional results can be found in e.g. [4,6–8,15,21,28,29]. In a similar direction, one can also see [12,14] (and references therein) where the authors investigate the growth rate of the maximal overlap in a string (i.e. the growth rate of the length of the longest repeated substring). We also refer to [2,3,25,37] for relatively close problems.

Recently, in [9], the results of Arratia and Waterman were generalized to α -mixing systems with exponential decay (and ψ -mixing with polynomial decay) and it was proved that if the Rényi entropy exists then for $\mu \otimes \mu$ -almost every (x, y)

$$\lim_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} = \frac{2}{H_2(\mu)}.$$

Furthermore, it was also shown in this paper that a generalization of the longest common substring problem for dynamical systems is to study the behaviour of the shortest distance between two orbits, which is, for a dynamical system (X, T, μ) , the behaviour, when n goes to infinity, of

$$m_n(x, y) = \min_{i, j=0, \dots, n-1} (d(T^i x, T^j y)).$$

Moreover, a relation between m_n and the correlation dimension of the invariant measure was proved.

It is natural to try and obtain the same type of results for random dynamical systems since they could model more precisely physical phenomena. For random sequences, this could correspond for example to a modification (e.g. a small perturbation) on the probability with which the letters of the alphabet are drawn (i.e. random sequences in random environments). For dynamical systems, this could correspond to adding some random noise or small perturbations while iterating the same transformation, or iterating different transformations drawn randomly within a family of transformations (see e.g. [24] for an introduction to random dynamical systems).

In [13], the behaviour of the longest common substring of encoded sequences (and of the shortest distance between observed orbits) were studied and a relation with the Rényi entropy of the pushforward measure was proved. In particular, it allows the authors to obtain annealed results on the shortest distance between orbits of random dynamical systems.

Obtaining quenched results is much more delicate, in particular because generally the random maps do not have a common invariant measure. The first family of random dynamical systems to study and where one can hope to obtain results are random subshifts of finite type. Indeed, good mixing properties have been proved (see e.g. [10,22,23,38]) which allows to get other statistical properties (e.g. [19,34,35] for the distribution of hitting times, [18] for extreme value laws). Following this idea and the setting of these papers, we study here the behaviour of the longest common substrings for random subshifts of finite type (in probabilistic language, this corresponds to the longest common substring for random sequences in random environments) and prove a link with the Rényi entropy of the stationary measure.

The paper is organized as follows. In Section 1, we will define random subshifts of finite type, explain our assumptions and give an upper bound (Theorem 2) and a lower bound (Theorem 3 and Theorem 4) for the growth rate of the longest common substring for random subshifts. In Section 2, we will apply our results to random Bernoulli shifts and random Gibbs measures. The proof of the theorems will be given in Section 3.

1. Statement of the main results

We first give the definition of a random subshift of finite type. Let $(\Omega, \theta, \mathbb{P})$ be an invertible ergodic measure preserving system, set $X = \{1, \dots, N\}^{\mathbb{N}}$ for some $N \in \mathbb{N}$ and let $\sigma : X \rightarrow X$ denote the shift. Let $b : \Omega \rightarrow \{1, \dots, N\}$ be a random variable. Let $A = \{A(\omega) = (a_{ij}(\omega)) : \omega \in \Omega\}$ be a random transition matrix, i.e. for any $\omega \in \Omega$, $A(\omega)$ is a $b(\omega) \times b(\theta\omega)$ -matrix with entries in $\{0, 1\}$, at least one non-zero entry in each row and each column and such that $\omega \mapsto a_{ij}(\omega)$ is measurable for any $i \in \mathbb{N}$ and $j \in \mathbb{N}$. For any $\omega \in \Omega$ define the subset of the integers $X_\omega = \{1, \dots, b(\omega)\}$ and

$$\mathcal{E}_\omega = \{x = (x_0, x_1, \dots) : x_i \in X_{\theta^i \omega} \text{ and } a_{x_i x_{i+1}}(\theta^i \omega) = 1 \text{ for all } i \in \mathbb{N}\} \subset X,$$

$$\mathcal{E} = \{(\omega, x) : \omega \in \Omega, x \in \mathcal{E}_\omega\} \subset \Omega \times X.$$

We consider the random dynamical system coded by the skew-product $S : \mathcal{E} \rightarrow \mathcal{E}$ given by $S(\omega, x) = (\theta\omega, \sigma x)$. Let ν be an S -invariant probability measure with marginal \mathbb{P} on Ω and let $(\mu_\omega)_\omega$ denote its decomposition on \mathcal{E}_ω , that is, $d\nu(\omega, x) = d\mu_\omega(x) d\mathbb{P}(\omega)$. The measures μ_ω are called the *sample measures*. Note $\mu_\omega(A) = 0$ if $A \cap X_\omega = \emptyset$. We denote by $\mu = \int \mu_\omega d\mathbb{P}$ the marginal of ν on X .

We emphasize that the sample measures are not invariant. However, since θ is invertible, by σ -invariance of ν and almost everywhere uniqueness of the decomposition $d\nu = d\mu_\omega d\mathbb{P}$, we get for \mathbb{P} -almost every $\omega \in \Omega$,

$$(\sigma^i)_* \mu_\omega = \mu_{\theta^i \omega} \quad \text{for all } i \in \mathbb{N}. \tag{1}$$

For $y \in X$ we denote by $C_n(y) = \{z \in X : y_i = z_i \text{ for all } 0 \leq i \leq n - 1\}$ the n -cylinder that contains y . Set $\mathcal{F}_0^n(X)$ as the sigma-algebra in X generated by all the n -cylinders.

As explain in the introduction, for two sequences $x, y \in X$, we are interested in the asymptotic behaviour of the longest common substring, that is the behaviour of

$$M_n(x, y) = \max\{m : x_{i+k} = y_{j+k} \text{ for } k = 1, \dots, m \text{ and for some } 0 \leq i, j \leq n - m\}.$$

We will show it is linked to the Rényi entropy of the stationary measure μ . Thus, we define the lower and upper Rényi entropies of the measure μ :

$$\underline{H}_2(\mu) = \liminf_{k \rightarrow \infty} \frac{\log \sum \mu(C_k)^2}{-k} \quad \text{and} \quad \overline{H}_2(\mu) = \limsup_{k \rightarrow \infty} \frac{\log \sum \mu(C_k)^2}{-k},$$

where the sums are taken over all k -cylinders. When the limit exists we denote by $H_2(\mu)$ the common value.

To obtain our results, we will need information on the decay of the measure of cylinders, thus we define

$$h_0 = \liminf_{k \rightarrow +\infty} \frac{\log \int_\Omega \max_{C_k} \mu_\omega(C_k) d\mathbb{P}}{-k},$$

where the max is taken over all k -cylinders.

We will assume the following: there is a constant $a \in [0, 1)$ and a function $\alpha(g)$ satisfying $\alpha(g) = \mathcal{O}(a^g)$ such that for all $n, m, A \in \mathcal{F}_0^n(X)$ and $B \in \mathcal{F}_0^m(X)$:

(I) the marginal measure μ satisfies

$$|\mu(A \cap \sigma^{-g-n} A) - \mu(A)^2| \leq \alpha(g);$$

(II) (fibered exponential α -mixing) for \mathbb{P} -almost every $\omega \in \Omega$

$$|\mu_\omega(A \cap \sigma^{-g-n} B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)| \leq \alpha(g).$$

One can observe that assumption (I) is weaker than α -mixing since in the intersection we only deal with the same cylinder A . We recall that the measure μ is α -mixing if:

(I-a) (exponential α -mixing) the marginal measure μ satisfies

$$|\mu(A \cap \sigma^{-g-n} B) - \mu(A)\mu(B)| \leq \alpha(g)$$

for all $m, n, A \in \mathcal{F}_0^n(X)$ and $B \in \mathcal{F}_0^m(X)$.

Before stating our results, we will consider the annealed case:

Theorem 1 (Theorem 4.4 [13]). *If $0 < H_2(\mu)$, then*

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} \leq \frac{2}{H_2(\mu)} \quad \text{for } \nu \otimes \nu\text{-almost every } ((\omega, x), (\tilde{\omega}, y)) \in \mathcal{E} \times \mathcal{E}.$$

Moreover, if hypothesis (I-a) holds, then

$$\underline{\lim}_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} \geq \frac{2}{H_2(\mu)} \quad \text{for } \nu \otimes \nu\text{-almost every } ((\omega, x), (\tilde{\omega}, y)) \in \mathcal{E} \times \mathcal{E}.$$

First of all, we observe that the statement of this theorem is slightly different that the one of Theorem 4.4 in [13] since they consider more general dynamical systems and not only random subshifts of finite type. Nevertheless, one can adapt easily their results and proof to obtain the theorem as stated here.

One could wonder why the Rényi entropy of μ appears in these results (and not the Rényi entropy of ν for example). In fact, when studying M_n , we are not interested on the behaviour of the whole orbits $S^n(\omega, x)$ but only its projection on X (called an observation of the dynamical system). More precisely, if $\pi : \mathcal{E} \rightarrow X$ denotes the canonical projection (i.e. $\pi(\omega, x) = x$), we study the behaviour of the image (or observation) of the orbits, that is $\pi(S^n(\omega, x))$. The idea of observing dynamical systems was developed in [32,33] to obtain annealed results for return times in random dynamical systems and for the shortest distance between random orbits in [13]. Moreover, it was proved, that when observing dynamical systems, these quantities are linked with the dimension (or in our case the Rényi entropy) of the pushforward measure $\pi_*\nu$ (where $\pi_*\nu(\cdot) = \nu(\pi^{-1}(\cdot))$). Furthermore, in our random setting the pushforward measure $\pi_*\nu$ and the measure μ are equals (e.g. [32, Proof of Theorem 8]) and thus $H_2(\pi_*\nu) = H_2(\mu)$.

Unfortunately, these technics only give annealed results, thus, in this paper, we will use different tools to obtain quenched results.

Remark 1. We note that since S is a dynamical system, one could apply (under the right assumptions) the results of [9] to study the shortest distance between orbits $m_n((\omega, x), (\tilde{\omega}, y)) = \min_{i,j=0,\dots,n-1} (d(S^i(\omega, x), S^j(\tilde{\omega}, y)))$ and link it to the correlation dimension of ν . Nevertheless, it will not give us precise informations on $M_n(x, y)$ since $m_n((\omega, x), (\tilde{\omega}, y))$ takes into account the distance between elements of the orbits of (ω, x) and $(\tilde{\omega}, y)$ while M_n only considers elements of the orbits of x and y .

We present now the first main result of this section which gives an upper bound for the growth rate of the longest common substring.

Theorem 2. *If $0 < H_2(\mu) \leq 2h_0$ and if hypothesis (I) and (II) hold, then for \mathbb{P} -almost every $\omega \in \Omega$,*

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} \leq \frac{2}{H_2(\mu)} \quad \text{for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x, y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

One can notice that in the deterministic case [9] and in the annealed case, no mixing assumptions are needed to obtain the upper bound. As one can see in the proof of this theorem, the main problem and difference with the deterministic case is that the sample measures are not invariant which is the main reason to use mixing to obtain the upper bound (and the lower).

Moreover, one can observe that assuming $H_2(\mu) \leq 2h_0$ is not a too restrictive assumption. Indeed, in the deterministic case this hypothesis is always satisfied (see e.g. [20] in the proof of Theorem 1(IV)). In the random setting, this assumption prohibits for example to have some sample measures with an extreme behaviour (relatively with the others).

To obtain a lower bound, we will need stronger assumptions: we will need α -mixing for the measure μ and we will require some mixing properties for the base transformation $(\Omega, \theta, \mathbb{P})$.

First of all, we will treat the case when $(\Omega, \theta, \mathbb{P})$ is a ρ -mixing two-sided shift, i.e. $\Omega = \mathcal{A}^{\mathbb{Z}}$ for some alphabet \mathcal{A} , θ is the shift and:

(III) (exponential ρ -mixing) For all n and for all $\psi \in L^2(\mathcal{F}_{-\infty}^n(\Omega))$ and $\phi \in L^2(\mathcal{F}_0^\infty(\Omega))$

$$\left| \int_{\Omega} \psi \cdot \phi \circ \theta^{n+s} d\mathbb{P} - \int_{\Omega} \psi d\mathbb{P} \int_{\Omega} \phi d\mathbb{P} \right| \leq \rho(s) \|\psi\|_2 \|\phi\|_2$$

with $\rho(n) = \mathcal{O}(a^n)$.

Moreover, we will need that the sample measure μ_ω of a cylinder of size n does not depend on all the terms of ω :

(IV) there exists a function h with $h(n) = \mathcal{O}(n)$ such that for \mathbb{P} -almost every ω and every cylinder $C \in \mathcal{F}_0^n(X)$, the function $\omega \mapsto \mu_\omega(C)$ belongs to $L^2(\mathcal{F}_{-h(n)}^{h(n)}(\Omega))$.

One can observe that it is quite simple to check if assumption (IV) is satisfied, however this assumption is restrictive and only enables us to work with some special family of sample measures. Nevertheless, if the system $(\Omega, \theta, \mathbb{P})$ satisfies some stronger mixing assumption we will be able to work with more general families of sample measures. Thus, after the statement of the next theorem we will give an alternative couple of assumptions which also allows us to obtain a lower bound for the growth rate of the longest common substring.

Theorem 3. *If $0 < \underline{H}_2(\mu) \leq \overline{H}_2(\mu) < 2h_0$ and if hypothesis (I-a), (II), (III) and (IV) hold, then, for \mathbb{P} -almost every $\omega \in \Omega$,*

$$\liminf_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} \geq \frac{2}{H_2(\mu)} \quad \text{for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x, y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

Moreover, if the Rényi entropy exists, we get for \mathbb{P} -almost every $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} = \frac{2}{H_2(\mu)} \quad \text{for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x, y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

In Section 2.1, we will apply this result to random Bernoulli shifts.

Remark 2 (Infinite alphabets). One can observe in the proof of Theorem 3, that stronger mixing assumptions for the stationary measure and the sample measures allow us to work with infinite alphabets. More precisely, if in Theorem 3, one replaces assumptions (I-a) and (II) by

(I') (exponential ϕ -mixing) the marginal measure μ satisfies

$$|\mu(A \cap \sigma^{-g-n} B) - \mu(A)\mu(B)| \leq \alpha(g)\mu(A);$$

and

(II') (fibered exponential ϕ -mixing) for \mathbb{P} -almost every $\omega \in \Omega$

$$|\mu_\omega(A \cap \sigma^{-g-n} B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)| \leq \alpha(g)\mu(A),$$

then the same conclusions are satisfied.

To deal with more general random subshifts (and in particular random Gibbs measures in Section 2.2) we will need a stronger mixing assumption on the base $(\Omega, \theta, \mathbb{P})$ (satisfied for example for Anosov diffeomorphisms [26]):

(III') (exponential 3-mixing) There exists a Banach space \mathcal{B} such that for all $\psi, \phi, \varphi \in \mathcal{B}$, for all $n \in \mathbb{N}^*$ and $m \in \mathbb{N}^*$, we have

$$\left| \int_{\Omega} \psi \cdot \phi \circ \theta^n \cdot \varphi \circ \theta^{n+m} d\mathbb{P} - \int_{\Omega} \psi d\mathbb{P} \int_{\Omega} \phi d\mathbb{P} \int_{\Omega} \varphi d\mathbb{P} \right| \leq \|\psi\|_{\mathcal{B}} \|\phi\|_{\mathcal{B}} \|\varphi\|_{\mathcal{B}} \rho(\min(n, m))$$

with $\rho(n) = \mathcal{O}(a^n)$ and $\|\cdot\|_{\mathcal{B}}$ is the norm in the Banach space \mathcal{B} . We are now able to replace assumption (IV) by a less restrictive assumption:

(IV') There exists $\xi \geq 0$ such that for every $n \in \mathbb{N}$ and every cylinder $C \in \mathcal{F}_0^n(X)$, the functions $\psi_1 : \omega \mapsto \mu_\omega(C)$ and $\psi_2 : \omega \mapsto \max_{C_n} \mu_\omega(C_n)$ (where the max is taken over all n -cylinders) belong to the Banach space \mathcal{B} and

$$\|\psi_1\|_{\mathcal{B}} \leq \xi^n \quad \text{and} \quad \|\psi_2\|_{\mathcal{B}} \leq \xi^n.$$

Moreover, if the base $(\Omega, \theta, \mathbb{P})$ satisfies exponential 4-mixing, it will allow us to weaken our mixing assumption for the marginal measure μ and use assumption (I):

(III'') (exponential 4-mixing) There exists a Banach space \mathcal{B} such that for all $\psi, \phi, \varphi, \nu \in \mathcal{B}$, for all $n \in \mathbb{N}^*$, $m \in \mathbb{N}^*$ and $l \in \mathbb{N}^*$, we have

$$\left| \int_{\Omega} \psi \cdot \phi \circ \theta^n \cdot \varphi \circ \theta^{n+m} \cdot \nu \circ \theta^{n+m+l} d\mathbb{P} - \int_{\Omega} \psi d\mathbb{P} \int_{\Omega} \phi d\mathbb{P} \int_{\Omega} \varphi d\mathbb{P} \int_{\Omega} \nu d\mathbb{P} \right| \leq \|\psi\|_{\mathcal{B}} \|\phi\|_{\mathcal{B}} \|\varphi\|_{\mathcal{B}} \|\nu\|_{\mathcal{B}} \rho(\min(n, m, l))$$

with $\rho(n) = \mathcal{O}(a^n)$ and $\|\cdot\|_{\mathcal{B}}$ is the norm in the Banach space \mathcal{B} .

In Section 2.2, we will check these assumptions for random Gibbs measures and will chose the Banach space \mathcal{B} to be the space of Hölder continuous functions.

With these assumptions, we obtain the same results as in Theorem 3:

Theorem 4. *If $0 < \underline{H}_2(\mu) \leq \overline{H}_2(\mu) < 2h_0$ and if*

- *hypothesis (I-a), (II), (III') and (IV') are satisfied, or*
- *hypothesis (I), (II), (III'') and (IV') are satisfied,*

then the conclusions of Theorem 3 hold.

We will now apply our results to random Bernoulli shifts and random Gibbs measures (these examples follow [34,35], where assumptions (I-a) and (II) where proved to obtain a quenched exponential distribution of hitting times).

2. Examples

2.1. Random Bernoulli shifts

Let $s \geq 1$ and (Ω, θ) be a subshift of finite type on the symbolic space $\{0, 1, \dots, s\}^{\mathbb{Z}}$ and let \mathbb{P} be a Gibbs measure from a Hölder potential.

Let $b \geq 1$ and make the shift $\{0, 1, \dots, b\}^{\mathbb{N}}$ a random subshift by putting on it the random Bernoulli measures constructed as follows. Let $W = (w_{ij})$ be a $s \times b$ stochastic matrix with entries in $(0, 1)$. Set $p_j(\omega) = w_{\omega_0, j}$. The random Bernoulli measure μ_ω is defined by

$$\mu_\omega([x_0 \dots x_n]) = p_{x_0}(\omega)p_{x_1}(\theta\omega) \dots p_{x_n}(\theta^n \omega).$$

First of all, hypothesis (IV) is satisfied since $\mu_\omega([x_0 \dots x_n])$ only depends on $\omega_0, \dots, \omega_n$. Since μ_ω are Bernoulli measures, one can observe that for all $m, n, A \in \mathcal{F}_0^m$ and $B \in \mathcal{F}_0^n$:

$$|\mu_\omega(A \cap \sigma^{-g-n} B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)| = 0$$

for every $g \geq 1$ and every $\omega \in \Omega$. Thus, property (I-a) is satisfied.

Moreover, it was proved in [34] that assumption (II) is satisfied. Since the Gibbs measure \mathbb{P} is exponentially ψ -mixing, it is exponentially ρ -mixing and (III) is satisfied. Thus, if $0 < \underline{H}_2(\mu) \leq 2h_0$ one can apply Theorem 2 and if besides that $\overline{H}_2(\mu) < 2h_0$ then one can apply Theorem 3.

For example, when the base is i.i.d., we can compute the Rényi entropy. Indeed

$$\begin{aligned} \mu([x_0 \dots x_n]) &= \int \mu_\omega([x_0 \dots x_n]) d\mathbb{P}(\omega) \\ &= \int p_{x_0}(\omega)p_{x_1}(\theta\omega) \dots p_{x_n}(\theta^n \omega) d\mathbb{P}(\omega) \\ &= \int p_{x_0}(\omega) d\mathbb{P}(\omega) \int p_{x_1}(\theta\omega) d\mathbb{P}(\omega) \dots \int p_{x_n}(\theta^n \omega) d\mathbb{P}(\omega) \\ &= \int p_{x_0}(\omega) d\mathbb{P}(\omega) \int p_{x_1}(\omega) d\mathbb{P}(\omega) \dots \int p_{x_n}(\omega) d\mathbb{P}(\omega). \end{aligned}$$

Thus,

$$\sum_{C_n} \mu(C_n)^2 = \left(\sum_{x_0} \left(\int p_{x_0}(\omega) d\mathbb{P} \right)^2 \right)^n$$

and

$$H_2(\mu) = -\log \left(\sum_{x_0} \left(\int p_{x_0}(\omega) d\mathbb{P} \right)^2 \right).$$

A similar computation gives us

$$h_0 = -\log\left(\int \max_{x_0} p_{x_0}(\omega) d\mathbb{P}\right).$$

So, if $H_2(\mu) < 2h_0$, we have for \mathbb{P} -almost every $\omega \in \Omega$,

$$\lim_{n \rightarrow \infty} \frac{M_n(x, y)}{\log n} = \frac{2}{-\log(\sum_{x_0} (\int p_{x_0}(\omega) d\mathbb{P})^2)}$$

for $\mu_\omega \otimes \mu_\omega$ -almost every $(x, y) \in X \times X$.

In this case, whether the condition $H_2(\mu) < 2h_0$ holds or not can be easily checked. For example, this condition will be satisfied if the letter with the maximum weight is always the same. Indeed, assuming that it exists $b_0 \in \{0, \dots, b\}$ such that $\max_{x_0} p_{x_0}(\omega) = p_{b_0}$ for every $\omega \in \Omega$, we observe that

$$\sum_{x_0} \left(\int p_{x_0}(\omega) d\mathbb{P}\right)^2 = p_{b_0}^2 + \sum_{x_0 \neq b_0} \left(\int p_{x_0}(\omega) d\mathbb{P}\right)^2 > p_{b_0}^2$$

and thus

$$H_2(\mu) < -\log(p_{b_0}^2) = 2h_0.$$

Also, the condition $H_2(\mu) < 2h_0$ will be satisfied if all the letters have a relatively close probability, i.e. if it exists a constant P such that $P < p_j(\omega) < P\sqrt{b+1}$ for every $j \in \{0, \dots, b\}$ and every $\omega \in \Omega$. Indeed, in this case, we have

$$\begin{aligned} \left(\int \max_{x_0} p_{x_0}(\omega) d\mathbb{P}\right)^2 &< \left(\int P\sqrt{b+1} d\mathbb{P}\right)^2 = P^2(b+1) = \sum_{x_0} \left(\int P d\mathbb{P}\right)^2 \\ &< \sum_{x_0} \left(\int p_{x_0}(\omega) d\mathbb{P}\right)^2 \end{aligned}$$

and thus $H_2(\mu) < 2h_0$. This could be applied to small perturbations of a uniform Bernoulli shift, i.e., $p_j(\omega) = \frac{1}{b+1} + \delta_j(\omega)$ with $\frac{1-\sqrt{b+1}}{(b+1)(1+\sqrt{b+1})} < \delta_j(\omega) < \frac{\sqrt{b+1}-1}{(b+1)(1+\sqrt{b+1})}$ for every $j \in \{0, \dots, b\}$ and every $\omega \in \Omega$ (one can easily check that in this case $P < p_j(\omega) < P\sqrt{b+1}$ with $P = \frac{2}{(b+1)(1+\sqrt{b+1})}$).

2.2. Random Gibbs measures

In this section we will give details on a family of shifts which satisfy our assumptions.

We will use the approach detailed in [38] which is concerned with shifts on \mathbb{N} , for example the full shift. We note that this extends a little beyond the full shift, to the so-called BIP setting.

We assume that $(\Omega, \mathbb{P}, \theta)$ is an invertible measure preserving system and let $X = \mathbb{N}^{\mathbb{N}}$ and let $\sigma : X \rightarrow X$ denote the shift. For $r \in (0, 1)$, let d_r be the usual symbolic metric on X , i.e., $d_r(x, y) = r^k$ where $x_i = y_i$ for $i = 0, \dots, k-1$, but $x_k \neq y_k$.

Assume that $\phi : X \times \Omega \rightarrow \mathbb{R}$ is a function which is almost surely Hölder continuous, which is to say, for

$$V_n^\omega(\phi) := \sup\{|\phi_\omega(x) - \phi_\omega(y)| : x_i = y_i, i = 0, \dots, n-1\},$$

there is some $r \in (0, 1)$ and $\kappa(\omega) \geq 0$ such that $\int \log \kappa d\mathbb{P} < \infty$ where $V_n^\omega(\phi) \leq \kappa(\omega)r^n$.

Define $S_n \phi_\omega(x) := \sum_{k=0}^{n-1} \phi_{\theta^k \omega} \circ \sigma^k(x)$. If x, y are in the same m -cylinder for $m \geq n$, then $|S_n \phi_\omega(x) - S_n \phi_\omega(y)| \leq r^{m-n} \sum_{k=0}^{n-1} r^k \kappa(\theta^{n-k} \omega)$. As in the proof of [17, Lemma 7.2], the assumption on the integrability of $\log \kappa$ implies that the above limit is finite a.s., say $\sum_{k=0}^{n-1} r^k \kappa(\theta^{n-k} \omega) \leq c_\omega$. However, it is also pointed out in [38] that if κ is integrable, then we have an a.s. uniform upper bound, say C_ϕ on $\sum_{k=0}^{n-1} r^k \kappa(\theta^{n-k} \omega)$. Given a Hölder function ψ , then we define

$$D_\omega(\psi) := \sup_{x, y \in X_\omega} \left\{ \frac{|\psi(x) - \psi(y)|}{d_r(x, y)} \right\}.$$

Now we define the random Ruelle operator by

$$\mathcal{L}_\omega \psi(x) = \sum_{\sigma y=x} e^{\phi_\omega(y)} \psi(y),$$

where $\psi : X' \rightarrow [0, \infty]$ where $X' \subset X$ is such that \mathcal{L}_ω is well-defined. As in [16,38], it can be shown that there exists some constant λ_ω and some measurable function ρ_ω which is uniformly bounded from below, such that $\mathcal{L}_\omega \rho_\omega = \lambda_\omega \rho_{\theta\omega}$ a.s. and such that $\log \rho$ satisfies the same smoothness properties as ϕ , i.e. we have the same κ and r in the variation. This allows us to replace ϕ with

$$\varphi_\omega(x) := \phi_\omega(x) + \log \rho_\omega - \log \rho_{\theta\omega}(\sigma x) - \log \lambda_\omega.$$

Letting \mathcal{L}_ω denote the corresponding transfer operator, one consequence of this is that $\mathcal{L}_\omega 1 = 1$. Note also that random equilibrium states for ϕ and φ coincide.

Now we have the property that

$$\int \mathcal{L}_\omega^n(\psi) \cdot \gamma \, d\mu_\omega = \int \psi \cdot \gamma \circ \sigma^n \, d\mu_{\theta^{-n}\omega} \tag{2}$$

for appropriate observables ψ, γ .

We will make the following almost sure assumptions on our system (which are satisfied for subshifts of finite type with Hölder potentials):

1. $\int \kappa \, d\mathbb{P} < \infty$, so $\sum_{k=0}^\infty r^k \kappa(\theta^{n-k}\omega)$ is a.s. uniformly bounded, independently of ω .
2. There exists a measure μ_ω where $\mathcal{L}_\omega^* \mu_\omega = \mu_{\theta^{-1}\omega}$, i.e., (2) holds for L^1 observables.
3. Big images: there exists some $C_{BIP} > 0$ such that for any n -cylinder U and $\omega \in \Omega$, $\inf \mu_{\theta^n \omega}(\sigma^n U) > C_{BIP}$.
4. There exist $C > 0$, and $g(n) \rightarrow 0$ as $n \rightarrow \infty$ such that

$$\left\| \mathcal{L}_\omega^n(\psi) - \int \psi \, d\mu_\omega \right\|_\infty \leq Cg(n)D_\omega(\psi).$$

Under these conditions, it was proved in [35, Proposition 6.1] that the sample measures satisfy (II).

When $\theta : \Omega \rightarrow \Omega$ is a subshift of finite type on a finite alphabet, with a Gibbs measure for a Hölder potential, it is known that assumption (III'') is satisfied with \mathcal{B} being the space of Hölder continuous functions [26,36].

For $\alpha > 0$, let the norm $\|\cdot\|_\alpha$ be defined by $\|\cdot\|_\alpha = |\cdot|_\alpha + |\cdot|_\infty$ where $|f|_\alpha = \sup\{\frac{V_n(f)}{\alpha^n} : n \geq 0\}$.

It was also proved in [35, Lemma 6.2] that for any $\beta \in (0, 1]$, there exist $\alpha \in (0, 1]$ and $C_\beta > 0$, such that for every cylinder C in $\mathcal{F}_0^n(X)$, the map $\psi_1 : \omega \mapsto \mu_\omega(C)$ is α -Hölder and $\|\psi_1\|_\alpha \leq C_\beta r^{-\beta n}$. Thus, $\|\psi_1\|_\alpha \leq \xi^n$ for some $\xi \geq 0$. Moreover, since for every real-valued functions f, g we have $|\max f(x) - \max g(x)| \leq \max |f(x) - g(x)|$, we obtain that the map $\psi_2 : \omega \mapsto \max_{C_n} \mu_\omega(C_n)$ is α -Hölder and $\|\psi_2\|_\alpha \leq \xi^n$. Thus, (IV') is satisfied.

Assumption (I-a) has been proved in [35, Section 6.2]. However, our proof contains a mistake since both terms in the right-hand side of the first equation in page 149 should be with the Hölder norm. In fact, we will prove that the sample measures satisfy (I). One can observe that to obtain Theorem 2.2 in [35], (I-a) was only used in equation (4.3) and could be substituted by (I).

Following the proof of [30, Proposition 2.4], we fix our set $A \in \mathcal{F}_0^n$ and take both w and v to be $w(\omega) = w_A(\omega) = \mu_\omega(A) - \mu(A)$ (this normalisation by $\mu(A)$ simplifies the calculations). Note that $|w|_\infty \leq 1$. Let $\gamma \in (0, \alpha)$. For $k \in \mathbb{N}$, we approximate v, w by v_k, w_k , depending only on coordinates $x_{-k}, \dots, x_0, \dots, x_k$ such that $|v - v_k| \leq \gamma^k |v|_\gamma$ and $|w - w_k| \leq \gamma^k |w|_\gamma$. So that proof yields that for $n \geq 2k$,

$$\begin{aligned} \left| \int v \circ \theta^{n+g} w \, d\mathbb{P} \right| &\leq \left| \int (v - v_k) \circ \theta^{n+g} w \, d\mathbb{P} \right| + \left| \int v_k \circ \theta^{n+g} (w - w_k) \, d\mathbb{P} \right| + \left| \int v_k \circ \theta^{n+g} w_k \, d\mathbb{P} \right| \\ &\leq \gamma^k |v|_\gamma |w|_\infty + \gamma^k |w|_\gamma |v|_\infty + K' \rho^{n+g-2k} |v|_\infty |w_k|_\infty \\ &\leq 2\gamma^k C_\beta r^{-\beta n} + K' \rho^{n+g-2k}. \end{aligned}$$

So taking $k = \lfloor \frac{n+g}{3} \rfloor$, if we choose $\gamma \in (0, \alpha)$ so that $\gamma^{\frac{1}{3}} r^{-\beta} < 1$, we obtain

$$\left| \int v \circ \theta^{n+g} w \, d\mathbb{P} \right| \leq 2C_\beta \gamma^{\frac{g}{3}} + K' \rho^{\frac{n+g}{3}}. \tag{3}$$

Moreover, we can observe that by (II)

$$\begin{aligned}
 |\mu(A \cap \sigma^{-g-n} A) - \mu(A)^2| &= \left| \int \mu_\omega(A \cap \sigma^{-g-n} A) d\mathbb{P} - \mu(A)^2 \right| \\
 &\leq \left| \int \mu_\omega(A) \mu_{\theta^{g+n}\omega}(A) d\mathbb{P} - \mu(A)^2 \right| + \alpha(g) \\
 &= \alpha(g) + \left| \int v \circ \sigma^{n+g} w d\mathbb{P} \right|.
 \end{aligned} \tag{4}$$

Thus, by (3) and (4), (I) is verified.

Finally, we showed that if the fiber maps satisfy conditions 1–4 and the base transformation is a subshift of finite type on a finite alphabet with a Gibbs measure for some Hölder potential, then assumptions (I), (II), (III'') and (IV') are satisfied. Thus, if $0 < \underline{H}_2(\mu) \leq \overline{H}_2(\mu) < 2h_0$, one can apply Theorem 4.

3. Proofs

In this section, we will prove our theorems. Both proofs follow the line of [9] but diverge at some point since the samples measures are not invariant but satisfy (1).

Proof of Theorem 2. For simplicity we assume $\alpha(g) = e^{-g}$. Let $\varepsilon > 0$ and define

$$k_n = \left\lceil \frac{2 \log n + d \log \log n}{\underline{H}_2(\mu) - \varepsilon} \right\rceil,$$

where d is a constant to be chosen later.

Let us also denote

$$A_{i,j}(y) = \sigma^{-i} [C_{k_n}(\sigma^j y)]$$

and

$$S_n(x, y) = \sum_{i,j=0,\dots,n-1} \mathbb{1}_{A_{i,j}(y)}(x).$$

Let $\omega \in \Omega$ such that (1) is satisfied. Using Markov's inequality we obtain

$$\mu_\omega \otimes \mu_\omega(\{(x, y) : M_n(x, y) \geq k_n\}) = \mu_\omega \otimes \mu_\omega(\{(x, y) : S_n(x, y) \geq 1\}) \leq \mathbb{E}_\omega(S_n). \tag{5}$$

Moreover, the invariance formula (1) of the sample measures gives us

$$\begin{aligned}
 \mathbb{E}_\omega(S_n) &= \int \int \sum_{i,j=0,\dots,n-1} \mathbb{1}_{C_{k_n}(\sigma^j y)}(\sigma^i x) d\mu_\omega(x) d\mu_\omega(y) \\
 &= \sum_{i,j=0,\dots,n-1} \int \int \mathbb{1}_{C_{k_n}(\sigma^j y)}(x) d\mu_{\theta^i \omega}(x) d\mu_\omega(y) \\
 &= \sum_{i,j=0,\dots,n-1} \int \mu_{\theta^i \omega}(C_{k_n}(\sigma^j y)) d\mu_\omega(y) \\
 &= \sum_{i,j=0,\dots,n-1} \int \mu_{\theta^i \omega}(C_{k_n}(y)) d\mu_{\theta^j \omega}(y) \\
 &= \sum_{i,j=0,\dots,n-1} \sum_{C_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}).
 \end{aligned}$$

One can notice that, since the sample measures are not invariant, we cannot estimate the previous sum directly as in the deterministic case [9]. Thus, this is where our proof will differ and where we will use the mixing assumptions which where not necessary in the deterministic proof. First of all, using Markov’s inequality, we observe that

$$\mathbb{P}\left(\omega : \mathbb{E}_\omega(S_n) \geq \frac{1}{\log n}\right) \leq \log n \cdot \int \sum_{i,j=0,\dots,n-1} \sum_{C_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P}(\omega)$$

To study the behaviour of the integral on the right hand side of the previous inequality, we will divide the sum in two terms, when i and j are far from one another and when they are not. Let us define $m = \gamma \log n$ where $\gamma > 0$ will be chosen later.

When i and j are close from one another, we have, using that $\mu_{\theta^j \omega}$ is a probability measure and the invariance of \mathbb{P}

$$\begin{aligned} \sum_{|i-j| \leq m} \sum_{C_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P}(\omega) &\leq \sum_{|i-j| \leq m} \int \max_{C_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \cdot \sum_{C_{k_n}} \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P}(\omega) \\ &= \sum_{|i-j| \leq m} \int \max_{C_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) d\mathbb{P}(\omega) \\ &\leq 2mn \int \max_{C_{k_n}} \mu_\omega(C_{k_n}) d\mathbb{P}(\omega). \end{aligned}$$

When i and j are far from one another, we can use the mixing assumptions (I) and (II) to obtain

$$\begin{aligned} \sum_{|i-j| > m} \sum_{C_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P}(\omega) &\leq 2 \sum_{j > i+m} \sum_{C_{k_n}} \left(\alpha(m - k_n) + \int \mu_{\theta^i \omega}(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) d\mathbb{P}(\omega) \right) \\ &\leq 2n^2 N^{k_n} \alpha(m - k_n) + 2 \sum_{j > i+m} \sum_{C_{k_n}} \mu(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) \\ &\leq 2n^2 N^{k_n} \alpha(m - k_n) + n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2. \end{aligned}$$

Thus, we obtain, for n large enough,

$$\begin{aligned} \mathbb{P}\left(\omega : \mathbb{E}_\omega(S_n) \geq \frac{1}{\log n}\right) &\leq \log n \left(2mn \int \max_{C_{k_n}} \mu_\omega(C_{k_n}) d\mathbb{P}(\omega) + 2n^2 N^{k_n} \alpha(m - k_n) + n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2 \right) \\ &\leq \log n \left(2mn e^{-k_n(h_0 - \varepsilon/2)} + 2n^2 N^{k_n} e^{-(m - k_n)} + n^2 e^{-k_n(H_2(\mu) - \varepsilon)} \right), \end{aligned}$$

where the last inequality came from the definition of h_0 and $H_2(\mu)$.

Then, choosing $d > 0$ large enough and $\gamma > 0$ large enough, we have, by definition of k_n and since $H_2(\mu) \leq 2h_0$, that

$$\mathbb{P}\left(\omega : \mathbb{E}_\omega(S_n) \geq \frac{1}{\log n}\right) \leq \mathcal{O}\left(\frac{1}{\log n}\right).$$

Choosing a subsequence $\{n_\kappa\}_{\kappa \in \mathbb{N}}$ such that $n_\kappa = \lceil e^{\kappa^2} \rceil$ we have that

$$\mathbb{P}\left(\omega : \mathbb{E}_\omega(S_{n_\kappa}) \geq \frac{1}{\log n_\kappa}\right) \leq \frac{1}{\kappa^2}.$$

Since the last quantity is summable in κ , the Borel–Cantelli lemma gives that for \mathbb{P} -almost every $\omega \in \Omega$, if κ is large enough then

$$\mathbb{E}_\omega(S_{n_\kappa}) \leq \frac{1}{\log n_\kappa} = \frac{1}{\kappa^2}.$$

Thus, this inequality together with (5) gives us that for \mathbb{P} -almost every $\omega \in \Omega$, if κ is large enough then

$$\mu_\omega \otimes \mu_\omega(\{(x, y) : M_{n_\kappa}(x, y) \geq k_{n_\kappa}\}) \leq \frac{1}{\kappa^2}.$$

As previously, since the last quantity is summable in κ , the Borel–Cantelli lemma gives that for $\mu_\omega \otimes \mu_\omega$ -almost every (x, y) , if κ is large enough then

$$M_{n_\kappa}(x, y) < k_{n_\kappa}$$

and then

$$\frac{M_{n_\kappa}(x, y)}{\log n_\kappa} \leq \frac{1}{\underline{H}_2(\mu) - \varepsilon} \left(2 + \frac{\log \log n_\kappa}{\log n_\kappa} \right).$$

Finally, taking the limit superior in the previous equation and observing that $(n_\kappa)_\kappa$ is increasing, $(M_n)_n$ is increasing and $\lim_{\kappa \rightarrow +\infty} \frac{\log n_\kappa}{\log n_{\kappa+1}} = 1$, we have for $\mu_\omega \otimes \mu_\omega$ -almost every (x, y)

$$\overline{\lim}_{n \rightarrow +\infty} \frac{M_n(x, y)}{\log n} = \overline{\lim}_{\kappa \rightarrow +\infty} \frac{\log M_{n_\kappa}(x, y)}{\log n_\kappa} \leq \frac{2}{\underline{H}_2(\mu) - \varepsilon}.$$

Then the theorem is proved since ε can be chosen arbitrarily small. □

Proof of Theorem 3 and Theorem 4. For $\varepsilon > 0$, let us define

$$k_n = \left\lfloor \frac{1}{\overline{H}_2(\mu) + \varepsilon} (2 \log n + d \log \log n) \right\rfloor,$$

where d is a constant that we will choose later.

Let $\omega \in \Omega$ such that (1) is satisfied. As in the proof of Theorem 2, we have

$$\mathbb{E}_\omega(S_n) = \sum_{i, j=0, \dots, n-1} \sum_{C_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}).$$

Following the lines of the proof of Theorem 7 in [9], we have, by Chebyshev’s inequality,

$$\mu_\omega \otimes \mu_\omega((x, y) : M_n(x, y) < k_n) = \mu_\omega \otimes \mu_\omega((x, y) : S_n(x, y) = 0) \leq \frac{\text{var}_\omega(S_n)}{\mathbb{E}_\omega(S_n)^2}.$$

Thus, we need to control the variance of S_n . First of all, we observe that

$$\text{var}_\omega(S_n) = \sum_{0 \leq i, i', j, j' \leq n-1} \iint \mathbb{1}_{C_{k_n}(\sigma^j y)}(\sigma^i x) \mathbb{1}_{C_{k_n}(\sigma^{j'} y)}(\sigma^{i'} x) d\mu_\omega(x) d\mu_\omega(y) - \mathbb{E}_\omega(S_n)^2.$$

We will estimate the variance dividing the sum of $\text{var}(S_n)$ into 4 terms. Let $g = \log(n^\beta)$ where β is a constant that we will choose later.

For $i' - i > g + k_n$, we use the invariance formula (1) and the mixing assumption (II) to obtain:

$$\begin{aligned} & \iint \mathbb{1}_{C_{k_n}(\sigma^j y)}(\sigma^i x) \mathbb{1}_{C_{k_n}(\sigma^{j'} y)}(\sigma^{i'} x) d\mu_\omega(x) d\mu_\omega(y) \\ &= \iint \mathbb{1}_{C_{k_n}(\sigma^j y)}(x) \mathbb{1}_{C_{k_n}(\sigma^{j'} y)}(\sigma^{i'-i} x) d\mu_{\theta^i \omega}(x) d\mu_\omega(y) \\ &\leq \alpha(g) + \int \mu_{\theta^i \omega}(C_{k_n}(\sigma^j y)) \mu_{\theta^{i'} \omega}(C_{k_n}(\sigma^{j'} y)) d\mu_\omega(y). \end{aligned}$$

If, moreover, $j' - j > g + k_n$, using again the mixing assumption (II), we have

$$\int \mu_{\theta^i \omega}(C_{k_n}(\sigma^j y)) \mu_{\theta^{i'} \omega}(C_{k_n}(\sigma^{j'} y)) d\mu_\omega(y)$$

$$\begin{aligned}
 &= \int \mu_{\theta^i \omega}(C_{k_n}(y)) \mu_{\theta^{i'} \omega}(C_{k_n}(\sigma^{j'-j} y)) d\mu_{\theta^j \omega}(y) \\
 &= \sum_{C_{k_n}, C'_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^j \omega}(C_{k_n} \cap \sigma^{-(j'-j)} C'_{k_n}) \\
 &\leq \alpha(g) + \sum_{C_{k_n}, C'_{k_n}} \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}).
 \end{aligned}$$

However, if $j' - j \leq g + k_n$, we obtain:

$$\int \mu_{\theta^i \omega}(C_{k_n}(\sigma^j y)) \mu_{\theta^{i'} \omega}(C_{k_n}(\sigma^{j'} y)) d\mu_{\omega}(y) \leq \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) \int \mu_{\theta^i \omega}(C_{k_n}(y)) d\mu_{\theta^j \omega}(y).$$

By symmetry, the case where $i' - i \leq g + k_n$ and $j' - j > g + k_n$ will be treated as the previous one.

Finally, when $|i - i'| \leq g + k_n$ and $|j - j'| \leq g + k_n$, we have:

$$\begin{aligned}
 \iint \mathbb{1}_{C_{k_n}(\sigma^j y)}(\sigma^i x) \mathbb{1}_{C_{k_n}(\sigma^{j'} y)}(\sigma^{i'} x) d\mu_{\omega}(x) d\mu_{\omega}(y) &\leq \iint \mathbb{1}_{C_{k_n}(\sigma^j y)}(\sigma^i x) d\mu_{\omega}(x) d\mu_{\omega}(y) \\
 &= \int \mu_{\theta^i \omega}(C_{k_n}(y)) d\mu_{\theta^j \omega}(y).
 \end{aligned}$$

Then, one can gather these estimates to obtain

$$\begin{aligned}
 &\mu_{\omega} \otimes \mu_{\omega}((x, y) : M_n(x, y) < k_n) \\
 &\leq \frac{2n^4 \alpha(g) + 2(g + k_n) \sum_{0 \leq i, j, i' \leq n-1} \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) \int \mu_{\theta^i \omega}(C_{k_n}(y)) d\mu_{\theta^j \omega}(y)}{\mathbb{E}_{\omega}(S_n)^2} \\
 &\quad + \frac{(g + k_n)^2 \sum_{0 \leq i, j \leq n-1} \int \mu_{\theta^i \omega}(C_{k_n}(y)) d\mu_{\theta^j \omega}(y)}{\mathbb{E}_{\omega}(S_n)^2} \\
 &= \frac{2n^4 \alpha(g)}{\mathbb{E}_{\omega}(S_n)^2} + \frac{2(g + k_n) \sum_{0 \leq i' \leq n-1} \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) + (g + k_n)^2}{\mathbb{E}_{\omega}(S_n)}. \tag{6}
 \end{aligned}$$

This is where the proof diverge completely from the deterministic case. Indeed, as in the proof of Theorem 2, we cannot treat directly the previous estimate (which was possible in the deterministic case) and an extra care is needed. To deal with the term with the maximum, we use Markov's inequality to obtain

$$\begin{aligned}
 \mathbb{P}\left(\sum_{0 \leq i' \leq n-1} \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) \geq 1\right) &\leq \int_{\Omega} \sum_{0 \leq i' \leq n-1} \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) d\mathbb{P}(\omega) \\
 &= n \int_{\Omega} \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P}(\omega) \\
 &\leq n e^{k_n(h_0 - \varepsilon)}. \tag{7}
 \end{aligned}$$

Since $\overline{H}_2(\mu) < 2h_0$, one can choose ε small enough such that $n e^{k_n(h_0 - \varepsilon)} \leq n^{-\varepsilon}$ for every n large enough.

To deal with the expectation in the denominator in (6), we will need the following lemma (which proof can be found after the proof of the theorem).

Lemma 5. *Let $\frac{3}{4} < \delta < 1$. Under the assumptions of Theorem 3 or Theorem 4, we have*

$$\mathbb{P}\left(\left|\mathbb{E}_{\omega}(S_n) - n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right| \geq \left(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right)^{\delta}\right) = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Thus, using this lemma with (7), we have

$$\mathbb{P}\left(\left|\mathbb{E}_{\omega}(S_n) - n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right| \geq \left(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right)^{\delta} \cup \sum_{0 \leq i' \leq n-1} \max_{C_{k_n}} \mu_{\theta^{i'} \omega}(C_{k_n}) \geq 1\right) = \mathcal{O}\left(\frac{1}{\log n}\right).$$

Choosing a subsequence $\{n_\kappa\}_{\kappa \in \mathbb{N}}$ such that $n_\kappa = \lceil e^{\kappa^2} \rceil$, the Borel–Cantelli lemma gives that for \mathbb{P} -almost every $\omega \in \Omega$, if κ is large enough then

$$\sum_{0 \leq i' \leq n_\kappa - 1} \max_{C_{kn_\kappa}} \mu_{\theta^{i'} \omega}(C_{kn_\kappa}) \leq 1 \tag{8}$$

and

$$\left| \mathbb{E}_\omega(S_{n_\kappa}) - n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2 \right| \leq \left(n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2 \right)^\delta.$$

Thus, if κ is large enough

$$\mathbb{E}_\omega(S_{n_\kappa}) \geq n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2 - \left(n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2 \right)^\delta \geq \left(n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2 \right)^\delta. \tag{9}$$

Thus, (6) together with (8) and (9) gives us that for \mathbb{P} -almost every $\omega \in \Omega$, if κ is large enough then

$$\begin{aligned} \mu_\omega \otimes \mu_\omega((x, y) : M_{n_\kappa}(x, y) < k_{n_\kappa}) &\leq \frac{2n_\kappa^4 \alpha(g)}{(n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2)^{2\delta}} + \frac{2(g + k_{n_\kappa}) + (g + k_{n_\kappa})^2}{(n_\kappa^2 \sum_{C_{kn_\kappa}} \mu(C_{kn_\kappa})^2)^\delta} \\ &\leq \frac{2n_\kappa^4 \alpha(g)}{(\log n_\kappa)^{-2b\delta}} + \frac{2(g + k_{n_\kappa}) + (g + k_{n_\kappa})^2}{(\log n_\kappa)^{-b\delta}}, \end{aligned}$$

where the last inequality came from the definition of $\overline{H}_2(\mu)$ and our choice of k_n . Finally, choosing β large enough in the definition of g and choosing $d < 0$ small enough, we obtain that if κ is large enough

$$\mu_\omega \otimes \mu_\omega((x, y) : M_{n_\kappa}(x, y) < k_{n_\kappa}) \leq \frac{1}{\log n_\kappa} = \frac{1}{\kappa^2}.$$

Since the last quantity is summable in κ , the Borel–Cantelli lemma gives that for $\mu_\omega \otimes \mu_\omega$ -almost every (x, y) , if κ is large enough then

$$M_{n_\kappa}(x, y) \geq k_{n_\kappa}$$

and then

$$\frac{M_{n_\kappa}(x, y)}{\log n_\kappa} \geq \frac{1}{\overline{H}_2(\mu) + \varepsilon} \left(2 + b \frac{\log \log n_\kappa}{\log n_\kappa} \right).$$

Finally, using the same arguments as in the proof of Theorem 2, we have for \mathbb{P} -almost every ω

$$\lim_{n \rightarrow +\infty} \frac{M_n(x, y)}{\log n} = \lim_{\kappa \rightarrow +\infty} \frac{\log M_{n_\kappa}(x, y)}{\log n_\kappa} \geq \frac{2}{\overline{H}_2(\mu) + \varepsilon}$$

for $\mu_\omega \otimes \mu_\omega$ -almost every (x, y) .

Then the theorems are proved since ε can be chosen arbitrarily small. □

Proof of Lemma 5. As in the previous proof, we take $k_n = \lfloor \frac{1}{\overline{H}_2(\mu) + \varepsilon} (2 \log n + d \log \log n) \rfloor$ and $g = \log(n^\beta)$ where $d < 0$ and $\beta > 0$ are constants to be chosen later.

First of all, we use Markov’s inequality

$$\begin{aligned} \mathbb{P} \left(\left| \mathbb{E}_\omega(S_n) - n^2 \sum_{C_{kn}} \mu(C_{kn})^2 \right| \geq \left(n^2 \sum_{C_{kn}} \mu(C_{kn})^2 \right)^\delta \right) \\ = \mathbb{P} \left(\left| \mathbb{E}_\omega(S_n) - n^2 \sum_{C_{kn}} \mu(C_{kn})^2 \right|^2 \geq \left(n^2 \sum_{C_{kn}} \mu(C_{kn})^2 \right)^{2\delta} \right) \end{aligned}$$

$$\begin{aligned} &\leq \frac{\int |\mathbb{E}_\omega(S_n) - n^2 \sum_{C_{k_n}} \mu(C_{k_n})|^2 d\mathbb{P}(\omega)}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} \\ &= \frac{\int (\mathbb{E}_\omega(S_n)^2 + (n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^2 - 2\mathbb{E}_\omega(S_n)n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2) d\mathbb{P}(\omega)}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}}. \end{aligned} \tag{10}$$

Firstly, we will treat the last term on the previous numerator, using the mixing assumptions (I) and (II)

$$\begin{aligned} \int \mathbb{E}_\omega(S_n) d\mathbb{P} &= \sum_{C_{k_n}} \sum_{i,j=0,\dots,n-1} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P} \\ &\geq \sum_{C_{k_n}} \sum_{|i-j| \geq g+k_n} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) d\mathbb{P} \\ &\geq 2 \sum_{C_{k_n}} \sum_{j \geq i+g+k_n} \left(\int \mu_{\theta^i \omega}(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) d\mathbb{P} - \alpha(g) \right) \\ &\geq \sum_{C_{k_n}} \sum_{|i-j| \geq g+k_n} (\mu(C_{k_n})^2 - 2\alpha(g)) \\ &\geq n(n - (g + k_n)) \left(\sum_{C_{k_n}} \mu(C_{k_n})^2 - 2\alpha(g)N^{k_n} \right) \\ &\geq n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2 - n(g + k_n) \sum_{C_{k_n}} \mu(C_{k_n})^2 - 2n^2\alpha(g)N^{k_n}. \end{aligned} \tag{11}$$

To get an estimate on (10), we need to study the term $\int \mathbb{E}_\omega(S_n)^2 d\mathbb{P}$. One can observe that

$$\int \mathbb{E}_\omega(S_n)^2 d\mathbb{P} = \sum_{C_{k_n}, C'_{k_n}} \sum_{i,j,i',j'=0,\dots,n-1} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P}.$$

We will separate the study of this integral depending on the relative distance and position between i, j, i' and j' and consider 5 different cases.

Case 1: i, j, i' and j' are all far from one another, i.e. at least at a distance greater that $g + k_n$. We will assume that $i < j < i' < j'$ (when the relative position is different, everything can be done identically because of the symmetry) and that $j - i > g + k_n, i' - j > g + k_n, j' - i' > g + k_n$. Using the mixing assumptions (I-a) and (II) (a similar estimate is obtained when (III'') is satisfied) we obtain

$$\begin{aligned} &\int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \\ &\leq \int (\mu_{\theta^i \omega}(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) + \alpha(g)) (\mu_{\theta^{i'} \omega}(C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n}) + \alpha(g)) d\mathbb{P} \\ &\leq \alpha(g)^2 + \alpha(g) \mu(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) + \alpha(g) \mu(C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n}) \\ &\quad + \int \mu_{\theta^i \omega}(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n}) d\mathbb{P} \\ &\leq \alpha(g)^2 + 2\alpha(g) + \int (\alpha(g) + \mu_{\theta^i \omega}(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n} \cap \sigma^{-(i'-i)} (C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n}))) d\mathbb{P} \\ &\leq \alpha(g)^2 + 3\alpha(g) + \mu(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n} \cap \sigma^{-(i'-i)} (C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n})) \\ &\leq \alpha(g)^2 + 4\alpha(g) + \mu(C_{k_n} \cap \sigma^{-(j-i)} C_{k_n}) \mu(C'_{k_n} \cap \sigma^{-(j'-i')} C'_{k_n}) \\ &\leq 2\alpha(g)^2 + 6\alpha(g) + \mu(C_{k_n})^2 \mu(C'_{k_n})^2. \end{aligned} \tag{12}$$

Case 2: only two indices are close. We will assume that $i \leq j \leq i' \leq j'$ and that $j - i > g + k_n$, $j' - i' > g + k_n$, $j' - i' \leq g + k_n$. Since the cylinders form a partition and that the sample measures are probability measures, we have

$$\begin{aligned} & \sum_{C_{k_n}, C'_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \\ & \leq \sum_{C_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \max_{C'_{k_n}} \mu_{\theta^{i'} \omega}(C'_{k_n}) d\mathbb{P}. \end{aligned} \tag{13}$$

When the indices are in a different position and/or the two close indices are not j' and i' , the same idea can be used. However, one need to choose carefully with which index to take the maximum so that one index disappears with one sum and we obtain a similar term as (13) where the 3 remaining indices are far from each other. Then, we use the mixing assumptions (III') and (IV') (a similar estimate is obtained when (III) and (IV) are satisfied) to get

$$\begin{aligned} & \sum_{C_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \max_{C'_{k_n}} \mu_{\theta^{i'} \omega}(C'_{k_n}) d\mathbb{P} \\ & \leq \sum_{C_{k_n}} \left(\int \mu_{\omega}(C_{k_n}) d\mathbb{P} \int \mu_{\omega}(C_{k_n}) d\mathbb{P} \int \max_{C'_{k_n}} \mu_{\omega}(C'_{k_n}) d\mathbb{P} + \rho(g) \xi^{3k_n} \right) \\ & = \rho(g) \xi^{3k_n} N^{k_n} + \int \max_{C'_{k_n}} \mu_{\omega}(C'_{k_n}) d\mathbb{P} \sum_{C_{k_n}} \mu(C_{k_n})^2. \end{aligned} \tag{14}$$

Case 3: three indices are close and one is far from them. We will assume that $i \leq j \leq i' \leq j'$ and that $j - i \leq g + k_n$, $j' - i' \leq g + k_n$, $j' - i' > g + k_n$. Since $\mu_{\theta^j \omega}(C_{k_n}) \leq 1$ and $\mu_{\theta^i \omega}$ is a probability measure we have

$$\begin{aligned} & \sum_{C_{k_n}, C'_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \\ & \leq \sum_{C'_{k_n}} \int \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P}. \end{aligned}$$

When the indices are in a different position, one can use the same idea so that we stay with two indices which are far from each other and measure the same cylinder. Thus we can use the mixing assumptions (I) an (II), to obtain

$$\begin{aligned} \sum_{C'_{k_n}} \int \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} & \leq \sum_{C'_{k_n}} \int (\mu_{\theta^{i'} \omega}(C'_{k_n} \cap \sigma^{(j'-i')} C'_{k_n}) + \alpha(g)) d\mathbb{P} \\ & \leq \sum_{C'_{k_n}} (\mu(C'_{k_n} \cap \sigma^{(j'-i')} C'_{k_n}) + \alpha(g)) \\ & \leq \alpha(g) N^{k_n} + \sum_{C'_{k_n}} \mu(C'_{k_n})^2. \end{aligned} \tag{15}$$

Case 4: two indices are close and both are far from the two other indices which are close from one another. We will assume that $i \leq j \leq i' \leq j'$ and that $j - i \leq g + k_n$, $j' - i' > g + k_n$, $j' - i' \leq g + k_n$. Since the sample measures are probability measures, we obtain

$$\begin{aligned} & \sum_{C_{k_n}, C'_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \\ & \leq \int \max_{C_{k_n}} \mu_{\theta^j \omega}(C_{k_n}) \max_{C'_{k_n}} \mu_{\theta^{i'} \omega}(C'_{k_n}) d\mathbb{P}. \end{aligned} \tag{16}$$

For the other relative positions, we can observe that

- if the measures with the two indices that are far from each other measure different cylinders, we obtain an estimate similar to (16);
- if the measures with the two indices that are far from each other measure the same cylinder, the case can be treated as case 3.

Then, using the mixing assumptions (III') and (IV') (a similar estimate is obtained when (III) and (IV) are satisfied), we have

$$\int \max_{C_{k_n}} \mu_{\theta^j \omega}(C_{k_n}) \max_{C'_{k_n}} \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \leq \rho(g) \xi^{2k_n} + \left(\int \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P} \right)^2. \tag{17}$$

Case 5: all the indices are close. We will assume that $i \leq j \leq i' \leq j'$ and that $j - i \leq g + k_n$, $j' - i' \leq g + k_n$, $j' - i' \leq g + k_n$. In this case, the relative position is irrelevant. Since the sample measures are probability measures, we obtain

$$\sum_{C_{k_n}, C'_{k_n}} \int \mu_{\theta^i \omega}(C_{k_n}) \mu_{\theta^j \omega}(C_{k_n}) \mu_{\theta^{i'} \omega}(C'_{k_n}) \mu_{\theta^{j'} \omega}(C'_{k_n}) d\mathbb{P} \leq \int \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P}. \tag{18}$$

Finally, (10) together with (11), (12), (14), (15), (17) and (18) gives us that there exists a constant c_1 such that

$$\begin{aligned} & \mathbb{P} \left(\left| \mathbb{E}_{\omega}(S_n) - n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2 \right| \geq \left(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2 \right)^{\delta} \right) \\ & \leq \frac{c_1}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} \left(n^3 (g + k_n) \int \max_{C'_{k_n}} \mu_{\omega}(C'_{k_n}) d\mathbb{P} \sum_{C_{k_n}} \mu(C_{k_n})^2 + n^2 (g + k_n)^2 \sum_{C'_{k_n}} \mu(C'_{k_n})^2 \right. \\ & \quad + n^2 (g + k_n)^2 \left(\int \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P} \right)^2 + n (g + k_n)^3 \int \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P} \\ & \quad \left. + n^3 (g + k_n) \left(\sum_{C_{k_n}} \mu(C_{k_n})^2 \right)^2 + \Gamma(g) \right), \end{aligned} \tag{19}$$

where

$$\begin{aligned} \Gamma(g) &= n^4 (\alpha(g)^2 + \alpha(g)) + n^3 (g + k_n) \rho(g) \xi^{3k_n} N^{k_n} + n^2 (g + k_n)^2 \alpha(g) N^{k_n} \\ & \quad + n^2 (g + k_n)^2 \rho(g) \xi^{2k_n} + n^4 \alpha(g) N^{k_n} \sum_{C_{k_n}} \mu(C_{k_n})^2. \end{aligned}$$

We recall that $k_n = \frac{1}{H_2(\mu) + \varepsilon} (2 \log n + d \log \log n)$ and that $g = \log(n^\beta)$. Thus, as in (7), we have

$$n \int_{\Omega} \max_{C_{k_n}} \mu_{\omega}(C_{k_n}) d\mathbb{P}(\omega) \leq n^{-\varepsilon}$$

for every n large enough.

Moreover, by definition of $\overline{H}_2(\mu)$ and our choice of k_n , we have for every n large enough

$$n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2 \geq (\log n)^{-d}.$$

First of all, we choose $\beta \gg 1$ so that for any n large enough

$$\Gamma(g) \leq 1.$$

Then, for the first term in (19), we have

$$\frac{n^3 (g + k_n) \int \max_{C'_{k_n}} \mu_{\omega}(C'_{k_n}) d\mathbb{P} \sum_{C_{k_n}} \mu(C_{k_n})^2}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} = \frac{n (g + k_n) \int \max_{C'_{k_n}} \mu_{\omega}(C'_{k_n}) d\mathbb{P}}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta-1}} \leq \frac{(g + k_n) n^{-\varepsilon}}{(\log n)^{-b(2\delta-1)}}.$$

For the second term in (19), we have

$$\frac{n^2(g+k_n)^2 \sum_{C'_{k_n}} \mu(C'_{k_n})^2}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} = \frac{(g+k_n)^2}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta-1}}$$

$$\leq \frac{(g+k_n)}{(\log n)^{-b(2\delta-1)}}.$$

For the third term in (19), we have

$$\frac{n^2(g+k_n)^2 \left(\int \max_{C_{k_n}} \mu_\omega(C_{k_n}) d\mathbb{P} \right)^2}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} \leq \frac{n^{-2\varepsilon}(g+k_n)^2}{(\log n)^{-2b\delta}}.$$

For the fourth term in (19), we have

$$\frac{n(g+k_n)^3 \int \max_{C_{k_n}} \mu_\omega(C_{k_n}) d\mathbb{P}}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} \leq \frac{n^{-\varepsilon}(g+k_n)^3}{(\log n)^{-2b\delta}}.$$

And, for the fifth term in (19), we have

$$\frac{n^3(g+k_n)(\sum_{C_{k_n}} \mu(C_{k_n})^2)^2}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta}} = \frac{(g+k_n)(\sum_{C_{k_n}} \mu(C_{k_n})^2)^{1/2}}{(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2)^{2\delta-3/2}}$$

$$\leq \frac{(g+k_n)}{(\log n)^{-b(2\delta-3/2)}}.$$

Finally, putting all these estimates together in (19), choosing $d \ll -1$ and since $3/4 < \delta < 1$, we obtain

$$\mathbb{P}\left(\left|\mathbb{E}_\omega(S_n) - n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right| \geq \left(n^2 \sum_{C_{k_n}} \mu(C_{k_n})^2\right)^\delta\right) = \mathcal{O}\left(\frac{1}{\log n}\right).$$

□

Acknowledgements

The author would like to thank Rodrigo Lambert for various comments on a first draft of the paper, Mike Todd for fruitful discussions and for fixing the mistake found in [35] and the referee for useful suggestions to improve the paper.

References

- [1] M. Abadi and L. Cardeno. Renyi entropies and large deviations for the first-match function. *IEEE Trans. Inf. Theory* **61** (4) (2015) 1629–1639. MR3332970 <https://doi.org/10.1109/TIT.2015.2406695>
- [2] M. Abadi and R. Lambert. From the divergence between two measures to the shortest path between two observables. *Ergod. Theory Dyn.* **39** (7) (2019) 1729–1744. MR3961671 <https://doi.org/10.1017/etds.2017.114>
- [3] M. Abadi and N. Vergne. Poisson approximation for search of rare words in DNA sequences. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** (2008) 223–244. MR2438672
- [4] R. Arratia, L. Gordon and M. Waterman. An extreme value theory for sequence matching. *Ann. Statist.* **14** (3) (1986) 971–993. MR0856801 <https://doi.org/10.1214/aos/1176350045>
- [5] R. Arratia and M. Waterman. An Erdős–Rényi law with shifts. *Adv. Math.* **55** (1985) 13–23. MR0772069 [https://doi.org/10.1016/0001-8708\(85\)90003-9](https://doi.org/10.1016/0001-8708(85)90003-9)
- [6] R. Arratia and M. Waterman. Critical phenomena in sequence matching. *Ann. Probab.* **13** (4) (1985) 1236–1249. MR0806221
- [7] R. Arratia and M. Waterman. The Erdős–Rényi strong law for pattern matching with a given proportion of mismatches. *Ann. Probab.* **17** (3) (1989) 1152–1169. MR1009450
- [8] R. Arratia and M. Waterman. A phase transition for the score in matching random sequences allowing deletions. *Ann. Appl. Probab.* **4** (1) (1994) 200–225. MR1258181
- [9] V. Barros, L. Liao and J. Rousseau. On the shortest distance between orbits and the longest common substring problem. *Adv. Math.* **334** (2019) 311–339. MR3897435 <https://doi.org/10.1016/j.aim.2019.01.001>
- [10] T. Bogenschütz and V. M. Gundlach. Ruelles’ transfer operator for random subshifts of finite type. *Ergodic Theory Dynam. Systems* **15** (1995) 413–447. MR1336700 <https://doi.org/10.1017/S0143385700008464>
- [11] G. Ciuperca, V. Girardin and L. Lhote. Computation and estimation of generalized entropy rates for denumerable Markov chains. *IEEE Trans. Inf. Theory* **57** (2011) 4026–4034. MR2840440 <https://doi.org/10.1109/TIT.2011.2133710>

- [12] P. Collet, C. Giardina and F. Redig. Matching with shift for one-dimensional Gibbs measures. *Ann. Appl. Probab.* **19** (4) (2009) 1581–1602. MR2538081 <https://doi.org/10.1214/08-AAP588>
- [13] A. Coutinho, R. Lambert and J. Rousseau. Matching strings in encoded sequences. *Bernoulli* **26** (3) (2020) 2021–2050. MR4091100 <https://doi.org/10.3150/19-BEJ1181>
- [14] L. Debowski. Maximal repetition and zero entropy rate. *IEEE Trans. Inf. Theory* **64** (4) (2018) 2212–2219. MR3782249 <https://doi.org/10.1109/TIT.2017.2733535>
- [15] A. Dembo, S. Karlin and O. Zeitouni. Critical phenomena for sequence matching with scoring. *Ann. Probab.* **22** (4) (1994) 1993–2021. MR1331213
- [16] M. Denker, Y. Kifer and M. Stadlbauer. Conservativity of random Markov fibred systems. *Ergodic Theory Dynam. Systems* **28** (2008) 67–85. MR2380303 <https://doi.org/10.1017/S0143385707000594>
- [17] M. Denker, Y. Kifer and M. Stadlbauer. Thermodynamic formalism for random countable Markov shifts. *Discrete Contin. Dyn. Syst.* **22** (2008) 131–164. MR2410952 <https://doi.org/10.3934/dcds.2008.22.131>
- [18] A. C. M. Freitas, J. M. Freitas and S. Vaienti. Extreme value laws for non stationary processes generated by sequential and random dynamical systems. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** (3) (2017) 1341–1370. MR3689970 <https://doi.org/10.1214/16-AIHP757>
- [19] N. Haydn and M. Todd. Return times at periodic points in random dynamics. *Nonlinearity* **30** (1) (2017) 73–89. MR3604604 <https://doi.org/10.1088/0951-7715/30/1/73>
- [20] N. Haydn and S. Vaienti. The Rényi entropy function and the large deviation of short return times. *Ergodic Theory Dynam. Systems* **30** (1) (2010) 159–179. MR2586350 <https://doi.org/10.1017/S0143385709000030>
- [21] S. Karlin and F. Ost. Maximal length of common words among random letter sequences. *Ann. Probab.* **16** (1988) 535–563. MR0929062
- [22] Y. Kifer. Limit theorems for random transformations and processes in random environments. *Trans. Amer. Math. Soc.* **350** (4) (1998) 1481–1518. MR1451607 <https://doi.org/10.1090/S0002-9947-98-02068-6>
- [23] Y. Kifer. Thermodynamic formalism for random transformation revisited. *Stoch. Dyn.* **8** (1) (2008) 77–102. MR2399927 <https://doi.org/10.1142/S0219493708002238>
- [24] Y. Kifer and P.-D. Liu. Random dynamics. In *Handbook of Dynamical Systems* 379–499. B. Hasselblatt and A. Katok (Eds), 2006. MR2186245 [https://doi.org/10.1016/S1874-575X\(06\)80030-5](https://doi.org/10.1016/S1874-575X(06)80030-5)
- [25] I. Kontoyiannis, P. H. Algoet, Y. M. Suhov and A. J. Wyner. Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans. Inf. Theory* **44** (3) (1998) 1319–1327. MR1616653 <https://doi.org/10.1109/18.669425>
- [26] M. Kotani and T. Sunada. The pressure and higher correlations for an Anosov diffeomorphism. *Ergodic Theory Dynam. Systems* **21** (3) (2001) 807–821. MR1836433 <https://doi.org/10.1017/S0143385701001407>
- [27] T. Luczak and W. Szpankowski. A suboptimal lossy data compression based on approximate pattern matching. *IEEE Trans. Inf. Theory* **43** (1997) 1439–1451. MR1476778 <https://doi.org/10.1109/18.623143>
- [28] M. Mansson. On compound Poisson approximation for sequence matching. *Combin. Probab. Comput.* **9** (2000) 529–548. MR1816101 <https://doi.org/10.1017/S096354830000448X>
- [29] C. Neuhauser. A phase transition for the distribution of matching blocks. *Combin. Probab. Comput.* **5** (1996) 139–159. MR1400960 <https://doi.org/10.1017/S0963548300001930>
- [30] W. Parry and M. Pollicott. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* **187–188** (1990). MR1085356
- [31] G. Reinert, S. Schbath and M. Waterman. *Applied Combinatorics on Words On Words with Applications to Biological Sequences. Encyclopedia of Mathematics and Its Applications, Chapter Statistics* **105**. Cambridge University Press, Cambridge, 2005. MR2165687 <https://doi.org/10.1017/CBO9781107341005>
- [32] J. Rousseau. Hitting time statistics for observations of dynamical systems. *Nonlinearity* **27** (2014) 2377–2392. MR3266858 <https://doi.org/10.1088/0951-7715/27/9/2377>
- [33] J. Rousseau and B. Saussol. Poincaré recurrence for observations. *Trans. Amer. Math. Soc.* **362** (2010) 5845–5859. MR2661498 <https://doi.org/10.1090/S0002-9947-2010-05078-0>
- [34] J. Rousseau, B. Saussol and P. Varandas. Exponential law for random subshifts of finite type. *Stochastic Process. Appl.* **124** (2014) 3260–3276. MR3231619 <https://doi.org/10.1016/j.spa.2014.04.016>
- [35] J. Rousseau and M. Todd. Hitting times and periodicity in random dynamics. *J. Stat. Phys.* **161** (2015) 131–150. MR3392511 <https://doi.org/10.1007/s10955-015-1325-7>
- [36] R. Rühr and R. Shi. Quantitative Multiple pointwise convergence and effective multiple correlations. Preprint. Available at arXiv:1904.13055.
- [37] P. Shields. String matching: The ergodic case. *Ann. Probab.* **20** (3) (1992) 1199–1203. MR1175257
- [38] M. Stadlbauer. Coupling methods for random topological Markov chains. *Ergodic Theory Dynam. Systems* **37** (2017) 971–994. MR3628927 <https://doi.org/10.1017/etds.2015.61>
- [39] M. Waterman. *Introduction to Computational Biology: Maps, Sequences and Genomes*. Chapman and Hall, London, 1995.