

Matching strings in encoded sequences

ADRIANA COUTINHO¹, RODRIGO LAMBERT² and JÉRÔME ROUSSEAU^{3,4}

¹*Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, CEP 05508-090, São Paulo-SP, Brazil. E-mail: adcoutinho.s@gmail.com*

²*Faculdade de Matemática, Universidade Federal de Uberlândia, Av. João Naves de Avila, 2121, CEP 38408-100, Uberlândia-MG, Brazil. E-mail: rodrigolambert@ufu.br; url: sites.google.com/view/rodrigolambert*

³*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*

⁴*Departamento de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, 40170-110 Salvador, Brazil*

We investigate the length of the longest common substring for encoded sequences and its asymptotic behaviour. The main result is a strong law of large numbers for a re-scaled version of this quantity, which presents an explicit relation with the Rényi entropy of the source. We apply this result to the zero-inflated contamination model and the stochastic scrabble. In the case of dynamical systems, this problem is equivalent to the shortest distance between two observed orbits and its limiting relationship with the correlation dimension of the pushforward measure. An extension to the shortest distance between orbits for random dynamical systems is also provided.

Keywords: coding; correlation dimension; random dynamical systems; Rényi entropy; shortest distance; string matching

1. Introduction

Finding patterns on symbolic strings has been a widely studied subject matter on Genetics, Probability and Information Theory over the years. The investigations about how much information a n -string has on the whole realization of the process are naturally linked with the concept of redundancy and compression algorithms. On the other hand, the overlap between (some proportion of) two different strings can give us some knowledge about the similarity of the sources that generate those processes. Moreover, repetition and similarity are two well-exploited concepts in the study of DNA sequences.

In view of repetition, one of the earliest studied quantities was the well-known Ornstein–Weiss return time. A strong law of large numbers for this quantity and its explicit relationship with the entropy of the source was stated in [32] and the convergence in distribution has been widely studied (see, e.g., the reviews [2,24,37]). An interesting and intuitive link between return times and the notion of data compression schemes can be found in [41], and a consistent estimator for the entropy based on that quantity was provided in [26]. We remark also the first return of a string to its own n -cylinder (which is an outspread of the return times investigation), which can be found in [1,3,5–7,22,23,38], and references therein.

On the other hand, the notion of coincidence has been exploited on the context of *waiting times* [19,41,42]. In [42] it was proved an exponential limiting distribution for the waiting time (properly re-scaled), when the source-measure is ψ -mixing with exponential decay of correlations. We recall that in this paper the author considered two independent copies of the same process.

On the Erdős–Rényi scenario, the similarity between two sources has also been widely investigated (we refer the reader to [17,18,29,31] and references therein). As a recent example, we recall the shortest path between two observables defined in [4]. In this work, the authors proved an almost-sure linear increasing, a large deviation principle and a weak convergence for the shortest path function. All these results were linked to the divergence between two measures, which is essentially a measure of similarity between the two source-measures.

Holding on the same scenario, a remarkable matching quantity has been studied in [9]: $M_n(x, y)$, the length of the *longest matching consecutive subsequence* (or longest common substring) between two sequences. More precisely, if x and y are two realizations of the stochastic processes $(X_n)_{n \in \mathbb{N}}$ and $(Y_n)_{n \in \mathbb{N}}$,

$$M_n(x, y) = \max\{k : x_i^{i+k-1} = y_j^{j+k-1} \text{ for some } 0 \leq i, j \leq n - k\},$$

where x_i^{i+k-1} (respectively y_j^{j+k-1}) denotes the substring $x_i x_{i+1} \cdots x_{i+k-1}$ (respectively $y_j y_{j+1} \cdots y_{j+k-1}$).

If the two processes are independent and identically distributed, and generated by the same source \mathbb{P} , the authors proved that $M_n/(\log_{1/p} n) \rightarrow 2$ for almost every realization (x, y) , with $p = \mathbb{P}(X_0 = Y_0)$ [9]. Furthermore, they also proved that the same result holds for Markov chains, but with p being the largest eigenvalue of the matrix $[(p_{ij})^2]$, where $[p_{ij}]$ is the transition matrix. This result was recently generalized in [14] for α -mixing processes with exponential decay and ψ -mixing processes with polynomial decay with a limit depending on the Rényi entropy of \mathbb{P} and in [33] for random sequences in random environment. We recall that weak convergence theorems for sequence matching were also investigated over the last years (e.g., [29,31]).

Further generalizations of such quantity has also appeared on the literature. An interesting example was the sequence matching with scores introduced in [8]. In this paper, the authors consider that each symbol in the alphabet has a particular score (or weight). Therefore, each match score becomes a function which depends on the match size and the weights of the symbols as well. In the i.i.d. case and for Markov chains, they obtain a strong law of large numbers for the highest-scoring matching substring. A generalization for this statement (for instance, allowing incomplete matches) can be found in [17,18].

Following the direction of the pattern investigation between strings, one can ask if some of the above mentioned results hold if we transform our sequences following certain rules of modification. In other words: what happens if we consider encoded sequences as our interest objects of investigation?

In this paper, we study a version of the longest matching substring problem when the orbits are encoded by a measurable function (which we call *encoder* or *observation*, depending on the context). We call it the *longest common substring between encoded strings*. More precisely, let χ (respectively $\tilde{\chi}$) be an alphabet, $\chi^{\mathbb{N}}$ (respectively, $\tilde{\chi}^{\mathbb{N}}$) the space of all sequences with symbols in χ (respectively $\tilde{\chi}$) and let $f : \chi^{\mathbb{N}} \rightarrow \tilde{\chi}^{\mathbb{N}}$ be a measurable function (following the terminology of [25], we will call f an encoder (one can also see [39] where f is called a coder)). Given two

sequences $x, y \in \chi^{\mathbb{N}}$, we define the n -length of the longest common substring for the encoded pair $(f(x), f(y))$ by

$$M_n^f(x, y) = \max\{k : f(x)_i^{i+k-1} = f(y)_j^{j+k-1} \text{ for some } 0 \leq i, j \leq n - k\},$$

where $f(x)_i^{i+k-1}$ and $f(y)_j^{j+k-1}$ denotes the substrings (of the encoded sequences $f(x)$ and $f(y)$) of length k beginning in $f(x)_i$ and $f(x)_j$ respectively.

In the symbolic case, we prove an almost sure convergence for M_n^f . Namely, we provide necessary conditions on the encoder as well as in the source to prove that M_n^f grows logarithmically fast in n . It is in fact a law of large numbers with limiting rate linked with the Rényi entropy of the pushforward measure (denoted by $H_2(f_*\mathbb{P})$). Namely, if \mathbb{P} is the source-measure, then

$$\lim_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} = \frac{2}{H_2(f_*\mathbb{P})} \quad \mathbb{P} \otimes \mathbb{P}\text{-a.s.} \tag{*}$$

In the context of stochastic coding, (*) shows to be rather applicable. As a first illustration of this feature, we generalize the results from the *stochastic scrabble* given by [8], from a Markov chain to a general α -mixing process with exponential decay. The second application deals with the stochastic noise (or contamination encoder), which can be viewed in [16,21].

Recently, [14] showed that the problem of the longest common substring for stochastic processes is related to the shortest distance between two orbits and, to the best of our knowledge, this was the first article where this quantity was defined and studied. Following this idea, we can observe that, in dynamical systems, the correspondent of the longest common substring for the encoded pair is the shortest distance between observed orbits. Formally, let $f : X \rightarrow Y$ be a measurable function, called the observation. If we consider a dynamical system (X, T, μ) , we investigate the asymptotic behavior of

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

and prove that its limiting behavior is related to the correlation dimension of the pushforward measure $f_*\mu$ (denoted $C_{f_*\mu}$). If f is the identity in X , we recover the shortest distance between two orbits problem, studied in [14]. In that paper, the authors provide a law of large numbers and related it with the correlation dimension of the source measure. In the present paper, we generalize this result for a family of observations, concluding that the limiting rate is given by the dimension of the pushforward measure $f_*\mu$ (under suitable conditions on f). Namely, for rapidly mixing systems,

$$\lim_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} = \frac{2}{C_{f_*\mu}} \quad \mu \otimes \mu\text{-a.s.}, \tag{**}$$

provided that $C_{f_*\mu}$ exists.

In [30,34,35], the study of observed orbits (in particular, the study of return and hitting time) was used to obtain results for random dynamical systems. Following this idea, we combine (**)

with a particular observation f to obtain the following strong law of large numbers for random dynamical systems (provided that C_ν exists)

$$\lim_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} = \frac{2}{C_\nu} \quad \mu \otimes \mu\text{-a.s.}, \tag{***}$$

where $m_n^{\omega, \tilde{\omega}}(x, \tilde{x})$ is the shortest distance between two random orbits ($\{x, T_\omega x, \dots, T_\omega^n x\}$ and $\{\tilde{x}, T_{\tilde{\omega}} \tilde{x}, \dots, T_{\tilde{\omega}}^n \tilde{x}\}$) and C_ν is the correlation dimension of the stationary measure (we refer the reader to Section 4 for more details). We present then a collection of applications for this statement. The first one treats non i.i.d. random dynamical systems. The second deals with random perturbed dynamics. We finish the applications with random hyperbolic toral automorphisms.

The rest of this paper is organized as follows. In Section 2, we study the relation between the longest common substrings for encoded sequences and the Rényi entropy. We state precisely result (★) and apply it to the stochastic scrabble and the zero-inflated contamination model. In Section 3, we analyse the behaviour of the shortest distance between observed orbits of a dynamical system and present result (★★). Section 4 deals with the case of random dynamical systems and states result (★★★), as well as some applications.

2. Reaching Rényi entropy via string matching of encoded sequences

The present section is dedicated to study of the longest common substring of encoded sequences. We start by presenting some terminology and definitions, in order to introduce the problem.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, where $\Omega = \chi^{\mathbb{N}}$ for some alphabet χ , \mathcal{F} the sigma-algebra generated by the n -cylinders in Ω and \mathbb{P} is a stationary probability measure on \mathcal{F} . If σ is the left shift on Ω , we can see $(\Omega, \mathcal{F}, \mathbb{P}, \sigma)$ as a symbolic dynamical system with \mathbb{P} σ -invariant. Let $\tilde{\Omega} = \tilde{\chi}^{\mathbb{N}}$ for some alphabet $\tilde{\chi}$ and $\tilde{\mathcal{F}}$ the sigma-algebra generated by the n -cylinders in $\tilde{\Omega}$.

Definition 2.1. Let $f : \Omega \rightarrow \tilde{\Omega}$ be an encoder. Given two sequences $x, y \in \Omega$, we define the n -length of the longest common substring for the encoded pair $(f(x), f(y))$ by

$$M_n^f(x, y) = \max\{k : f(x)_i^{i+k-1} = f(y)_j^{j+k-1} \text{ for some } 0 \leq i, j \leq n - k\},$$

where $f(x)_i^{i+k-1}$ and $f(y)_j^{j+k-1}$ denote the substrings of length k beginning in $f(x)_i$ and $f(y)_j$ respectively.

For $y \in \Omega$ (respectively, $\tilde{\Omega}$) we denote by $C_n(y)$ the n -cylinder containing y , that is, the set of sequences $z \in \Omega$ (respectively, $\tilde{\Omega}$) such that $z_i = y_i$ for any $i = 0, \dots, n - 1$. We denote \mathcal{F}_0^n (respectively $\tilde{\mathcal{F}}_0^n$) the sigma-algebra on Ω (respectively, $\tilde{\Omega}$) generated by all n -cylinders.

Definition 2.2. The lower and upper Rényi entropies of a measure \mathbb{P} are defined as

$$\underline{H}_2(\mathbb{P}) = - \liminf_{k \rightarrow \infty} \frac{1}{k} \log \sum_{C_k} \mathbb{P}(C_k)^2 \quad \text{and} \quad \overline{H}_2(\mathbb{P}) = - \overline{\lim}_{k \rightarrow \infty} \frac{1}{k} \log \sum_{C_k} \mathbb{P}(C_k)^2,$$

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

In general, the existence of the Rényi entropy is not known. However, it was computed in some particular cases: Bernoulli shift, Markov chains and Gibbs measure of a Hölder-continuous potential [23]. The existence was also proved for ϕ -mixing measures [28], for weakly ψ -mixing processes [23] and for ψ_g -regular processes [6]. In Section 2.1, we will prove that for Markov chains, the Rényi entropy does not depend on the initial distribution but only on the transition matrix and that one can compute the Rényi entropy even if the measure is not stationary.

Definition 2.3. Consider the dynamical system $(\Omega, \mathbb{P}, \sigma)$. We say that it is α -mixing if there exists a function $\alpha : \mathbb{N} \rightarrow \mathbb{R}$ where $\alpha(g)$ converges to zero when g goes to infinity and such that

$$\sup_{A \in \mathcal{F}_0^n; B \in \mathcal{F}_0^m} |\mathbb{P}(A \cap \sigma^{-g-n} B) - \mathbb{P}(A)\mathbb{P}(B)| \leq \alpha(g), \tag{2}$$

for all $m, n \in \mathbb{N}$.

We say that the system is ψ -mixing if there exists a function $\psi : \mathbb{N} \rightarrow \mathbb{R}$ where $\psi(g)$ converges to zero when g goes to infinity and such that

$$\sup_{A \in \mathcal{F}_0^n; B \in \mathcal{F}_0^m} \left| \frac{\mathbb{P}(A \cap \sigma^{-g-n} B) - \mathbb{P}(A)\mathbb{P}(B)}{\mathbb{P}(A)\mathbb{P}(B)} \right| \leq \psi(g), \tag{3}$$

for all $m, n \in \mathbb{N}$. In the cases that $\alpha(g)$ or $\psi(g)$ decreases exponentially fast to zero, we say that the system has an exponential decay.

Now we are ready to present the main result of this section. It states that, under suitable conditions and large values of n , the longest common substring behaves like $\log n$, for almost all realizations.

Theorem 2.4. Consider $f : \Omega \rightarrow \tilde{\Omega}$ an encoder such that $\underline{H}_2(f_*\mathbb{P}) > 0$. For $\mathbb{P} \otimes \mathbb{P}$ -almost every $(x, y) \in \Omega \times \Omega$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} \leq \frac{2}{\underline{H}_2(f_*\mathbb{P})}. \tag{4}$$

Moreover, if

- (i) the system $(\Omega, \mathbb{P}, \sigma)$ is α -mixing with an exponential decay (or ψ -mixing with $\psi(g) = g^{-a}$ for some $a > 0$);
- (ii) $C_n \in \tilde{\mathcal{F}}_0^n$ implies $f^{-1}C_n \in \mathcal{F}_0^{h(n)}$, where $h(n) = o(n^\gamma)$, for some $\gamma > 0$,

then, for $\mathbb{P} \otimes \mathbb{P}$ -almost every $(x, y) \in \Omega \times \Omega$,

$$\underline{\lim}_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} \geq \frac{2}{\overline{H}_2(f_*\mathbb{P})}. \tag{5}$$

Therefore, if the Rényi entropy exists, we get for $\mathbb{P} \otimes \mathbb{P}$ -almost every $(x, y) \in \Omega \times \Omega$,

$$\lim_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} = \frac{2}{H_2(f_*\mathbb{P})}. \tag{*}$$

Remark 2.5. We emphasize that to obtain this result one cannot apply directly Theorem 7 of [14] since in general the pushforward measure $f_*\mathbb{P}$ is not stationary (see, e.g., Section 2.3).

Proof. For simplicity, we assume $\alpha(g) = e^{-g}$. The ψ -mixing case can be obtained by a simple modification. The proof of this theorem follows the lines of the proof of the Theorem 7 in [14], but an extra care is needed (mainly in the second part of the proof) since we are working with pre-image of cylinders (instead of cylinders in [14]).

In the first part of the proof, for $\epsilon > 0$ we denote

$$k_n = \left\lceil \frac{2 \log n + \log \log n}{\underline{H}_2(f_*\mathbb{P}) - \epsilon} \right\rceil.$$

Let us also denote

$$A_{i,j}^f(y) = \sigma^{-i} [f^{-1} C_{k_n}(f(\sigma^j y))]$$

and

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

We first show that the event $\{M_n^f \geq k_n\}$ occurs only finitely many times. It follows from definition of S_n^f and Markov's inequality that

$$\mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) \geq k_n\}) = \mathbb{P} \otimes \mathbb{P}(\{(x, y) : S_n^f(x, y) \geq 1\}) \leq \mathbb{E}(S_n^f).$$

Moreover, by computing the expected value of S_n^f we get

$$\begin{aligned} \mathbb{E}(S_n^f) &= \int \int \sum_{i,j=1,\dots,n} \mathbb{1}_{A_{i,j}^f(y)}(x) d\mathbb{P}(x) d\mathbb{P}(y) \\ &= \sum_{i,j=1,\dots,n} \int \mathbb{P}(f^{-1} C_{k_n}(f(\sigma^j y))) d\mathbb{P}(y) \\ &= n^2 \int f_*\mathbb{P}(C_{k_n}(f(y))) d\mathbb{P}(y). \end{aligned}$$

Thus,

$$\mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) \geq k_n\}) \leq n^2 \int f_*\mathbb{P}(C_{k_n}(f(y))) d\mathbb{P}(y).$$

For large values of n , by definition of $\underline{H}_2(f_*\mathbb{P})$ it holds

$$\int f_*\mathbb{P}(C_{k_n}(f(y))) d\mathbb{P}(y) = \sum_{C_{k_n}} f_*\mathbb{P}(C_{k_n})^2 \leq e^{-k_n(\underline{H}_2(f_*\mathbb{P})-\epsilon)}.$$

Moreover by definition of k_n ,

$$\mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) \geq k_n\}) \leq n^2 e^{-k_n(\underline{H}_2(f_*\mathbb{P})-\epsilon)} \leq \frac{1}{\log n}.$$

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

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

Since the last quantity is summable in κ , the Borel–Cantelli lemma gives that if κ is large enough, then for almost every pair (x, y) it holds

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

and then

$$\frac{M_{n_\kappa}^f(x, y)}{\log n_\kappa} \leq \frac{1}{\underline{H}_2(f_*\mathbb{P}) - \epsilon} \left(2 + \frac{1 + \log \log n_\kappa}{\log n_\kappa} \right). \tag{6}$$

We observe that for all n , there exists κ such that $e^\kappa \leq n \leq e^{\kappa+1}$. In addition, since $(M_n^f)_{n \in \mathbb{N}}$ is an increasing sequence, we get

$$\frac{M_{n_\kappa}^f(x, y)}{\log n_{\kappa+1}} \leq \frac{M_n^f(x, y)}{\log n} \leq \frac{M_{n_{\kappa+1}}^f(x, y)}{\log n_\kappa}. \tag{7}$$

Taking the limit superior in the above inequalities and observing that $\lim_{\kappa \rightarrow \infty} \frac{\log n_\kappa}{\log n_{\kappa+1}} = 1$ by (7) we obtain

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} = \overline{\lim}_{\kappa \rightarrow \infty} \frac{M_{n_\kappa}^f(x, y)}{\log n_\kappa}.$$

Thus, by (6) we have

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n^f(x, y)}{\log n} \leq \frac{2}{\underline{H}_2(f_*\mathbb{P}) - \epsilon}.$$

Since ϵ can be arbitrarily small, (4) is proved.

Despite some similarities with Theorem 7 in [14], we emphasize that second part of the present proof is quite different, in particular since the length of the encoded sequences may be changed by the encoder.

We will now prove (5). In order to do that denote, for $\epsilon > 0$,

$$k_n = \left\lfloor \frac{2 \log n + b \log \log n}{\overline{H}_2(f_*\mathbb{P}) + \epsilon} \right\rfloor$$

where b is a constant to be chosen.

Note that by definition of S_n^f we have

$$\begin{aligned} \mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) < k_n\}) &= \mathbb{P} \otimes \mathbb{P}(\{(x, y) : S_n^f(x, y) = 0\}) \\ &\leq \mathbb{P} \otimes \mathbb{P}(\{(x, y) : |S_n^f(x, y) - \mathbb{E}(S_n^f)| \geq |\mathbb{E}(S_n^f)|\}). \end{aligned}$$

By Chebyshev's inequality we deduce that

$$\mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) < k_n\}) \leq \frac{\text{var}(S_n^f)}{\mathbb{E}(S_n^f)^2}.$$

We have to estimate the variance of S_n^f .

We see at once that

$$\begin{aligned} \text{var}(S_n^f) &= \sum_{1 \leq i, i', j, j' \leq n} \text{cov}(\mathbb{1}_{A_{ij}^f}, \mathbb{1}_{A_{i'j'}^f}) \\ &= \sum_{1 \leq i, i', j, j' \leq n} \int \int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^j y)))(\sigma^i x)} \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^{j'} y)))(\sigma^{i'} x)} \\ &\quad - n^4 \left(\sum_{C_{k_n}} f_*\mathbb{P}(C_{k_n})^2 \right)^2. \end{aligned} \tag{8}$$

Let $g = g(n) = (\log n)^\beta$, for some $\beta > \max\{1, \gamma\}$. There are four cases to consider.

Case 1: $|i - i'| > g + k_n$. Using the α -mixing condition, we have

$$\begin{aligned} &\int \left(\int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^j y)))(\sigma^{i-i'} x)} \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^{j'} y)))(x)} d\mathbb{P}(x) \right) d\mathbb{P}(y) \\ &\leq \alpha(g + k_n - h(k_n)) \\ &\quad + \int \left(\int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^j y)))(x)} d\mathbb{P}(x) \int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^{j'} y)))(x)} d\mathbb{P}(x) \right) d\mathbb{P}(y) \\ &= \alpha(g + k_n - h(k_n)) + \int f_*\mathbb{P}(C_{k_n}(f(\sigma^j y))) f_*\mathbb{P}(C_{k_n}(f(\sigma^{j'} y))) d\mathbb{P}(y). \end{aligned} \tag{9}$$

To estimate the first term of the sum above, we analyse two cases.

Case 1.1: $|j - j'| > g + k_n$. In this case, we have

$$\begin{aligned}
 & \int f_*\mathbb{P}(C_{k_n}(f(\sigma^j y))) f_*\mathbb{P}(C_{k_n}(f(\sigma^{j'} y))) d\mathbb{P}(y) \\
 &= \int f_*\mathbb{P}(C_{k_n}(f(\sigma^{j-j'} y))) f_*\mathbb{P}(C_{k_n}(f(y))) d\mathbb{P}(y) \\
 &= \sum_{C_{k_n}, C'_{k_n}} \int_{f^{-1}(C_{k_n}) \cap \sigma^{j-j'}(f^{-1}(C'_{k_n}))} f_*\mathbb{P}(C_{k_n}) f_*\mathbb{P}(C'_{k_n}) d\mathbb{P}(y) \\
 &= \sum_{C_{k_n}, C'_{k_n}} f_*\mathbb{P}(C_{k_n}) f_*\mathbb{P}(C'_{k_n}) \mathbb{P}(f^{-1}(C_{k_n}) \cap \sigma^{j-j'}(f^{-1}(C'_{k_n}))).
 \end{aligned}$$

Using the α -mixing condition in the last expression, we get that

$$\begin{aligned}
 & \int f_*\mathbb{P}(C_{k_n}(f(\sigma^j y))) f_*\mathbb{P}(C_{k_n}(f(\sigma^{j'} y))) d\mathbb{P}(y) \\
 & \leq \sum_{C_{k_n}, C'_{k_n}} f_*\mathbb{P}(C_{k_n}) f_*\mathbb{P}(C'_{k_n}) (f_*\mathbb{P}(C_{k_n}) f_*\mathbb{P}(C'_{k_n})) \\
 & \quad + \sum_{C_{k_n}, C'_{k_n}} f_*\mathbb{P}(C_{k_n}) f_*\mathbb{P}(C'_{k_n}) (\alpha(g + k_n - h(k_n))) \\
 & = \alpha(g + k_n - h(k_n)) + \left(\sum_{C_{k_n}} f_*\mathbb{P}(C_{k_n})^2 \right)^2.
 \end{aligned} \tag{10}$$

Case 1.2 $|j - j'| \leq g + k_n$. By Hölder's inequality it follows that,

$$\begin{aligned}
 & \int f_*\mathbb{P}(C_{k_n}(f(\sigma^j y))) f_*\mathbb{P}(C_{k_n}(f(\sigma^{j'} y))) d\mathbb{P}(y) \\
 & \leq \left(\int f_*\mathbb{P}(C_{k_n}(f(\sigma^j y)))^2 d\mathbb{P}(y) \right)^{1/2} \left(\int f_*\mathbb{P}(C_{k_n}(f(\sigma^{j'} y)))^2 d\mathbb{P}(y) \right)^{1/2} \\
 & = \sum_{C_{k_n}} f_*\mathbb{P}(C_{k_n})^3 \\
 & \leq \left(\sum_{C_{k_n}} f_*\mathbb{P}(C_{k_n})^2 \right)^{3/2}
 \end{aligned} \tag{11}$$

where the last inequality comes from the subadditivity of the function $z(x) = x^{2/3}$.

Case 2. $|i - i'| \leq g + k_n$:

Case 2.1 $|j - j'| > g + k_n$:

By symmetry, this case is analogous to the Case 1.2.

Case 2.2. $|j - j'| \leq g + k_n$:

$$\begin{aligned} & \int \int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^j y)))}(\sigma^i x) \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^{j'} y)))}(\sigma^{i'} x) d\mathbb{P}(x) d\mathbb{P}(y) \\ & \leq \int \int \mathbb{1}_{f^{-1}(C_{k_n}(f(\sigma^j y)))}(\sigma^i x) d\mathbb{P}(x) d\mathbb{P}(y) \\ & = \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2. \end{aligned} \tag{12}$$

Putting the estimates (10), (11), (12) together in (8) we get

$$\begin{aligned} \frac{\text{var}(S_n^f)}{\mathbb{E}(S_n^f)^2} & \leq \frac{3n^4 \alpha (g + k_n - h(k_n)) + 4n^3 (g + k_n) (\sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^{3/2}}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2} \\ & \quad + \frac{4n^2 (g + k_n)^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2}. \end{aligned} \tag{13}$$

We estimate each term on the right separately. Using the definition of k_n and of the Rényi entropy, for n large enough, we have for the first term

$$\frac{3n^4 \alpha (g + k_n - h(k_n))}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2} \leq \frac{3n^4 \alpha (g + k_n - h(k_n))}{(\log n)^{-2b}}.$$

By hypothesis, $h(k_n) = o((\log n)^\gamma)$. Therefore, by definition of g and k_n , for n large enough we have $g + k_n - h(k_n) > \log(n^4)$. Hence,

$$\frac{3n^4 \alpha (g + k_n - h(k_n))}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2} \leq 3(\log n)^{2b}. \tag{14}$$

To estimate the second term, we obtain

$$\begin{aligned} & \frac{4n^3 (g + k_n) (\sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^{3/2}}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2} \\ & \leq 4(g + k_n) (\log n)^{b/2} \\ & \leq 4(\log n)^{\beta+b/2} + \frac{2(\log n)^{1+b/2} + b \log(\log n) (\log n)^{b/2}}{\overline{H}_2(f_* \mathbb{P}) + \epsilon}. \end{aligned} \tag{15}$$

Finally for the third term, we get

$$\begin{aligned} & \frac{4n^2(g+k_n)^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2}{(n^2 \sum_{C_{k_n}} f_* \mathbb{P}(C_{k_n})^2)^2} \\ & \leq 4(g+k_n)^2 (\log n)^b \\ & \leq 8(\log n)^{2\beta+b} + \frac{4(\log n)^{2+b} + b^2(\log(\log n))^2(\log n)^b}{(\overline{H}_2(f_* \mathbb{P}) + \epsilon)^2}. \end{aligned} \tag{16}$$

Taking $b < -4\beta$ and substituting (14), (15) and (16) into (13), we obtain

$$\mathbb{P} \otimes \mathbb{P}(\{(x, y) : M_n^f(x, y) < k_n\}) \leq \mathcal{O}((\log n)^{-1}). \tag{17}$$

Thus, taking a subsequence $\{n_\kappa\}_\kappa = \lceil e^{\kappa^2} \rceil$ as in the proof of (4), we can use Borel–Cantelli lemma to obtain (5).

Finally, if the Rényi entropy exists, by (4) and (5) we conclude the proof of the theorem. \square

In what follows, we compute the Rényi entropy for Markov chains and then we apply the above stated theorem to some well-known cases of probability’s literature. The first one is a contamination encoder that flips to zero some symbols of the sequence and the second one gives a weight on each symbol of χ .

2.1. Rényi entropy of Markov chains

In the sequel, we present an entropy invariance statement by change of initial distribution. In particular, we will use this result in the example of the stochastic scrabble (Section 2.3) to compute the entropy of the pushforward measure.

Theorem 2.6. *Let $(X_n)_{n \in \mathbb{N}}$ be a Markov chain in a finite alphabet χ , with irreducible and aperiodic transition matrix $P = [(p_{ij})]$ and stationary measure μ . For any Markov measure ν with initial distribution π and transition matrix P it holds*

$$H_2(\nu) = H_2(\mu) = -\log p$$

where p is the largest eigenvalue of the matrix $[(p_{ij})^2]$.

Proof. First of all, we observe that the second equality is a well-known result (see, e.g., [23] Section 2.2). For the first equality, we will show the following two inequalities

$$\overline{H}_2(\nu) \leq H_2(\mu) \tag{18}$$

and

$$\underline{H}_2(\nu) \geq H_2(\mu). \tag{19}$$

For convenience here, we will adopt the following notation for strings of stochastic processes: $\{X_n^m = x_n^m\} = \{X_n = x_n, X_{n+1} = x_{n+1}, \dots, X_m = x_m\}$ for every non-negative integers n, m such that $n \leq m$ and for any realization $x = x_0^\infty$.

We will use corollary (3.13) from [20], which states that there exists $\gamma \in (0, 1)$ such that for all $k > 1$

$$\sup_{x_k \in \mathcal{X}} |v(X_k = x_k) - \mu(x_k)| \leq \gamma^k.$$

A straightforward computation gives that for every $n > k > 1$

$$\sup_{x_0, x_k \in \mathcal{X}} |v(X_k = x_k \mid X_0 = x_0) - \mu(x_k)| \leq \gamma^k$$

and for every $x_k^n \in \mathcal{X}^{n-k+1}$

$$|v(X_k^n = x_k^n) - \mu(x_k^n)| \leq c\gamma^k \mu(x_k^n)$$

with $c = (\inf_{x_0} \{\mu(x_0)\})^{-1} < +\infty$.

Let $(a_n)_{n \in \mathbb{N}}$ be a non-decreasing and unbounded sequence in n taking values on the non-negative integers and such that $n \geq a_n = o(n)$. Without loss of generality we will only consider the strings x_0^n such that $v(X_0^n = x_0^n) > 0$. On the one hand, we get

$$\begin{aligned} v(X_0^n = x_0^n) &\leq v(X_{a_n}^n = x_{a_n}^n) \\ &\leq [c\gamma^{a_n} \mu(x_{a_n}^n) + \mu(x_{a_n}^n)]. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{n} \log \sum_{x_0^n} v(X_0^n = x_0^n)^2 &\leq \frac{2}{n} \log(c\gamma^{a_n} + 1) + \frac{1}{n} \log \sum_{x_0^n} \mu(x_{a_n}^n)^2 \\ &= \frac{2}{n} \log(c\gamma^{a_n} + 1) + \frac{1}{n} \log \sum_{x_0^{a_n-1}} \sum_{x_{a_n}^n} \mu(x_{a_n}^n)^2 \\ &\leq \frac{2}{n} \log(c\gamma^{a_n} + 1) + \frac{1}{n} \log |\mathcal{X}|^{a_n} + \frac{1}{n} \log \sum_{x_{a_n}^n} \mu(x_{a_n}^n)^2. \end{aligned}$$

One can observe that the two first terms in the last line vanish as $n \rightarrow \infty$. Moreover, by stationarity of μ we obtain

$$\begin{aligned} \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x_0^n} v(X_0^n = x_0^n)^2 &\leq \overline{\lim}_{n \rightarrow \infty} \frac{1}{n} \log \sum_{x_{a_n}^n} \mu(x_{a_n}^n)^2 \\ &= \lim_{n \rightarrow \infty} \frac{1}{n - a_n} \log \sum_{x_0^{n-a_n}} \mu(x_0^{n-a_n})^2 = H_2(\mu) \end{aligned}$$

which gives us (18).

On the other hand, first notice that for strings such that $\nu(X_0^n = x_0^n) > 0$, we have for n large enough

$$\begin{aligned} \nu(X_0^n = x_0^n) &= \pi(x_0)P_{x_0x_1} \cdots P_{x_{a_n-1}x_{a_n}} P_{x_{a_n}x_{a_n+1}} \cdots P_{x_{n-1}x_n} \\ &\geq \pi(x_0)\rho^{a_n} \frac{1}{\nu(X_{a_n} = x_{a_n})} \nu(X_{a_n} = x_{a_n}^n) \\ &\geq \frac{\pi(x_0)\rho^{a_n}}{\mu(x_{a_n}) + \gamma^{a_n}} [\mu(x_{a_n}^n)(1 - \gamma^{a_n})] \\ &\geq d\rho^{a_n} [\mu(x_{a_n}^n)(1 - \gamma^{a_n})] \end{aligned}$$

where $\rho := \min_{P_{ij} > 0} P_{ij}$ and $d = \frac{1}{2} \min_{\pi(x_0) > 0} \pi(x_0)$.

Now

$$\begin{aligned} \frac{1}{n} \log \sum_{x_0^n} \nu(X_0^n = x_0^n)^2 &\geq \frac{2}{n} \log(d\rho^{a_n}) + \frac{1}{n} \log \sum_{x_0^n} [\mu(x_{a_n}^n)(1 - \gamma^{a_n})]^2 \\ &\geq \frac{2}{n} \log(d\rho^{a_n}) + \frac{2}{n} \log(1 - \gamma^{a_n}) + \frac{1}{n} \log \sum_{x_{a_n}^n} [\mu(x_{a_n}^n)]^2. \end{aligned}$$

Taking the limit inferior and observing that the first two terms in the last line vanish and the third one converges to $H_2(\mu)$ as n diverges, we obtain (19). This last statement concludes the proof. \square

2.2. The zero-inflated contamination model

Let $\{\xi_i\}_{i \in \mathbb{N}}$ be a sequence of i.i.d. random variables taking values on $\{0, 1\}$, independently of \mathbb{P} , and governed by a Bernoulli measure such that $\mu(\xi_i = 1) = 1 - \epsilon$, where $\epsilon \in (0, 1)$ is the noise parameter. Let $f_\xi : \Omega \rightarrow \Omega$ be a perturbation given by $f_\xi(z) = \{\xi_i z_i\}_{i \in \mathbb{N}}$. This defines the zero inflated contamination model (see [16,21]).

Then, if $\underline{H}_2(f_\xi * \mathbb{P}) > 0$, for $\mathbb{P} \otimes \mathbb{P}$ -almost every $(z, t) \in \Omega \times \Omega$,

$$\overline{\lim}_{n \rightarrow \infty} \frac{M_n^{f_\xi}(z, t)}{\log n} \leq \frac{2}{\underline{H}_2(f_\xi * \mathbb{P})}.$$

Moreover, if the system $(\Omega, \mathbb{P}, \sigma)$ is α -mixing with an exponential decay, for $\mathbb{P} \otimes \mathbb{P}$ -almost every $(z, t) \in \Omega \times \Omega$,

$$\underline{\lim}_{n \rightarrow \infty} \frac{M_n^{f_\xi}(z, t)}{\log n} \geq \frac{2}{\overline{H}_2(f_\xi * \mathbb{P})}.$$

Indeed, for k large enough $f_\xi^{-1}C_k \in \mathcal{F}_0^{m_\epsilon(k)}$, where $m_\epsilon(k)$ is the proportion of 1's in the k -cylinder $C_k(\xi)$. Let $\mu^{\otimes \mathbb{N}}$ denote the product measure that governs the stochastic process $\{\xi_i\}_{i \in \mathbb{N}}$.

One can observe that by the law of large numbers $\mu^{\otimes \mathbb{N}}$ -almost every realization ξ has an ϵ -proportion of zeros, i.e.

$$\lim_{k \rightarrow \infty} \frac{m_\epsilon(k)}{k} = 1 - \epsilon.$$

Thus, for $\mu^{\otimes \mathbb{N}}$ -almost every ξ , there exists $\epsilon_1 > 0$ such that $m_\epsilon(k) = o(k^{1+\epsilon_1})$ and thus one can apply Theorem 2.4.

Moreover, if \mathbb{P} is a Bernoulli measure we can explicitly compute the Rényi entropy of $f_\xi * \mathbb{P}$. Namely, by using the binomial theorem, for k large enough we get

$$\begin{aligned} \sum_{C_k} [\mathbb{P}(f_\xi^{-1} C_k)]^2 &= \sum_{j=1}^{m_\epsilon(k)} \binom{m_\epsilon(k)}{j} p^{2j} (1-p)^{2(m_\epsilon(k)-j)} \\ &= [p^2 + (1-p)^2]^{m_\epsilon(k)}. \end{aligned}$$

Therefore the Rényi entropy is given by

$$\begin{aligned} H_2(f_\xi * \mathbb{P}) &= - \lim_{k \rightarrow \infty} \frac{m_\epsilon(k)}{k} \log(p^2 + (1-p)^2) \\ &= -(1-\epsilon) \log(p^2 + (1-p)^2). \end{aligned}$$

We observe that if $\chi = \{a_1, \dots, a_n\}$ is a finite alphabet and $\mathbb{P}(X = a_i) = p_i$, by similar computations (and the multinomial theorem) we obtain

$$H_2(f_\xi * \mathbb{P}) = -(1-\epsilon) \log\left(\sum_i p_i^2\right) = (1-\epsilon)H_2(\mathbb{P}).$$

Thus, by Theorem 2.4, we get that for $\mu^{\mathbb{N}}$ -almost every realization of $\{\xi_i\}_{i \in \mathbb{N}}$ it holds

$$\frac{M_n^{f_\xi}}{\log n} \xrightarrow{n \rightarrow \infty} \frac{2}{(1-\epsilon)H_2(\mathbb{P})} \quad \mathbb{P} \otimes \mathbb{P}\text{-a.s.}$$

The case $f_\xi = \text{Id}$ is equivalent to $\epsilon = 0$ (no contamination), and if ϵ is close to 1 we expect to observe larger values for $M_n^{f_\xi}$ (in view of Theorem 2.4). This can be summarized with the following assertion: the more contamination, the more coincidences appear between the encoded strings. This is a rather intuitive feature of the string matching problem, which indicates that sequences which had lost much information tends to present more similarity.

2.3. Highest-scoring matching substring

In this example, we will consider the case in which a shorter match can be better scored than a long one, depending on the symbols that compose the matched strings. For this, we assume that each string is scored according to the symbols that compose it. In this sense, suppose that

each letter $a \in \chi$ is associated to a weight $v(a) \in \mathbb{N}^*$. We also denote the score of a string z_0^{m-1} by $V(z_0^{m-1}) = \sum_{j=0}^{m-1} v(z_j)$. If x and y are two realizations of the χ -valued stochastic processes $(X_n)_n$ and $(Y_n)_n$,

$$V_n(x, y) = \max_{0 \leq i, j \leq n-m} \{V(z_0^{m-1}) : \text{there exists } 1 \leq m \leq n \text{ such that } z_0^{m-1} = x_i^{i+m-1} = y_j^{j+m-1}\}$$

is the n th highest-scoring matching substring [8]. The authors also named it stochastic Scrabble, because of the namesake board game. For two copies independently generated by the same Markov source \mathbb{P} with positive transition probabilities $[p_{ij}]$, they stated the following result:

$$\lim_{n \rightarrow \infty} \frac{V_n}{\log n} = \frac{2}{-\log p} \quad \mathbb{P} \times \mathbb{P}\text{-a.s.}, \quad (20)$$

where $p \in (0, 1)$ is the largest root of $\det(P - \lambda^V) = 0$, with $P = [p_{ij}^2]$ and $\lambda^V = [\delta_{ij} \lambda^{v(i)}]$.

One can observe that this result (20) can be obtained as particular case of Theorem 2.4. Indeed, inspired by [8], we can construct a specific encoder f that stretches the sequences depending on the weights of its letters. Formally

$$\begin{aligned} f : \chi^{\mathbb{N}} &\rightarrow \chi^{\mathbb{N}} \\ x_0^\infty &\mapsto \underbrace{x_0 x_0 \cdots x_0}_{v(x_0)} \underbrace{x_1 x_1 \cdots x_1}_{v(x_1)} \cdots \underbrace{x_n x_n \cdots x_n}_{v(x_n)} \cdots \end{aligned} \quad (21)$$

With this particular encoder, we get that $M_n^f(x, y) = V_n(x, y)$ and thus to get (20) we need to compute $H_2(f_*\mathbb{P})$ and check that conditions (i) and (ii) of Theorem 2.4 are satisfied.

We recall that if (X_n) is a Markov chain in $\chi = \{1, 2, \dots, d\}$, we can see $f(X_n)$ as a Markov Chain in $\tilde{\chi}$, which is a $(\sum_{i \in \chi} v(i))$ -sized alphabet, given by

$$\tilde{\chi} = \{1_1, 1_2, \dots, 1_{v(1)}, 2_1, 2_2, \dots, 2_{v(2)}, \dots, d_1, d_2, \dots, d_{v(d)}\}.$$

In this context, we will consider that $f : \chi^{\mathbb{N}} \rightarrow \tilde{\chi}^{\mathbb{N}}$. Furthermore, if $Q = [Q_{ij}]$, $1 \leq i, j \leq d$ is the transition matrix for (X_n) we get that the transition matrix Q^* for the chain $(f(X_n))$ on $\tilde{\chi}$ is given by

$$\begin{aligned} Q_{i_\ell i_{\ell+1}}^* &= 1 && \text{if } 1 \leq \ell \leq v(i) - 1 \text{ and } 1 \leq i, j \leq d; \\ Q_{i_{v(i)} j_1}^* &= Q_{ij} && \text{if } 1 \leq i, j \leq d; \\ Q_{ij}^* &= 0 && \text{otherwise.} \end{aligned}$$

Notice that, if $v_{\min} = \min_{i \in \chi} \{v(i)\}$ is the minimum weight, we get for any cylinder C_n ,

$$f^{-1}C_n \in \mathcal{F}_0^{\lfloor \frac{n}{v_{\min}} \rfloor},$$

and since $n/v_{\min} = o(n^{1+\epsilon})$ for all $\epsilon > 0$, condition (ii) of Theorem 2.4 is then satisfied. We recall that an irreducible and aperiodic positive recurrent Markov chain is an α -mixing process

with exponential decay of correlation (see, e.g., Theorem 4.9 in [27]) which implies condition (i).

Finally, to obtain (20), we need to compute $H_2(f_*\mathbb{P})$. As in [8], to assure aperiodicity for the encoded process $f(X_n)$ we assume that $gdc\{v(1), v(2), \dots, v(d)\} = 1$.

Moreover, by Theorem (2.6), we know that the Rényi entropy of its stationary measure μ is given by $H_2(\mu) = -\log p$, where p is the largest positive eigenvalue of the matrix $[(Q^*)^2_{ij}]$, $1 \leq i, j \leq (\sum_{i \in \chi} v(i))$ (it was proved in [8] that this p is the same as the one defined in (20)). Moreover, we observe that $f_*\mathbb{P}$ is a Markov measure with initial distribution π and transition matrix Q^* , where π is defined by $\pi(i_1) = \mathbb{P}(X_0 = i)$ and $\pi(i_j) = 0$ for any $i \in \chi$ and $1 < j \leq v(i)$. It is important to notice that in general, $f_*\mathbb{P}$ is not stationary.

Thus, by Theorem (2.6), we have $H_2(\mu) = H_2(f_*\mathbb{P})$ and we can combine it with equation (\star) in Theorem 2.4 to conclude that, for $\mathbb{P} \times \mathbb{P}$ almost every pair of realizations, as n diverges it holds

$$\frac{V_n}{\log n} \longrightarrow \frac{2}{-\log p}.$$

We remark that this example generalizes [8] to α -mixing processes with exponential decay and ψ -mixing with polynomial decay, since we can apply Theorem 2.4 to this encoder f , and then obtain information on the highest scoring V_n .

3. Shortest distance between observed orbits

In [14] it was explained that, in the case of dynamical systems, investigating the longest common substring is similar to the study of the shortest distance between orbits. Mixing this idea with the fact that studying statistical properties of observations of dynamical systems could be more significant for experimentalists (see, e.g., [15,34,35]), we will analyse in this section the behaviour of the shortest distance between two observed orbits.

Let (X, \mathcal{A}, T, μ) be a dynamical system where (X, d) is a metric space, \mathcal{A} is a σ -algebra on X , $T : X \rightarrow X$ is a measurable map and μ an invariant probability measure on (X, \mathcal{A}) that is, $\mu(T^{-1}(A)) = \mu(A)$, for all $A \in \mathcal{A}$.

Definition 3.1. Let $f : X \rightarrow Y \subset \mathbb{R}^N$ be a measurable function, called the observation. We define the shortest distance between two observed orbits as follows

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

For a measure ν on X , we define the lower and upper correlation dimension of ν by

$$\underline{C}_\nu = \liminf_{r \rightarrow 0} \frac{\log \int_X \nu(B(x, r)) d\nu(x)}{\log r} \quad \text{and} \quad \overline{C}_\nu = \limsup_{r \rightarrow 0} \frac{\log \int_X \nu(B(x, r)) d\nu(x)}{\log r}.$$

If the limit exists, we denote by C_ν the common value.

We will show that the shortest distance between two observed orbits is related with the correlation dimension of the pushforward measure $f_*\mu$. Recall that the pushforward measure is given by $f_*\mu(\cdot) := \mu(f^{-1}(\cdot))$.

Theorem 3.2. *Let (X, \mathcal{A}, T, μ) be a dynamical system. Consider an observation $f : X \rightarrow Y$ such that $\underline{C}_{f_*\mu} > 0$. Then for $\mu \otimes \mu$ -almost every $(x, y) \in X \times X$*

$$\overline{\lim}_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} \leq \frac{2}{\underline{C}_{f_*\mu}}. \tag{22}$$

We recall that the condition $\underline{C}_{f_*\mu} = 0$ can lead to unknown values for the above limit. However, one can observe that if $m_n^f = 0$ on a set of positive measure, our result implies immediately that $\underline{C}_{f_*\mu} = 0$. The following simple example illustrates this fact.

Example 3.3. Let $X = [0, 1]$ and $\mu = Leb$ the Lebesgue measure on X . Given $A \subset X$ with $\mu(A) > 0$ we define a function $f : X \rightarrow X$ by

$$f(x) = \begin{cases} x, & \text{if } x \in A^c \\ c, & \text{if } x \in A \end{cases}$$

where $c \in [0, 1]$ is a constant. For any transformation T which is μ -invariant, we have $m_n^f(x, y) = 0$ for every $x, y \in A$, and thus $\underline{C}_{f_*\mu} = 0$. One can also observe that if T is ergodic, for n sufficiently large $m_n^f(x, y) = 0$ for almost every x, y . Indeed, by Poincaré recurrence theorem, we obtain that, for almost every x, y , the orbits of x and y will visit A , i.e. it exist $n_1, n_2 \in \mathbb{N}$ such that $T^{n_1}(x) \in A$ and $T^{n_2}(y) \in A$. Therefore, for n sufficiently large $m_n^f(x, y) = d(f(T^{n_1}(x)), f(T^{n_2}(y))) = 0$. In fact, with a simple computation, one can show that $C_{f_*\mu} = 0$.

Proof of Theorem 3.2. For $\epsilon > 0$, we define

$$k_n = \frac{2 \log n + \log \log n}{\underline{C}_{f_*\mu} - \epsilon}.$$

We also define

$$A_{ij}^f(y) = T^{-i}[f^{-1}B(f(T^j y), e^{-k_n})]$$

and

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

Using Markov inequality, we get that

$$\mu \otimes \mu(\{(x, y) : m_n^f(x, y) < e^{-k_n}\}) = \mu \otimes \mu(\{(x, y) : S_n^f(x, y) > 0\}) \leq \mathbb{E}(S_n^f).$$

Using the invariance of μ , we can compute the expected value of S_n^f

$$\mathbb{E}(S_n^f) = n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y).$$

Thus, for n large enough, by definition of $\underline{C}_{f_*\mu}$ and k_n , we obtain

$$\mu \otimes \mu(\{(x, y) : m_n^f(x, y) \leq e^{-k_n}\}) \leq n^2 e^{-k_n(\underline{C}_{f_*\mu} - \epsilon)} = \frac{1}{\log n}.$$

Choosing a subsequence $\{n_\kappa\}_{\kappa \in \mathbb{N}}$ such that $n_\kappa = \lceil e^{\kappa^2} \rceil$, we can use Borel–Cantelli lemma as in the proof of (4) to obtain

$$\overline{\lim}_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} \leq \frac{2}{\underline{C}_{f_*\mu} - \epsilon}.$$

Since ϵ can be arbitrarily small, the proof is complete. □

As in [14], to obtain an equality in (22), we will need more assumptions on the system.

(H1) Let $\mathcal{H}^\alpha(X, \mathbb{R})$ be the space of Hölder observables. For all $\psi, \phi \in \mathcal{H}^\alpha(X, \mathbb{R})$ and for all $n \in \mathbb{N}^*$, we have:

$$\left| \int_X \psi \circ f(T^n x) \phi \circ f(x) d\mu(x) - \int_X \psi \circ f d\mu \int_X \phi \circ f d\mu \right| \leq \|\psi \circ f\|_\alpha \|\phi \circ f\|_\alpha \theta_n$$

with $\theta_n = a^n$ and $a \in [0, 1)$.

(HA) There exist $r_0 > 0$, $\xi \geq 0$ and $\beta > 0$ such that for $f_*\mu$ -almost every $y \in \mathbb{R}^N$ and any $r_0 > r > \rho > 0$,

$$f_*\mu(B(y, r + \rho) \setminus B(y, r - \rho)) \leq r^{-\xi} \rho^\beta.$$

One can observe that, if f is Lipschitz, assuming hypothesis (H1) is weaker than assuming an exponential decay of correlations (for Hölder observables) for the system (X, \mathcal{A}, T, μ) . Indeed, note that if f is Lipschitz then $\psi \circ f$ is Hölder for every Hölder function ψ .

Theorem 3.4. *Let (X, \mathcal{A}, T, μ) be a dynamical system and consider a Lipschitz observation $f : X \rightarrow Y$ such that $\underline{C}_{f_*\mu} > 0$. If the system satisfies (H1) and (HA), then for $\mu \otimes \mu$ -almost every $(x, y) \in X \times X$*

$$\underline{\lim}_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} \geq \frac{2}{C_{f_*\mu}}.$$

Furthermore, if $C_{f_*\mu}$ exists, we get

$$\lim_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} = \frac{2}{C_{f_*\mu}} \tag{★★}$$

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

In what follows one will observe that the proof follows the lines of the symbolic case where M_n^f will be substitute by $-\log m_n^f$ and cylinders of size k will be substitute by balls of radius

e^{-k} . Thus, we will only write the main lines of the proof, giving more details when the proof diverge from the symbolic one.

To prove Theorem 3.4, the main difficulty and difference with the symbolic case is that we cannot apply mixing as simply. In particular, we can only apply mixing to Hölder observables and indicator functions are not even continuous. To overthrow this difficulty, we will first prove in the following lemma that a particular function is Hölder. In the proof of Theorem 3.4, we will apply the mixing property to this particular function.

Lemma 3.5. *Let (X, \mathcal{A}, T, μ) be a dynamical system with observation f . If it satisfies (HA), then there exist $0 < r_0 < 1$, $c \geq 0$ and $\zeta \geq 0$ such that for any $0 < r < r_0$, the function $\psi_1 : x \mapsto f_*\mu(B(x, r))$ belongs to $\mathcal{H}^\alpha(X, \mathbb{R})$ and*

$$\|\psi_1\|_\alpha \leq 2r^{-\zeta}.$$

Proof. Let $x, y \in X$ and $0 < r < r_0$, if $\|x - y\| < r$ we have

$$\|f_*\mu(B(y, r)) - f_*\mu(B(x, r))\| \leq f_*\mu(B(x, r + \|x - y\|)) - f_*\mu(B(x, r - \|x - y\|)).$$

Thus, by (HA),

$$\|f_*\mu(B(y, r)) - f_*\mu(B(x, r))\| \leq r^{-\xi} \|x - y\|^\beta.$$

On the other hand, if $\|x - y\| \geq r$ then

$$\|f_*\mu(B(y, r)) - f_*\mu(B(x, r))\| \leq 2 \leq \frac{2}{r} \|x - y\|.$$

Thus, ψ_1 is Hölder and $\|\psi_1\|_\alpha \leq 2r^{-\zeta}$ with $\zeta = \max\{1, \xi\}$. □

In the sequel, we present the proof of Theorem 3.4. This proof mainly follows the ideas of the proof of [14], Theorem 5.

Proof of Theorem 3.4. Without loss of generality, we will assume here that $\theta_n = e^{-n}$. Let $b < -4$. Given $\epsilon > 0$, we define

$$k_n = \frac{2 \log n + b \log \log n}{\bar{C}_{f_*\mu} + \epsilon}.$$

By Chebyshev's inequality, we get that

$$\mu \otimes \mu(\{(x, y) : m_n^f(x, y) \geq e^{-k_n}\}) \leq \frac{\text{var}(S_n^f)}{\mathbb{E}(S_n^f)^2}.$$

We now proceed to estimate the variance of S_n^f .

We see at once that

$$\begin{aligned} \text{var}(S_n^f) &= \sum_{1 \leq i, i', j, j' \leq n} \text{cov}(\mathbb{1}_{A_{ij}^f}, \mathbb{1}_{A_{i'j'}^f}) \\ &= \sum_{1 \leq i, i', j, j' \leq n} \int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(T^i x) \mathbb{1}_{f^{-1}B(f(T^{j'} y), e^{-kn})}(T^{i'} x) \\ &\quad - n^4 \left(\int f_* \mu(B(f(y), e^{-kn})) d\mu(y) \right)^2. \end{aligned} \tag{23}$$

One can observe that this equation is similar to (8), thus as in the symbolic case we would like to apply the mixing property to estimate the previous sum. However, in this case our assumption (H1) only allows us to use mixing with Hölder functions thus we will approximate our characteristic functions by Lipschitz (and thus Hölder) functions following the construction of the proof of Lemma 9 in [36].

Let $\rho > 0$ (to be defined properly later). Let $\eta_{e^{-kn}} : [0, \infty) \rightarrow \mathbb{R}$ be the $\frac{1}{\rho e^{-kn}}$ -Lipschitz function such that $\mathbb{1}_{[0, e^{-kn}]} \leq \eta_{e^{-kn}} \leq \mathbb{1}_{[0, (1+\rho)e^{-kn}]}$ and set $\varphi_{f(y), e^{-kn}}(x) = \eta_{e^{-kn}}(d(f(y), x))$. Since f is L -Lipschitz it follows that $\varphi_{f(y), e^{-kn}} \circ f$ is $\frac{L}{\rho e^{-kn}}$ -Lipschitz. Moreover, we have

$$\begin{aligned} \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(x) &= \mathbb{1}_{B(f(T^j y), e^{-kn})}(f(x)) \\ &= \mathbb{1}_{[0, e^{-kn}]}(d(f(T^j y), f(x))) \\ &\leq \eta_{e^{-kn}}(d(f(T^j y), f(x))) \\ &= \varphi_{f(T^j y), e^{-kn}}(f(x)). \end{aligned} \tag{24}$$

We are now able to apply the mixing property and as in the symbolic case, we will consider four different cases. Let us fix $g = g(n) = \log(n^\gamma)$ for some $\gamma > 0$ to be defined later.

Case 1: $|i - i'| > g$. By (H1) and (24), we obtain

$$\begin{aligned} &\int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(T^i x) \mathbb{1}_{f^{-1}B(f(T^{j'} y), e^{-kn})}(T^{i'} x) d\mu(x) d\mu(y) \\ &= \int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(T^{i-i'} x) \mathbb{1}_{f^{-1}B(f(T^{j'} y), e^{-kn})}(x) d\mu(x) d\mu(y) \\ &\leq \int \int \varphi_{f(T^j y), e^{-kn}}(f(T^{i-i'} x)) \varphi_{f(T^{j'} y), e^{-kn}}(f(x)) d\mu(x) d\mu(y) \\ &\leq \int \left(\int \varphi_{f(T^j y), e^{-kn}}(f(T^{i-i'} x)) d\mu(x) \int \varphi_{f(T^{j'} y), e^{-kn}}(f(x)) d\mu(x) \right) d\mu(y) \\ &\quad + \theta_g \|\varphi_{f(T^j y), e^{-kn}}\| \|\varphi_{f(T^{j'} y), e^{-kn}}\| \\ &\leq \frac{L^2}{\rho^2 e^{-2kn}} \theta_g + \int f_* \mu(B(f(T^j y), (1 + \rho)e^{-kn})) f_* \mu(B(f(T^{j'} y), (1 + \rho)e^{-kn})) d\mu(y). \end{aligned}$$

This estimate is similar to (9), however one needs to take extra care to estimate the second part (and not use mixing immediately as in the discrete case) since the radius of the balls is not e^{-k_n} anymore. Indeed, we need the radius to be e^{-k_n} so that when we will use mixing again we will obtain a term which will simplify with the last term in (8). To do so, we can observe that using (HA) we obtain

$$\begin{aligned} & \int f_*\mu(B(f(T^j y), (1 + \rho)e^{-k_n}))f_*\mu(B(f(T^{j'} y), (1 + \rho)e^{-k_n})) d\mu(y) \\ & \quad - \int f_*\mu(B(f(T^j y), e^{-k_n}))f_*\mu(B(f(T^{j'} y), e^{-k_n})) d\mu(y) \\ & \leq \int f_*\mu(B(f(T^j y), (1 + \rho)e^{-k_n}))(f_*\mu(B(f(T^{j'} y), (1 + \rho)e^{-k_n})) \\ & \quad - f_*\mu(B(f(T^{j'} y), e^{-k_n}))) d\mu(y) \\ & \quad + \int f_*\mu(B(f(T^{j'} y), e^{-k_n}))(f_*\mu(B(f(T^j y), (1 + \rho)e^{-k_n})) \\ & \quad - f_*\mu(B(f(T^j y), e^{-k_n}))) d\mu(y) \\ & \leq \int f_*\mu(B(f(T^j y), (1 + \rho)e^{-k_n}))e^{\xi k_n} \rho^\beta d\mu(y) \\ & \quad + \int f_*\mu(B(f(T^{j'} y), e^{-k_n}))e^{\xi k_n} \rho^\beta d\mu(y). \end{aligned}$$

Therefore, choosing $\rho = n^{-\delta}$ for some $\delta > 0$ to be defined later, we have for n large enough

$$\begin{aligned} & \int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-k_n})}(T^i x) \mathbb{1}_{f^{-1}B(f(T^{j'} y), e^{-k_n})}(T^{i'} x) d\mu(x) d\mu(y) \\ & \leq \frac{L^2}{\rho^2 e^{-2k_n}} \theta_g + 2e^{\xi k_n} \rho^\beta \int f_*\mu(B(f(T^j y), 2e^{-k_n})) d\mu(y) \\ & \quad + \int f_*\mu(B(f(T^j y), e^{-k_n}))f_*\mu(B(f(T^{j'} y), e^{-k_n})) d\mu(y). \end{aligned}$$

One can observe that in this estimate we have now an additional term that was not present in the symbolic setting (9) which is due to the need to approximate characteristic functions by Lipschitz functions. To deal with the third term of the last inequality, we need to consider two different cases.

Case 1.1: $|j - j'| > g$. We can use the mixing property (H1) to the particular function defined in Lemma 3.5

$$\begin{aligned} & \int f_*\mu(B(f(T^j y), e^{-k_n}))f_*\mu(B(f(T^{j'} y), e^{-k_n})) d\mu(y) \\ & \leq 4\theta_g e^{2\xi k_n} + \left(\int f_*\mu(B(f(y), e^{-k_n})) d\mu(y) \right)^2 \end{aligned}$$

and we obtain an estimate similar to (10).

Case 1.2: $|j - j'| \leq g$. Using Holder’s inequality together and the invariance of μ , as in (11), we have

$$\begin{aligned} & \int f_*\mu(B(f(T^j y), e^{-kn}))f_*\mu(B(f(T^{j'} y), e^{-kn})) d\mu(y) \\ & \leq \int f_*\mu(B(f(y), e^{-kn}))^2 d\mu(y). \end{aligned}$$

Case 2.1: $|i - i'| \leq g$ and $|j - j'| > g$. In this case, we obtain the same estimate as in the Case 1.2 using the following symmetry:

$$\mathbb{1}_{f^{-1}B(f(T^\ell y), e^{-kn})}(T^m x) = \mathbb{1}_{f^{-1}B(f(T^m x), e^{-kn})}(T^\ell y)$$

for all $\ell, m \in \mathbb{N}$ and all x and y .

Case 2.2: $|i - i'| \leq g$ and $|j - j'| \leq g$. The boundedness of the indicator function and invariance of μ give that,

$$\begin{aligned} & \int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(T^i x)\mathbb{1}_{f^{-1}B(f(T^{j'} y), e^{-kn})}(T^{i'} x) d\mu(x) d\mu(y) \\ & \leq \int \int \mathbb{1}_{f^{-1}B(f(T^j y), e^{-kn})}(T^i x) d\mu(x) d\mu(y) = \int f_*\mu(B(f(y), e^{-kn})) d\mu(y). \end{aligned}$$

Putting all the previous estimates in (23), we obtain

$$\begin{aligned} \frac{\text{var}(S_n^f)}{\mathbb{E}(S_n^f)^2} & \leq \frac{n^4 L^2 \rho^{-2} e^{2k_n \theta_g} + 4n^4 \theta_g e^{2\zeta k_n} + 2n^4 e^{\xi k_n} \rho^\beta \int f_*\mu(B(f(y), 2e^{-k_n})) d\mu(y)}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} \\ & \quad + \frac{4n^2 g^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y) + 4n^3 g \int f_*\mu(B(f(y), e^{-k_n}))^2 d\mu(y)}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2}. \end{aligned} \tag{25}$$

This estimate is comparable to (13) (in the symbolic setting), except for the third term coming from our approximation of characteristic functions, and the terms will be dealt with in a similar way. To help the reader understanding the following majorations, we can observe that

$$n^{-2}(\log n)^{-b} = e^{-k_n(\overline{C}_{f_*\mu} + \epsilon)} \lesssim \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y) \leq 1.$$

Recalling that $\rho = n^{-\delta}$, we can choose δ large enough (depending on $\xi, \beta, \overline{C}_{f_*\mu}, b$ and ϵ) so that

$$\frac{2n^4 e^{\xi k_n} \rho^\beta \int f_*\mu(B(f(y), 2e^{-k_n})) d\mu(y)}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} \leq \frac{1}{n}. \tag{26}$$

Recalling that $g = \log(n^\gamma)$, we can observe that we can choose γ large enough (depending on $\delta, \bar{C}_{f_*\mu}, \zeta, b$ and ϵ) so that

$$\frac{n^4 L^2 \rho^{-2} e^{2k_n \theta_g}}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} \leq \frac{1}{n} \tag{27}$$

and so that

$$\frac{4n^4 \theta_g e^{2\zeta k_n}}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} \leq \frac{1}{n}. \tag{28}$$

For the fourth term, we have

$$\begin{aligned} \frac{4n^2 g^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y)}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} &\leq \frac{4g^2}{n^2 e^{-k_n(\bar{C}_{f_*\mu} + \epsilon)}} \\ &\leq 4\gamma^2 (\log n)^{2+b}. \end{aligned} \tag{29}$$

To estimate the last term, one cannot use immediately the subadditivity as in (11), thus we will use the following lemma.

Lemma 3.6 (Lemma 14 [14]). *Let $Z \subset \mathbb{R}^N$ and let ν be a probability measure on Z . There exists a constant $K > 0$ depending only on N such that for every r small enough*

$$\int_Z \nu(B(y, r))^2 d\nu(y) \leq K \left(\int_Z \nu(B(y, r)) d\nu(y) \right)^{3/2}.$$

Applying the previous lemma with $Z = Y$ and $\nu = f_*\mu$ we obtain

$$\begin{aligned} \frac{4n^3 g \int f_*\mu(B(f(y), e^{-k_n}))^2 d\mu(y)}{(n^2 \int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^2} &\leq \frac{4gK}{n(\int f_*\mu(B(f(y), e^{-k_n})) d\mu(y))^{1/2}} \\ &\leq \frac{4gK}{n} e^{\frac{k_n(\bar{C}_{f_*\mu} + \epsilon)}{2}} \\ &\leq 4K\gamma (\log n)^{1+\frac{b}{2}}. \end{aligned} \tag{30}$$

Since $b < -4$ and substituting (27), (26), (28), (29) and (30) into (25), we get

$$\mu \otimes \mu(\{(x, y) : m_n^f(x, y) \geq e^{-k_n}\}) \leq \mathcal{O}((\log n)^{-1}).$$

Thus, taking a subsequence $n_\kappa = \lceil e^{k^2} \rceil$. As in the proof of Theorem 3.2, by Borel–Cantelli lemma we obtain

$$\varliminf_{n \rightarrow \infty} \frac{\log m_n^f(x, y)}{-\log n} = \varliminf_{\kappa \rightarrow \infty} \frac{\log m_{n_\kappa}^f(x, y)}{-\log n_\kappa} \geq \frac{2}{\underline{C}_{f_*\mu} + \epsilon}.$$

Since ϵ can be arbitrarily small, the theorem follows. □

Following the idea of [34,35] that the study of observation of dynamical systems can be used to study random dynamical systems, we will show in the next section that the previous result can be applied to obtain information on the shortest distance between two random orbits.

4. Shortest distance between orbits for random dynamical systems

Let $X \subset \mathbb{R}^N$ and let $(\Omega, \theta, \mathbb{P})$ be a probability measure preserving system, where Ω is a metric space and $B(\Omega)$ its Borelian σ -algebra. We first introduce the notion of random dynamical system.

Definition 4.1. A random dynamical system $\mathcal{T} = (T_\omega)_{\omega \in \Omega}$ on X over $(\Omega, B(\Omega), \mathbb{P}, \theta)$ is generated by maps T_ω such that $(\omega, x) \mapsto T_\omega(x)$ is measurable and satisfies:

$$T_\omega^0 = \text{Id} \quad \text{for all } \omega \in \Omega,$$

$$T_\omega^n = T_{\theta^{n-1}(\omega)} \circ \dots \circ T_{\theta(\omega)} \circ T_\omega \quad \text{for all } n \geq 1.$$

The map $S : \Omega \times X \rightarrow \Omega \times X$ defined by $S(\omega, x) = (\theta(\omega), T_\omega(x))$ is the dynamics of the random dynamical systems generated by \mathcal{T} and is called skew-product.

Definition 4.2. A probability measure μ is said to be an invariant measure for the random dynamical system \mathcal{T} if it satisfies

1. μ is S -invariant
2. $\pi_*\mu = \mathbb{P}$

where $\pi : \Omega \times X \rightarrow \Omega$ is the canonical projection.

Let $(\mu_\omega)_\omega$ denote the decomposition of μ on X , that is, $d\mu(\omega, x) = d\mu_\omega(x) d\mathbb{P}(\omega)$. We denote by $\nu = \int \mu_\omega d\mathbb{P}$ the marginal of ν on X .

Definition 4.3. We define the shortest distance between two random orbits by

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

As in the deterministic case, we need a hypothesis for the measure and an (annealed) exponential decay of correlations for the random dynamical system. Namely,

- (a) There exist $r_0 > 0$, $\xi \leq 0$ and $\beta > 0$ such that for almost every $y \in X$ and any $r_0 > r > \rho > 0$,

$$\nu(B(y, r + \rho) \setminus B(y, r - \rho)) \leq r^{-\xi} \rho^\beta.$$

(b) (Annealed decay of correlations) $\forall n \in \mathbb{N}^*$, ψ and ϕ Hölder observables from X to \mathbb{R} ,

$$\left| \int_{\Omega \times X} \psi(T_\omega^n(x))\phi(x) d\mu(\omega, x) - \int_{\Omega \times X} \psi d\mu \int_{\Omega \times X} \phi d\mu \right| \leq \|\psi\|_\alpha \|\phi\|_\alpha \theta_n$$

with $\theta_n = e^{-n}$.

Theorem 4.4. *Let \mathcal{T} be a random dynamical system on X over $(\Omega, B(\Omega), \mathbb{P}, \theta)$ with an invariant measure μ such that $\underline{C}_v > 0$. Then for $\mu \otimes \mu$ -almost every $(\omega, x, \tilde{\omega}, \tilde{x}) \in \Omega \times X \times \Omega \times X$,*

$$\overline{\lim}_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} \leq \frac{2}{\underline{C}_v}.$$

Moreover, if the random dynamical system satisfies assumptions (a) and (b), then

$$\underline{\lim}_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} \geq \frac{2}{\underline{C}_v}$$

and if C_v exists, then

$$\lim_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} = \frac{2}{C_v}. \tag{***}$$

To prove this theorem, we will just apply Theorem 3.2 and Theorem 3.4 to the skew-product S with a well-chosen observation, following the idea given in [34].

Proof. We use Theorem 3.2 and Theorem 3.4 for the dynamical system $(\Omega \times X, B(\Omega \times X), \mu, S)$ with the observation f defined by

$$f : \Omega \times X \rightarrow X$$

$$(\omega, x) \mapsto x.$$

Indeed, with this particular observation f , studying the observed orbit of (ω, x) under the skew-product S is similar to studying the random orbit of x with respect to ω since

$$f(S^n(\omega, x)) = f(\theta^n \omega, T_\omega^n(x)) = T_\omega^n(x).$$

Thus, for all z and $t \in \Omega \times X$ we can link the shortest distance between two observed orbits and the shortest distance between two random orbits. Set $z = (\omega, x)$ and $t = (\tilde{\omega}, \tilde{x})$ then

$$\begin{aligned} m_n^f(z, t) &= \min_{i, j=0, \dots, n-1} (d(f(S^i(\omega, x)), f(S^j(\tilde{\omega}, \tilde{x})))) \\ &= \min_{i, j=0, \dots, n-1} (d(T_\omega^i x, T_{\tilde{\omega}}^j \tilde{x})) \\ &= m_n^{\omega, \tilde{\omega}}(x, \tilde{x}). \end{aligned}$$

Moreover, we can identify the pushforward measure: $f_*\mu = \nu$. Therefore, in view of the lower and upper correlation dimensions, the following statement finishes the proof

$$\underline{C}_{f_*\mu} = \underline{C}_\nu \quad \text{and} \quad \overline{C}_{f_*\mu} = \overline{C}_\nu. \quad \square$$

In what follows, we present a collection of examples for which Theorem 4.4 holds. They illustrate some well-known random dynamical systems on the literature.

4.1. Non-i.i.d. random dynamical system

The first example is a non-i.i.d. random dynamical system for which it was computed recurrence rates in [30] and hitting times statistics in [34].

Consider the two linear maps which preserve Lebesgue measure Leb on $X = \mathbb{T}^1$, the one-dimensional torus:

$$\begin{aligned} T_1 : X &\rightarrow X & \text{and} & & T_2 : X &\rightarrow X \\ x &\mapsto 2x & & & x &\mapsto 3x. \end{aligned}$$

The following skew product gives the dynamics of the random dynamical system:

$$\begin{aligned} S : \Omega \times X &\rightarrow \Omega \times X \\ (\omega, x) &\mapsto (\theta(\omega), T_\omega(x)) \end{aligned}$$

with $\Omega = [0, 1]$, $T_\omega = T_1$ if $\omega \in [0, 2/5)$ and $T_\omega = T_2$ if $\omega \in [2/5, 1]$ where ω is the following piecewise linear map:

$$\theta(\omega) = \begin{cases} 2\omega & \text{if } \omega \in [0, 1/5) \\ 3\omega - 1/5 & \text{if } \omega \in [1/5, 2/5) \\ 2\omega - 4/5 & \text{if } \omega \in [2/5, 3/5) \\ 3\omega/2 - 1/2 & \text{if } \omega \in [3/5, 1]. \end{cases}$$

Note that the random orbit is constructed by choosing one of these two maps following a Markov process with the stochastic matrix

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}.$$

The associated skew-product S is $Leb \otimes Leb$ -invariant. It is easy to check that Lebesgue measure satisfies (a). Moreover, by [12] the skew product S has an exponential decay of correlations. Since in this example $\nu = Leb$, we have $C_\nu = 1$ and Theorem 4.4 implies that for $Leb \otimes Leb \otimes Leb \otimes Leb$ -almost every $(\omega, x, \tilde{\omega}, \tilde{x}) \in [0, 1] \times \mathbb{T}^1 \times [0, 1] \times \mathbb{T}^1$,

$$\lim_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} = 2.$$

4.2. Randomly perturbed dynamical systems

Consider a deterministic dynamical system (X, T, μ) where X is a compact Riemannian manifold, T is a map and μ is a T -invariant probability measure. We will present a random dynamical system constructed by perturbing the map T with a random additive noise. For $\epsilon > 0$, set $\Lambda_\epsilon = B(0, \epsilon)$ and let \mathcal{P}_ϵ be a probability measure on Λ_ϵ . For each $\omega \in \Lambda_\epsilon$, we denote the family of transformations $\{T_\omega\}_\omega$ where the map $T_\omega : X \rightarrow X$ are given by

$$T_\omega(x) = T(x) + \omega.$$

Denote \mathcal{T} the i.i.d. dynamical system on X over $(\Lambda_\epsilon^\mathbb{N}, \mathcal{P}_\epsilon^\mathbb{N}, \sigma)$. In the case where $X = \mathbb{T}^d$, for some expanding and piecewise expanding maps, if ϵ is sufficiently small, it was proved (see, e.g., [10,13,40]) that the random dynamical system has a stationary measure μ_ϵ absolutely continuous with respect to Lebesgue measure with density h_ϵ such that $0 < \underline{h}_\epsilon \leq h_\epsilon \leq \bar{h}_\epsilon < \infty$ and the system has an exponential decay of correlations. Thus, since the assumptions (a) and (b) are satisfied one can apply Theorem 4.4 and obtain information on the behavior of the shortest distance $m_n^{\omega, \hat{\omega}}$.

4.3. Random hyperbolic toral automorphisms

A linear toral automorphism is a map $T : \mathbb{T}^2 \rightarrow \mathbb{T}^2$ defined by the matrix action $x \mapsto Ax$, where the matrix A has integer entries and $\det A = \pm 1$. We say that T is hyperbolic if A has eigenvalues with modulus different from 1. For more simplicity, we will use the notation A for both the matrix and the associated automorphism.

For an hyperbolic toral automorphism A , we denote E_u^A the subspace spanned by e_u^A , the eigenvector associated to the eigenvalue whose absolute value is greater than 1 and we denote E_s^A the subspace spanned by e_s^A , the eigenvector associated to the eigenvalue whose absolute value is less than 1.

Following the definition from [11], we say that a pair (A_0, A_1) of hyperbolic toral automorphisms has the cone property if there exists an expansion cone \mathcal{E} such that

1. $A_i \mathcal{E} \subset \mathcal{E}$,
2. there exists $\lambda_\mathcal{E} > 1$ such that $|A_i x| \geq \lambda_\mathcal{E} |x|$ for $x \in \mathcal{E}$,
3. $E_u^{A_i} \cap \partial \mathcal{E} = 0$, where $\partial \mathcal{E}$ denote the boundary of \mathcal{E} ,

and there exists a contraction cone \mathcal{C} such that $\mathcal{C} \cap \mathcal{E} = 0$ and

1. $A_i^{-1} \mathcal{C} \subset \mathcal{C}$,
2. there exists $\lambda_\mathcal{C} < 1$ such that $|A_i^{-1} x| \geq \lambda_\mathcal{C}^{-1} |x|$ for $x \in \mathcal{C}$,
3. $E_s^{A_i} \cap \partial \mathcal{C} = 0$.

One can observe that for example a pair of hyperbolic toral automorphisms with positive entries, or a pair of hyperbolic toral automorphisms with negative entries, has the cone property.

Let $\Lambda = \{0, 1\}$ and $\theta = \sigma$ be the left shift on $\Lambda^\mathbb{N}$. Let A_0, A_1 two hyperbolic automorphisms satisfying the cone property. Let A_0 be chosen with a probability q and A_1 with a probability $1 - q$, i.e. $\mathbb{P} = \mathcal{P}^\mathbb{N}$ with $\mathcal{P}(0) = q$ and $\mathcal{P}(1) = 1 - q$.

Then, for the i.i.d. random dynamical system on \mathbb{T}^2 over $(\Lambda^{\mathbb{N}}, \mathcal{P}^{\mathbb{N}}, \sigma)$, the Lebesgue measure is stationary (and thus hypothesis (a) is satisfied) and the system has an exponential decay of correlations (see [11]).

Note that $\nu = \text{Leb} \otimes \text{Leb}$ implies that $C_\nu = 2$. Then, by Theorem 4.4 we get for $\mathbb{P} \otimes \text{Leb} \otimes \mathbb{P} \otimes \text{Leb}$ -almost every $(\omega, x, \tilde{\omega}, \tilde{x}) \in \Omega \times \mathbb{T}^2 \times \Omega \times \mathbb{T}^2$,

$$\lim_{n \rightarrow \infty} \frac{\log m_n^{\omega, \tilde{\omega}}(x, \tilde{x})}{-\log n} = 1.$$

Acknowledgements

This work is partially supported by FAPESB and CNPQ (grants 406324/2017-4, 439422/2018-3 and 435470/2018-3). AC is partially supported by CAPES. RL is partially supported by FAPESP (grant 2014/19805-1). JR was partially supported by FCT project PTDC/MAT-PUR/28177/2017, with national funds, and by CMUP (UID/MAT/00144/2019), which is funded by FCT with national (MCTES) and European structural funds through the programs FEDER, under the partnership agreement PT2020. This work is part of the Universal FAPEMIG project “Dinâmica de recorrência para shifts aleatórios, e processos com lapsos de memória” (process APQ-00987-18). The authors would like to thank the anonymous referee for helpful comments, corrections and suggestions.

References

- [1] Abadi, M., Gallo, S. and Rada-Mora, E.A. (2018). The shortest possible return time of β -mixing processes. *IEEE Trans. Inform. Theory* **64** 4895–4906. MR3819346 <https://doi.org/10.1109/tit.2017.2757494>
- [2] Abadi, M. and Galves, A. (2001). Inequalities for the occurrence times of rare events in mixing processes. The state of the art. *Markov Process. Related Fields* **7** 97–112. MR1835750
- [3] Abadi, M. and Lambert, R. (2013). The distribution of the short-return function. *Nonlinearity* **26** 1143–1162. MR3043376 <https://doi.org/10.1088/0951-7715/26/5/1143>
- [4] Abadi, M. and Lambert, R. (2019). From the divergence between two measures to the shortest path between two observables. *Ergodic Theory Dynam. Systems* **39** 1729–1744. MR3961671 <https://doi.org/10.1017/etds.2017.114>
- [5] Abadi, M. and Vaienti, S. (2008). Large deviations for short recurrence. *Discrete Contin. Dyn. Syst.* **21** 729–747. MR2399435 <https://doi.org/10.3934/dcds.2008.21.729>
- [6] Abadi, M.N. and Cardeño, L. (2015). Rényi entropies and large deviations for the first match function. *IEEE Trans. Inform. Theory* **61** 1629–1639. MR3332970 <https://doi.org/10.1109/TIT.2015.2406695>
- [7] Afraimovich, V., Chazottes, J.R. and Saussol, B. (2003). Pointwise dimensions for Poincaré recurrences associated with maps and special flows. *Discrete Contin. Dyn. Syst.* **9** 263–280. MR1952373 <https://doi.org/10.3934/dcds.2003.9.263>
- [8] Arratia, R., Morris, P. and Waterman, M.S. (1988). Stochastic Scrabble: Large deviations for sequences with scores. *J. Appl. Probab.* **25** 106–119. MR0929509 <https://doi.org/10.1017/s0021900200040687>
- [9] Arratia, R. and Waterman, M.S. (1985). An Erdős–Rényi law with shifts. *Adv. Math.* **55** 13–23. MR0772069 [https://doi.org/10.1016/0001-8708\(85\)90003-9](https://doi.org/10.1016/0001-8708(85)90003-9)

- [10] Aytaç, H., Freitas, J.M. and Vienti, S. (2015). Laws of rare events for deterministic and random dynamical systems. *Trans. Amer. Math. Soc.* **367** 8229–8278. MR3391915 <https://doi.org/10.1090/S0002-9947-2014-06300-9>
- [11] Ayyer, A. and Stenlund, M. (2007). Exponential decay of correlations for randomly chosen hyperbolic toral automorphisms. *Chaos* **17** 043116, 7. MR2380035 <https://doi.org/10.1063/1.2785145>
- [12] Baladi, V. (2000). *Positive Transfer Operators and Decay of Correlations*. *Advanced Series in Non-linear Dynamics* **16**. River Edge, NJ: World Scientific Co., Inc. MR1793194 <https://doi.org/10.1142/9789812813633>
- [13] Baladi, V. and Young, L.-S. (1993). On the spectra of randomly perturbed expanding maps. *Comm. Math. Phys.* **156** 355–385. MR1233850
- [14] Barros, V., Liao, L. and Rousseau, J. (2019). On the shortest distance between orbits and the longest common substring problem. *Adv. Math.* **344** 311–339. MR3897435 <https://doi.org/10.1016/j.aim.2019.01.001>
- [15] Boshernitzan, M.D. (1993). Quantitative recurrence results. *Invent. Math.* **113** 617–631. MR1231839 <https://doi.org/10.1007/BF01244320>
- [16] Collet, P., Galves, A. and Leonardi, F. (2008). Random perturbations of stochastic processes with unbounded variable length memory. *Electron. J. Probab.* **13** 1345–1361. MR2438809 <https://doi.org/10.1214/EJP.v13-538>
- [17] Dembo, A., Karlin, S. and Zeitouni, O. (1994). Critical phenomena for sequence matching with scoring. *Ann. Probab.* **22** 1993–2021. MR1331213
- [18] Dembo, A., Karlin, S. and Zeitouni, O. (1994). Limit distribution of maximal non-aligned two-sequence segmental score. *Ann. Probab.* **22** 2022–2039. MR1331214
- [19] Dembo, A. and Kontoyiannis, I. (1999). The asymptotics of waiting times between stationary processes, allowing distortion. *Ann. Appl. Probab.* **9** 413–429. MR1687410 <https://doi.org/10.1214/aoap/1029962749>
- [20] Ferrari, P.A. and Galves, A. (1997). *Acomplamento e Processos Estocásticos*. 21^o Colóquio Brasileiro de Matemática. [21st Brazilian Mathematics Colloquium]. Rio de Janeiro: Instituto de Matemática Pura e Aplicada (IMPA). MR1718548
- [21] Garcia, N.L. and Moreira, L. (2015). Stochastically perturbed chains of variable memory. *J. Stat. Phys.* **159** 1107–1126. MR3345412 <https://doi.org/10.1007/s10955-015-1227-8>
- [22] Grzegorek, P. and Kupsa, M. (2009). Return times in a process generated by a typical partition. *Non-linearity* **22** 371–379. MR2475551 <https://doi.org/10.1088/0951-7715/22/2/007>
- [23] Haydn, N. and Vienti, S. (2010). The Rényi entropy function and the large deviation of short return times. *Ergodic Theory Dynam. Systems* **30** 159–179. MR2586350 <https://doi.org/10.1017/S0143385709000030>
- [24] Haydn, N.T.A. (2013). Entry and return times distribution. *Dyn. Syst.* **28** 333–353. MR3170620 <https://doi.org/10.1080/14689367.2013.822459>
- [25] Kelbert, M. and Suhov, Y. (2013). *Information Theory and Coding by Example*. Cambridge: Cambridge Univ. Press. MR3137525 <https://doi.org/10.1017/CBO9781139028448>
- [26] Kontoyiannis, I., Algoet, P.H., Suhov, Yu.M. and Wyner, A.J. (1998). Nonparametric entropy estimation for stationary processes and random fields, with applications to English text. *IEEE Trans. Inform. Theory* **44** 1319–1327. MR1616653 <https://doi.org/10.1109/18.669425>
- [27] Levin, D.A., Peres, Y. and Wilmer, E.L. (2009). *Markov Chains and Mixing Times*. Providence, RI: Amer. Math. Soc. MR2466937
- [28] Łuczak, T. and Szpankowski, W. (1997). A suboptimal lossy data compression based on approximate pattern matching. *IEEE Trans. Inform. Theory* **43** 1439–1451. MR1476778 <https://doi.org/10.1109/18.623143>

- [29] Månsson, M. (2000). On compound Poisson approximation for sequence matching. *Combin. Probab. Comput.* **9** 529–548. MR1816101 <https://doi.org/10.1017/S096354830000448X>
- [30] Marie, P. and Rousseau, J. (2011). Recurrence for random dynamical systems. *Discrete Contin. Dyn. Syst.* **30** 1–16. MR2773129 <https://doi.org/10.3934/dcds.2011.30.1>
- [31] Neuhauser, C. (1996). A phase transition for the distribution of matching blocks. *Combin. Probab. Comput.* **5** 139–159. MR1400960 <https://doi.org/10.1017/S0963548300001930>
- [32] Ornstein, D.S. and Weiss, B. (1993). Entropy and data compression schemes. *IEEE Trans. Inform. Theory* **39** 78–83. MR1211492 <https://doi.org/10.1109/18.179344>
- [33] Rousseau, J. Longest common substring for random subshifts of finite type. Preprint. Available at [arXiv:1905.08131](https://arxiv.org/abs/1905.08131).
- [34] Rousseau, J. (2014). Hitting time statistics for observations of dynamical systems. *Nonlinearity* **27** 2377–2392. MR3266858 <https://doi.org/10.1088/0951-7715/27/9/2377>
- [35] Rousseau, J. and Saussol, B. (2010). Poincaré recurrence for observations. *Trans. Amer. Math. Soc.* **362** 5845–5859. MR2661498 <https://doi.org/10.1090/S0002-9947-2010-05078-0>
- [36] Saussol, B. (2006). Recurrence rate in rapidly mixing dynamical systems. *Discrete Contin. Dyn. Syst.* **15** 259–267. MR2191396 <https://doi.org/10.3934/dcds.2006.15.259>
- [37] Saussol, B. (2009). An introduction to quantitative Poincaré recurrence in dynamical systems. *Rev. Math. Phys.* **21** 949–979. MR2568049 <https://doi.org/10.1142/S0129055X090003785>
- [38] Saussol, B., Troubetzkoy, S. and Vaienti, S. (2002). Recurrence, dimensions, and Lyapunov exponents. *J. Stat. Phys.* **106** 623–634. MR1884547 <https://doi.org/10.1023/A:1013710422755>
- [39] Shields, P.C. (1996). *The Ergodic Theory of Discrete Sample Paths*. Graduate Studies in Mathematics **13**. Providence, RI: Amer. Math. Soc. MR1400225 <https://doi.org/10.1090/gsm/013>
- [40] Viana, M. (1997). Stochastic dynamics of deterministic systems. *Brazilian Math. Colloquium, IMPA*.
- [41] Wyner, A.D. and Ziv, J. (1989). Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory* **35** 1250–1258. MR1036629 <https://doi.org/10.1109/18.45281>
- [42] Wyner, A.J. (1999). More on recurrence and waiting times. *Ann. Appl. Probab.* **9** 780–796. MR1722282 <https://doi.org/10.1214/aoap/1029962813>

Received April 2019 and revised October 2019