Ergod. Th. & Dynam. Sys. (First published online 2017), page 1 of 16*
doi:10.1017/etds.2017.114

*Provisional—final page numbers to be inserted when paper edition is published

From the divergence between two measures to the shortest path between two observables

MIGUEL ABADI† and RODRIGO LAMBERT‡

† *Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, CEP 05508-090, São Paulo-SP, Brazil*
(e-mail: leugim@ime.usp.br)
‡ *Faculdade de Matemática, Universidade Federal de Uberlândia, Av. João Naves de Avila, 2121, CEP 38408-100, Uberlândia-MG, Brazil*
(e-mail: rodrigolambert@ufu.br)

(*Received* 20 *May* 2017 *and accepted in revised form* 13 *September* 2017)

Abstract. We consider two independent and stationary measures over $\chi^{\mathbb{N}}$, where χ is a finite or countable alphabet. For each pair of n-strings in the product space we define $T_n^{(2)}$ as the length of the shortest path connecting one of them to the other. Here the paths are generated by the underlying dynamic of the measures. If they are ergodic and have positive entropy we prove that, for almost every pair of realizations $(\mathbf{x}, \mathbf{y})$, $T_n^{(2)}/n$ is concentrated in one, as n diverges. Under mild extra conditions we prove a large-deviation principle. We also show that the fluctuations of $T_n^{(2)}$ converge (only) in distribution to a non-degenerate distribution. These results are all linked to a quantity that computes the similarity between those two measures. This is the so-called divergence between two measures, which is also introduced. Several examples are provided.

1. *Introduction: the shortest-path function*

Suppose one has to build a communication net consisting of nodes and links between nodes. A question of major interest is how to design the net such that it is easy to communicate between nodes without paying the cost of constructing a large number of links.

In this paper we study a quantity which describes the structural complexity of the net. Given two nodes, it gives the length of the shortest path from one node to another. We consider the case where the nodes are given by the partition into n-cylinders or n-strings of the phase space: specifically, we consider a finite or countable set χ. For each $n \in \mathbb{N}$, the nodes correspond to the partition of n-cylinders or n-strings of $\chi^{\mathbb{N}}$. We consider also two independent probability measures over $\Omega = \chi^{\mathbb{N}}$. The address node is chosen according to a measure μ and the source node according to a measure ν. We assume that both measures are ergodic and that μ is absolutely continuous with respect to ν, otherwise the communication could be impossible. We denote by $T_n^{(2)}$ the function that gives the length of the shortest path between two n-strings. The link between these two strings is driven by the shift operator σ over Ω. That is, for $\mathbf{x} = (x_0, x_1, \ldots) \in \Omega$ one gets $\sigma\mathbf{x} = (x_1, x_2, \ldots)$.

The cornerstone of this paper is the quantity that gives the minimum number of steps to get from one string to another.

Definition 1.1. The shortest-path function is defined by

$$T_n^{(2)}(\mathbf{x}, \mathbf{y}) = \inf\{k \geq 1 : y_0^{n-1} \cap \sigma^{-k}(x_0^{n-1}) \neq \emptyset\}.$$

Henceforth we write x_m^n as shorthand for $x_m x_{m+1} \ldots x_n$ for any $0 \leq m \leq n \leq \infty$. To illustrate this definition, let us take a look at the word ABRACADABRA in three different languages: let $\mathbf{x}, \mathbf{y}, \mathbf{z}$ be such that $x_0^{10} =$ ABRACADABRA (English), $y_0^{11} =$ AVRAKEHDABRA (Aramaic) and $z_0^{12} =$ ABBADAKEDABRA (Chaldean). Then $T_{11}^{(2)}(\mathbf{x}, \mathbf{y}) = 8$ since, considering the first 11 letters of $\mathbf{x}$ and $\mathbf{y}$, we have to shift y_0^{10} eight times to be able to connect it with x_0^{10}. Similarly, $T_{11}^{(2)}(\mathbf{x}, \mathbf{z}) = T_{11}^{(2)}(\mathbf{y}, \mathbf{z}) = 9$. Furthermore, we have $T_{11}^{(2)}(\mathbf{y}, \mathbf{x}) = T_{11}^{(2)}(\mathbf{z}, \mathbf{y}) = T_{11}^{(2)}(\mathbf{z}, \mathbf{x}) = 10$ and $T_{11}^{(2)}(\mathbf{x}, \mathbf{x}) = 7$, $T_{12}^{(2)}(\mathbf{y}, \mathbf{y}) = 11$, $T_{13}^{(2)}(\mathbf{z}, \mathbf{z}) = 12$.

The random variable $T_n^{(2)}$ is a two-dimensional version of the *shortest return function* $T_n(\mathbf{x}) = \inf\{k \geq 1 \mid x_0^{n-1} \cap \sigma^{-k}(x_0^{n-1})\}$. That is, it gives the length of the shortest path starting from and arriving at the same node. Its concentration phenomenon has already been studied in [**5**, **17**]. A large-deviation principle was related to the Rényi entropy in [**1**, **4**, **10**]. Limiting theorems for its fluctuations were presented in [**2**, **3**]. Since T_n considers starting and target sets being the same and $T_n^{(2)}$ allows them to come from different measures, T_n and $T_n^{(2)}$ are different in nature. In topological terms: T_n describes a local, while $T_n^{(2)}$ describes a global characteristic of the connection net.

In this paper we prove three fundamental theorems which describe the net through the statistical properties of $T_n^{(2)}$: concentration, large deviations and fluctuations.

First we prove that $T_n^{(2)}/n$ converges almost surely to one, as n diverges. Our result holds when μ has positive entropy and ν satisfies some specification property which prevents the net from being extremely sparse.

The concentration of $T_n^{(2)}/n$ leads us to study its large-deviation properties, namely, the rate of decay to zero of the probability of this ratio deviating from one. We compute this rate under the additional requirement that the measures satisfy certain regularity conditions. A similar condition was introduced and already related to the existence of a large-deviation principle for the shortest return function T_n in [**1**].

The limiting rate of the large-deviation function of $T_n^{(2)}$ is determined by a quantity that deserves attention in its own right. It gives a measure of similarity (or difference) between two measures (see Definition 2.1). In words, it is the expectation of the marginal distribution of order k of one of them with respect to the other. Since it is symmetric, their roles are exchangeable in this definition. We call it the divergence of order k. We also study some of its properties that are used later in the large-deviation principle for $T_n^{(2)}$ mentioned above. We provide several examples. In many cases the divergence results in an exponentially decreasing sequence on k and this leads us to consider its limiting rate. One of our main results establishes the existence of the limiting rate function which is far from obvious. We use a kind of sub-additivity property but with a telescopic technique rather than the classical linear one. We show that, in particular, when the two measures coincide, this limit corresponds to the Rényi entropy of the measure at argument $\beta = 2$ (see item (*e*) of examples (2.1)).

To describe the complexity of the net, we study the distribution of the shortest path function. We compute the distribution of a rescaled version of $T_n^{(2)}$ (namely, $n - T_n^{(2)}$) and prove that it converges to a non-degenerate distribution which depends on the stationary measures μ and ν. The limiting distribution may depend on an infinite number of parameters if the measures do. This limiting distribution also depends on the divergence between the measures. As an application of this theorem we compute the proportion of pairs of n-strings which do not overlap (which we call the *avoiding pairs* set).

When the subject is the distribution of $T_n^{(2)}$ we are not aware of any work which considers its behaviour in the context of stationary measures. There are some works which consider models of random graphs and present empirical data which adjust the distribution of the shortest path to Weibull or gamma distributions [**6**, **23**]. But even for classical models (e.g. Erdős–Rényi graphs), its full distribution, in a theoretical sense, has never been considered in the literature [**7**, **12**, **21**].

Since the random variables $T_n^{(2)}$ are defined on the same probability space, we further inquire into a stronger convergence. Our last result shows that $n - T_n^{(2)}$ does not even converge in probability, and a lower bound for the distance between two consecutive terms of the sequence $n - T_n^{(2)}$ is given.

Finally, we think it is important also to highlight the connection of the shortest path function with the study of Poincaré recurrence statistics. The waiting time function introduced by Wyner and Ziv in [**25**] has been well studied in the literature. Given two realizations $\mathbf{x}, \mathbf{y} \in \chi^{\mathbb{N}}$, it is the time expected until x_0^{n-1} appears in the realization $\mathbf{y}$ of another process:

$$W_n(\mathbf{x}, \mathbf{y}) = \inf\{k \geq 1 : y_k^{k+n-1} = x_0^{n-1}\}.$$

Now, we have that the shortest path function is the minimum of the waiting times of x_0^{n-1}, taking the minimum over all the realizations $\mathbf{z} \in \Omega$ that begin with y_0^{n-1}:

$$T_n^{(2)}(\mathbf{x}, \mathbf{y}) = \inf_{\mathbf{z}: z_0^{n-1} = y_0^{n-1}} W_n(\mathbf{x}, \mathbf{z}).$$

A number of classical results are known for W_n. When both strings are chosen with the same measure, Shields showed that for stationary ergodic Markov chains, $\ln W_n/n \to h$ for almost every pair of realizations, as n diverges and h is the Shannon entropy of the measure [**18**]. Nobel and Wyner [**15**] had proven a convergence in probability to the same limit. This result holds for α-mixing processes with a certain rate function α. Marlon and Shields extended it to weak Bernoulli processes [**14**]. Yet Shields [**18**] constructed an example of a very weak Bernoulli process in which the limit does not hold. Finally, Wyner [**24**] proved that the distribution of $W_n(\mathbf{x})\mu(\mathbf{x})$ converges to the exponential law for ψ-mixing measures. When both strings are chosen with possibly different measures, and the second one is a Markov chain, Kontoyiannis [**13**] showed that $\ln W_n/n \to h(\mu) + h(\mu||\nu)$ for $h(\mu||\nu)$ the relative entropy of μ with respect to ν.

This paper is organized as follows. In §2 we introduce the divergence concept. Properties, examples and the proof of its existence are also included in this section. In §3 we prove the concentration of the shortest path function. A large-deviation principle is proved in §4. The convergence of the shortest path distribution is presented in §5. An application to calculating the self-avoiding pairs of strings is also given. The non-convergence in probability is shown in §6.

2. *The divergence between two measures*

Let μ and ν be two probability measures over the same measurable space $(\Omega, \mathcal{F})$. It is natural to ask if these two measures are related in any sense, and if there is some function to scale this relation. There are in the literature several quantities devoted to answering these questions. We highlight here the *mutual information* and the *relative entropy* (or *Kullback–Leibler divergence*), which have been extensively discussed in the literature (see, for instance, [**9**]).

The present section is dedicated to a quantity which describes the degree of similarity between two given measures. As far as we know, it has never been considered in the literature. Its definition is as follows.

Definition 2.1. The k-divergence between μ and ν is defined by

$$\mathbb{E}_{\mu,\nu}(k) = \sum_{\omega \in \chi^k} \mu\nu(\omega)$$

(where by $\mu\nu(\omega)$ we always mean $\mu(\omega)\nu(\omega)$).

Let μ_k (ν_k) be the projection of μ (ν) over the first k coordinates of the space. The k-divergence is the mean of μ_k with respect to ν_k, or *vice versa*. It is also the inner product of the $|\chi|^k$-vectors with entries given by the probabilities $\mu(\omega)$ and $\nu(\omega)$, $\omega \in \chi^k$ (in any arbitrary ordering of the strings ω). Notice that $\mathbb{E}_{\mu,\nu}(k)$ is symmetric ($\mathbb{E}_{\mu,\nu}(k) = \mathbb{E}_{\nu,\mu}(k)$), and that it is not null if, and only if, the supports of the two measures have non-empty intersection. The previous sentence can be interpreted as follows: if the two measures do not communicate, the similarity between them is zero.

The next result says that the operation of opening a gap does not produce a smaller result. As a corollary, we conclude that the k-divergence is not increasing in k. For simplicity, hereafter for $\omega \in \chi^n$ and $\xi \in \chi^m$ we denote by $\omega\xi$ the $(n+m)$-string constructed by concatenation of ω and ξ. Formally, $\omega \cap \sigma^{-n}\xi$.

LEMMA 2.1. *Let $i + g + j = k$ be non-negative integers. Then*

$$\mathbb{E}_{\mu,\nu}(k) \leq \sum_{\omega \in \chi^i; \zeta \in \chi^j} \mu\nu(\omega \cap \sigma^{-(i+g)}\zeta). \tag{1}$$

Proof. Consider $\omega \in \chi^i$, $\xi \in \chi^g$ and $\zeta \in \chi^j$. Let us write the cylinder $\omega\sigma^{-i}\xi\sigma^{-(i+g)}\zeta \in \mathcal{F}_0^{k-1}$ by concatenating the three cylinders above. By removing the string ξ in $\mu(\omega\xi\zeta)$,

$$\sum_{\xi \in \chi^g} \mu\nu(\omega\xi\zeta) \leq \sum_{\xi \in \chi^g} \mu(\omega \cap \sigma^{-(i+g)}\zeta)\nu(\omega\xi\zeta)$$
$$= \mu\nu(\omega \cap \sigma^{-(i+g)}\zeta).$$

Summing over ω and ζ in the last display, we get (1). □

As a direct consequence of the above proposition, we get that the k-divergence is monotonic in k.

COROLLARY 2.1. *If $k < l$, then $\mathbb{E}_{\mu,\nu}(k) \geq \mathbb{E}_{\mu,\nu}(l)$.*

Proof. This follows by taking $i = k$, $g = 0$ and $j = l - k$. □

The above corollary proves that the k-divergence is a non-increasing function in k. In many cases, it decreases at an exponential rate. It is natural to ask about the existence of the limiting rate function, that is,

$$\underline{\mathcal{R}} = \liminf_{k\to\infty} \left(-\frac{1}{k} \log \mathbb{E}_{\mu,\nu}(k)\right), \quad \overline{\mathcal{R}} = \limsup_{k\to\infty} \left(-\frac{1}{k} \log \mathbb{E}_{\mu,\nu}(k)\right).$$

If both limits are equal, we denote this function by $\mathcal{R}$†.

In what follows, we provide some examples. They illustrate cases for the existence (or not) of $\mathcal{R}$. Finally, we state the main result of this section, a general condition in which the limiting rate exists. Furthermore, in §4 we will relate this limiting rate to a large-deviation principle for $T_n^{(2)}$.

Examples 2.1.

(a) Let μ and ν be two independent measures with disjoint supports. Then $\mathbb{E}_{\mu,\nu}(k) = 0$ for all k, and therefore $\mathcal{R} = \infty$.

(b) Suppose that μ and ν concentrate their mass in a unique realization $\mathbf{x}$ of the process. Then, for any $\mathbf{y} \in \chi^{\mathbb{N}}$, $\mu\nu(y_0^{k-1}) = 1$ if, and only if, $y_0^{k-1} = x_0^{k-1}$ (and zero otherwise). Thus we get $\mathcal{R} = 0$.

(c) If both measures μ and ν have independent and identically distributed marginals we get

$$\mathbb{E}_{\mu,\nu}(k) = \sum_{\omega\in\chi^k} \mu\nu(\omega) = \left[\sum_{x_0\in\chi} \mu\nu(x_0)\right]^k = \mathbb{E}_{\mu,\nu}(1)^k.$$

Therefore, the limit $\mathcal{R}$ exists and is given by

$$\mathcal{R} = -\log E_{\mu,\nu}(1).$$

(d) Let μ be a product of Bernoulli measures with parameter p and ν a product of Bernoulli measures with parameter $1 - p$. Then

$$\mathbb{E}_{\mu,\nu}(k) = \sum_{x_0^{k-1}\in\chi^k} \prod_{i=0}^{k-1} \mu\nu(x_i) = 2^k p^k (1-p)^k.$$

Then we get

$$\mathcal{R} = -\log[2p(1-p)].$$

(e) If $\mu = \nu$, we get that $\mathcal{R} = H_\mu(2)$, where

$$H_\mu(\beta) = -\lim_{k\to\infty} \frac{1}{k(\beta-1)} \log \sum_{\omega\in\chi^k} \mu^\beta(\omega)$$

is the Rényi entropy of the measure μ (provided that it exists).

† Throughout this paper logarithms can be taken in any base.

(f) A case where the limiting rate does not exist: a sequence that does not satisfy the law of large numbers. Let $\chi = \{0, 1\}$, and let μ be a measure concentrated on the realization

$$\mathbf{x} = 0^2 1^{2^2} 0^{2^3} 1^{2^4} \cdots 0^{2^k} 1^{2^{k+1}} \cdots ,$$

where a^j means the j-string $aa \cdots a \in \chi^j$. On the other hand, let ν be a product of Bernoulli measures with $p \neq 1/2$. By a direct computation, we get that $\mathbb{E}_{\mu,\nu}(k) = \nu(x_0^{k-1})$. Since the proportion of 0s and 1s in x_0^{k-1} does not converge as k goes to infinity, we get $\underline{\mathcal{R}} \neq \overline{\mathcal{R}}$.

(g) Let ν be an ergodic, positive entropy measure. By the Shannon–McMillan–Breiman theorem, $-(1/k) \log \nu(x_0^{k-1})$ converges to $h(\nu)$, for almost every $\mathbf{x} \in \chi^{\mathbb{N}}$, where $h(\nu)$ is the entropy of ν. Let $\mathbf{x}$ be one such sequence. Let μ be a measure concentrated on $\mathbf{x}$. Then $\mathbb{E}_{\mu,\nu}(k) = \nu(x_0^{k-1})$ and $\mathcal{R} = h(\nu)$.

The following theorem gives sufficient conditions for the existence of the limiting rate $\mathcal{R}$. Its proof uses a kind of sub-additivity property. But here, instead of the classical linear iteration of the sub-additivity property, we use a geometric iteration. To prove the existence of the limiting rate function $\mathcal{R}$ we use a kind of g-regular condition which is a version of the condition introduced in [**1**]. This condition was used to prove a large-deviation principle for the shortest return function T_n of a string to itself. That principle related the deviations of T_n to the Rény entropies of the measure. Examples which show its generality and other properties can also be found there.

In what follows, we present two quantities that will be very useful throughout this section. Let g be a fixed non-negative integer. For the measure μ (respectively, ν), set

$$\psi^+_{\mu,g}(i, j) = \sup_{\omega \in \chi^i, \xi \in \chi^j} \frac{\mu(\sigma^{-(i+g)}\xi \mid \omega)}{\mu(\xi)}, \tag{2}$$

and then

$$\psi^+_g = \max\{\psi^+_{\mu,g}, \psi^+_{\nu,g}\}.$$

We are now ready to state the main result of the present section. It provides a general condition for the existence of the limiting rate $\mathcal{R}$.

THEOREM 2.1. *Suppose that there exist positive constants $K > 0$ and $\epsilon > 0$ such that*

$$\log \psi^+_g(i, j) \leq K \frac{i + j}{[\log(i + j)]^{1+\epsilon}}. \tag{3}$$

Then $\mathcal{R}$ does exist.

For instance, if μ and ν have independent marginals, we take $g = 0$. Immediate calculations give $\psi^+_{\mu,g}(i, j) = \psi^+_{\nu,g}(i, j) = 1$ for all i and j, and condition (3) is satisfied. Moreover, if μ and ν are stationary measures of irreducible, aperiodic, positive recurrent Markov chains in a finite alphabet χ, then the Markov property gives

$$\psi^+_{m,0}(i, j) = \sup_{\omega \in \chi, \xi \in \chi} \frac{m(\sigma^{-1}\xi \mid \omega)}{m(\xi)}, \quad m = \mu, \nu,$$

which is finite since χ is finite and (3) is satisfied. Abadi and Cardeño [**1**] constructed several examples of processes of renewal type, with g equal zero and one with exponential or sub-exponential measure of cylinders which satisfies (3). Measures μ which satisfy the classical ψ-mixing condition, for each g fixed, have $\psi^+_{\mu,g}$ constant and thus (3) holds.

We now present the proof of the theorem.

Proof. Let us take $\omega \in \chi^i$, $\xi \in \chi^g$ and $\zeta \in \chi^j$. As in the proof of Lemma 2.1, $\sum_{\xi\in\chi^g} \mu\nu(\omega\xi\zeta) \le \mu\nu(\omega \cap \sigma^{-(i+g)}\zeta)$. Further, by (2),

$$\mu(\omega \cap \sigma^{-(i+g)}\zeta) \le \psi^+_g(i, j)\mu(\omega)\mu(\zeta).$$

The same holds for ν. Write $f(k) = \log \mathbb{E}_{\mu,\nu}(k)$. Write also $c_g(i, j) = 2 \log \psi^+_g(i, j)$. Summing up in ω and ζ and taking logarithms, by the inequalities above, we conclude that for all i, j,

$$\begin{aligned} f(i + g + j) &\le \log(\psi^+_g(i, j)\psi^+_g(i, j)) + \log \sum_{\omega\in\chi^i} \mu\nu(\omega) + \log \sum_{\zeta\in\chi^j} \mu\nu(\zeta) \\ &= c_g(i, j) + f(i) + f(j). \end{aligned} \tag{4}$$

Now we use a kind of sub-additivity argument. Let $(n_t)_{t\in\mathbb{N}}$ an increasing sequence of non-negative integers such that

$$\liminf_{n\to\infty} \frac{f(n)}{n} = \lim_{t\to\infty} \frac{f(n_t)}{n_t}. \tag{5}$$

Consider the sequence $\tilde{n}_t = n_t + g$, with $t \in \mathbb{N}$. Fix t. Firstly, for any positive integer $n \ge \tilde{n}_t$, write $n = \tilde{n}_t m + r$, with positive integers m, r such that $0 \le r < \tilde{n}_t$. Apply (4) with $i = \tilde{n}_t m - g$, $j = r$ and gap g to get

$$f(n) \le c_g(\tilde{n}_t m - g, r) + f(\tilde{n}_t m - g) + f(r). \tag{6}$$

Now we write m in base 2. For this, there exist a positive integer $\ell(m)$ and non-negative integers $\ell_1 < \ell_2 < \cdots < \ell_{\ell(m)}$ such that $m = \sum_{s=1}^{\ell(m)} 2^{\ell_s}$. Iterating (4) with $i = \tilde{n}_t \sum_{s=1}^{u-1} 2^{\ell_s} - g$ and $j = \tilde{n}_t 2^{\ell_u} - g$, for $u = 2, \ldots, \ell(m)$, we have that for the middle term in the right-hand side of (6),

$$f(\tilde{n}_t m - g) \le \sum_{u=2}^{\ell(m)} c_g\left(\tilde{n}_t \sum_{s=1}^{u-1} 2^{\ell_s} - g, \tilde{n}_t 2^{\ell_u} - g\right) + \sum_{u=1}^{\ell(m)} f(\tilde{n}_t 2^{\ell_u} - g). \tag{7}$$

The first sum on the right-hand side is zero if $\ell(m) = 1$. Finally, we decompose the argument in the last summation. For any $n \in \mathbb{N}$ of the form $n = \tilde{n}_t 2^\ell - g$, we apply (4) with $i = j = \tilde{n}_t 2^{\ell-1} - g$ to get

$$f(\tilde{n}_t 2^\ell - g) \le c_g(\tilde{n}_t 2^{\ell-1} - g, \tilde{n}_t 2^{\ell-1} - g) + 2f(\tilde{n}_t 2^{\ell-1} - g).$$

An iteration of the above inequality leads to

$$f(\tilde{n}_t 2^\ell - g) \le \sum_{s=0}^{\ell-1} 2^{\ell-s-1} c_g(\tilde{n}_t 2^s - g, \tilde{n}_t 2^s - g) + 2^\ell f(\tilde{n}_t - g). \tag{8}$$

Observe that $\tilde{n}_t - g = n_t$. Taking (6)–(8) together, we conclude that the limit superior of $f(n)/n$ is upper bounded by the limit superior of $\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}$, where

$$\mathrm{I} = \frac{c_g(\tilde{n}_t m - g, r)}{\tilde{n}_t m - g + r},$$
$$\mathrm{II} = \frac{1}{\tilde{n}_t m} \sum_{u=2}^{\ell(m)} c_g\left(\tilde{n}_t \sum_{s=1}^{u-1} 2^{\ell_s} - g, \tilde{n}_t 2^{\ell_u} - g\right),$$
$$\mathrm{III} = \frac{1}{\tilde{n}_t m} \sum_{u=1}^{\ell(m)} \sum_{s=0}^{\ell_u - 1} 2^{\ell_u - s - 1} c_g(\tilde{n}_t 2^s - g, \tilde{n}_t 2^s - g),$$
$$\mathrm{IV} = \frac{f(r)}{\tilde{n}_t m} + \frac{\sum_{u=1}^{\ell(m)} 2^{\ell_u} f(n_t)}{\tilde{n}_t m}.$$

As m diverges, the first term in IV vanishes since $0 \le r < \tilde{n}_t$. The second is bounded by $f(n_t)/n_t$. I goes to zero by (3). We recall that $\mathrm{II} = 0$ if $\ell(m) = 1$. Otherwise we also use condition (3) to get the upper bound

$$\frac{K}{m} \sum_{u=2}^{\ell(m)} \frac{\sum_{s=1}^{u} 2^{\ell_s}}{\left[\log\left(\tilde{n}_t \sum_{s=1}^{u} 2^{\ell_s} - 2g\right)\right]^{1+\epsilon}}.$$

The inner summation is trivially bounded by $3 \le \sum_{s=1}^{u} 2^{\ell_s} \le 2^{\ell_u + 1}$. Since $m = \sum_{u=1}^{\ell(m)} 2^{\ell_u}$, it follows that $\mathrm{II} \le 2K/[\log n_t]^{1+\epsilon}$. Finally, again using (3), we get that III is upper-bounded by

$$\frac{K}{m} \sum_{u=1}^{\ell(m)} 2^{\ell_u} \sum_{s=0}^{\ell_u - 1} \frac{1}{[\log(\tilde{n}_t 2^{s+1} - 2g)]^{1+\epsilon}}.$$

Changing the constant K, we can take the logarithm to base 2 here. The argument in the logarithm is lower-bounded by $n_t 2^{s+1}$. Now we use the fact that the sum of a decreasing sequence is bounded above by its first term plus the definite integral by the first and last terms in the sum. Thus, the rightmost sum in the above display is bounded by

$$\sum_{s=1}^{\infty} \frac{1}{[s + \log n_t]^{1+\epsilon}} \le \frac{1}{[1 + \log n_t]^{1+\epsilon}} + \frac{1}{\epsilon[1 + \log n_t]^{\epsilon}},$$

which goes to zero as t diverges. Summarizing, we conclude that

$$\limsup_{n \to \infty} \frac{f(n)}{n} \le \frac{f(n_t)}{n_t} + \frac{K}{[\log n_t]^{1+\epsilon}} + \frac{1}{\epsilon[1 + \log n_t]^{\epsilon}}.$$

The inequality holds for every t. If we take $t \to \infty$ then we conclude the proof since $f(n_t)/n_t$ converges by hypothesis. □

3. *Concentration*

The present section is devoted to the study of the asymptotics for $T_n^{(2)}$. Intuition says that the bigger is n, the more difficult it is to connect two n-strings. Thus we expect $T_n^{(2)}$ to increase. The question is at what rate. The main result of this section says that T_n/n

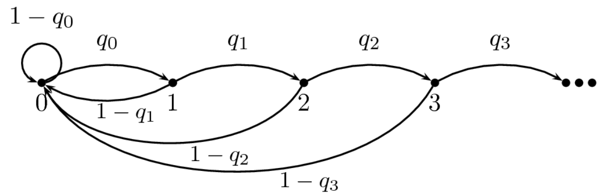

FIGURE 1. House of cards Markov chain $(Y_n)_{n\geq 0}$.

converges almost surely to one. The proof is divided into two parts. The first part proves that the limit inferior is lower-bounded by one, and the second states that the limit superior is upper-bounded by one. For the latter one we assume that the process satisfies the very weak specification property (see Definition 3.1 below).

There are several definitions of specification in the literature of dynamical systems. The first one was introduced by Bowen [**8**]. Many others have appeared, mainly following him but with some differences of nomenclature [**11**, **19**], or in weaker forms (see, for instance, [**20**, **22**, **26**]). Basically they mean that, for any given set of strings, they can be observed (at least in one single realization of the process) with bounded gaps between them. Sometimes the realization is required to be periodic. For the convenience of the reader, we present our condition here. It is easy to see that it is satisfied for a large class of stochastic processes and is less restrictive than the previous ones. Examples are provided below.

Definition 3.1. $(\chi^{\mathbb{N}}, \mu, \sigma)$ is said to have the very weak specification property (VWSP) if there exists a function $g : \mathbb{N} \to \mathbb{N}$ with $\lim_{n\to\infty} g(n)/n = 0$, that satisfies the following: for any pair of strings $\omega, \xi \in \chi^n$, there exists an $\mathbf{x} \in \chi^{\mathbb{N}}$ such that

$$x_0^{n-1} = \omega \quad \text{and} \quad x_{n+g(n)}^{2n+g(n)-1} = \xi.$$

Examples 3.1.

(a) Any process with complete grammar satisfies Definition 3.1 with $g(n) = 0$. We recall that a probability measure μ defined over $\chi^{\mathbb{N}}$ is said to have complete grammar if, for all $n \in \mathbb{N}$, we get $\mu(\omega) > 0$ for all $\omega \in \chi^n$.

(b) An irreducible and aperiodic Markov chain over a finite alphabet χ and stationary measure μ satisfies the VWSP with $g(n) \leq |\chi|$.

(c) We first construct a renewal process $(X_n)_{n\in\mathbb{N}}$ as an image of the house of cards Markov chain $(Y_n)_{n\in\mathbb{N}}$ with irreducible and aperiodic transition matrix Q given by

$$Q(y, 0) = 1 - q_y,$$
$$Q(y, y+1) = q_y,$$

$y \in \{0, 1, 2, \dots\}$. Figure 1 shows the transitions of this process.
Let $X_n = 0$ if $Y_n \neq 0$, and $X_n = 1$ if $Y_n = 0$, indicating the 'renewal' of $(Y_n)_{n\geq 0}$. Take $q_y = 1$, for all $n^2 \leq y \leq n^2 + n$ for some $n \in \mathbb{N}$, and any other $0 < q_y < 1$ for the remaining coefficients to guarantee that the Markov chain is positive recurrent. Obviously $(X_n)_{n\geq 0}$ does not have complete grammar. It is easy to see that $g(n) \leq \sqrt{n}$

and that this bound is actually sharp, taking $\omega = \xi = 01^{n^2-1} \in \chi^{n^2}$. Thus $(X_n)_{n\geq 0}$ satisfies the VWSP. On the other hand, the stationary measure of the house of cards Markov chain itself is an example that does not satisfy the VWSP.

We can now state the concentration theorem, which is the main result of this section.

THEOREM 3.1. *If μ has positive entropy, then:*
(a) $\liminf_{n\to\infty} (T_n^{(2)}/n) \geq 1$, *$\mu \times \nu$-almost everywhere.*
(b) *In addition, if ν satisfies the VWSP, then*

$$\lim_{n\to\infty} \frac{T_n^{(2)}}{n} = 1, \quad \mu \times \nu\text{-almost everywhere.}$$

Before proving the above result, let us introduce a family of sets and a result that will be useful for the proof.

Definition 3.2. For each $k \in \{1, \ldots, n-1\}$ define the set of pairs $(\mathbf{x}, \mathbf{y}) \in \chi^{\mathbb{N}} \times \chi^{\mathbb{N}}$ such that the first k symbols of x_0^{n-1} coincide exactly with the last k symbols of y_0^{n-1}. Namely,

$$R_n^{(2)}(k) = \{(\mathbf{x}, \mathbf{y}) \in \chi^{\mathbb{N}} \times \chi^{\mathbb{N}} : y_{n-k}^{n-1} = x_0^{k-1}\}.$$

For instance, if $\mathbf{x}$, $\mathbf{y}$ are such that $y_0^5 = 011100$ and $x_0^5 = 111100$, then $(\mathbf{x}, \mathbf{y}) \in R_4^{(2)}(3)$.

The following result establishes a connection between the shortest-path function and the $R_n^{(2)}(i)$-sets.

LEMMA 3.1. *For $k < n$,*

$$\{T_n^{(2)} \leq k\} \subseteq \bigcup_{i=n-k}^{n-1} R_n^{(2)}(i).$$

In addition, if ν satisfies the complete grammar condition, then the equality holds.

Proof. By definition, $(\mathbf{x}, \mathbf{y})$ belongs to $\{T_n^{(2)} \leq k\}$ if, and only if, there are $\mathbf{z} \in \chi^{\mathbb{N}}$ and $1 \leq i \leq k$ such that $z_0^{n-1} = y_0^{n-1}$ and $z_i^{i+n-1} = x_0^{n-1}$. In particular, since $n - 1 \geq i$, we have $y_i^{n-1} = x_0^{n-i-1}$, which in turns says that $(\mathbf{x}, \mathbf{y}) \in R_n^{(2)}(n-i)$. For the equality, notice that for any pair $(\mathbf{x}, \mathbf{y}) \in R_n^{(2)}(i)$, $i \in \{n-k, \ldots, n-1\}$, we get that $x_0^{i-1} = y_{n-i}^{n-1}$. The complete grammar condition ensures that there exists $\mathbf{z} \in \chi^{\mathbb{N}}$ such that $z_0^{n-1} = y_0^{n-1}$ and $z_i^{i+n-1} = x_0^{n-1}$. This concludes the proof. □

The next lemma gives the key connection with the divergence of μ and ν.

Henceforth, by $\mathbb{P}$ we mean the product measure $\mu \times \nu$.

LEMMA 3.2. *For $1 \leq k < n$, and ν a stationary measure,*

$$\mathbb{P}(R_n^{(2)}(k)) = \mathbb{E}_{\mu,\nu}(k).$$

Proof. If $1 \leq k < n$, then

$$\mathbb{P}(R_n^{(2)}(k)) = \mathbb{P}(x_0^{k-1} = y_{n-k}^{n-1}) = \sum_{\omega \in \chi^k} \mu\nu(\omega).$$

Since the last term above is equal to $\mathbb{E}_{\mu,\nu}(k)$, the proof is complete. □

We are now in a position to prove Theorem 3.1.

Proof of Theorem 3.1. For item (a), let $h > 0$ be the entropy of μ. Since μ is ergodic, the Shannon–McMillan–Breiman theorem says that

$$-\lim_{n\to\infty}\frac{1}{n}\log\mu(x_0^{n-1}) = h,$$

for μ-almost every $\mathbf{x}\in\chi^{\mathbb{N}}$. By Egorov's theorem, for every $0<\epsilon<h$, there exists a subset Ω_ϵ of Ω where this convergence is uniform and $\mu(\Omega_\epsilon)\geq 1-\epsilon$. That is, for all $\epsilon>0$, there exists a $k_0(\epsilon)$ such that for all $k>k_0(\epsilon)$,

$$e^{-k(h+\epsilon)} < \mu(x_0^{k-1}) < e^{-k(h-\epsilon)}, \tag{9}$$

for all $\mathbf{x}\in\Omega_\epsilon$. Forming the product with ν, and using Lemmas 3.1 and 3.2, we get

$$\begin{aligned}\mathbb{P}(\{T_n^{(2)}\leq(1-\epsilon)n\}\cap\Omega_\epsilon\times\Omega) &\leq \sum_{j=\lfloor\epsilon n\rfloor}^{n-1}\mathbb{P}(R_n^{(2)}(j)\cap\Omega_\epsilon\times\Omega)\\ &= \sum_{j=\lfloor\epsilon n\rfloor}^{n-1}\sum_{\omega\in\chi^j\cap\Omega_\epsilon}\mu\nu(\omega)\\ &\leq \sum_{j=\lfloor\epsilon n\rfloor}^{n-1} e^{-j(h-\epsilon)},\end{aligned}$$

where the last inequality was obtained using (9). A direct computation gives

$$\sum_{n=1}^{\infty}\mathbb{P}(T_n^{(2)}\leq(1-\epsilon)n)\leq\frac{1}{1-e^{-(h-\epsilon)}}.$$

By the Borel–Catelli lemma, $\{T_n^{(2)}\leq(1-\epsilon)n\}$ occurs only finitely many times. We conclude

$$\liminf_{n\to\infty}\frac{T_n^{(2)}}{n}\geq 1-\epsilon,\quad \mu\times\nu\text{-almost surely in }\Omega_\epsilon\times\Omega. \tag{10}$$

Since ϵ is arbitrary, this completes the proof of (a).

We now prove item (b). Since Definition 3.1 implies that $T_n^{(2)}\leq n+g(n)$, we divide both sides by n and get

$$\limsup_{n\to\infty}\frac{T_n^{(2)}}{n}\leq 1,\quad \mu\times\nu\text{-almost everywhere.}$$

Combining this with (a), we conclude the proof of item (b). □

4. *Large deviations*

In the previous section we showed that $T_n^{(2)}/n$ concentrates its mass at 1, as n diverges. Here we present the deviation rate for this limit. Since the VWSP implies that

$$\mathbb{P}\left(\frac{T_n^{(2)}}{n}>1+\epsilon\right)=0\quad\text{for all }n>n_0(\epsilon),$$

it is only meaningful to consider the lower deviation.

Definition 4.1. We define the lim inf and lim sup for the lower deviation rate, respectively, as

$$\underline{\Delta}(\epsilon) = \liminf_{n\to\infty} \frac{1}{n}\left| \log \mathbb{P}\left(\frac{T_n^{(2)}}{n} < 1-\epsilon \right) \right|$$

and

$$\overline{\Delta}(\epsilon) = \limsup_{n\to\infty} \frac{1}{n}\left| \log \mathbb{P}\left(\frac{T_n^{(2)}}{n} < 1-\epsilon \right) \right|.$$

If $\underline{\Delta} = \overline{\Delta}$ we write simply Δ.

We recall that the complete grammar condition ensures that $T_n^{(2)} \le n$.

THEOREM 4.1. *Let μ and ν be two stationary probability measures defined over $\chi^{\mathbb{N}}$.*

(a) *Then $\underline{\Delta}(\epsilon) \ge \epsilon\underline{\mathcal{R}}$ and $\overline{\Delta}(\epsilon) \le \epsilon\overline{\mathcal{R}}$.*

(b) *Suppose that ν has complete grammar. Then the equalities hold in (a).*

The ψ_g-regularity of the measure ensures the existence of $\mathcal{R}$.

COROLLARY 4.1. *Under conditions of Theorem 4.1, suppose that ν has complete grammar. Then $\Delta(\epsilon) = \epsilon\mathcal{R}$.*

Proof of Theorem 4.1. By Lemma 3.1, we have

$$\left\{ \frac{T_n^{(2)}}{n} < 1-\epsilon \right\} \subseteq \bigcup_{j=\lceil n\epsilon\rceil}^{n-1} R_n^{(2)}(j),$$

with equality if the process has complete grammar. In this case, considering just the first set in the union, we also have

$$R_n^{(2)}(\lceil n\epsilon\rceil) \subseteq \left\{ \frac{T_n^{(2)}}{n} < 1-\epsilon \right\}.$$

Thus, by Lemma 3.2,

$$\mathbb{E}_{\mu,\nu}(\lceil n\epsilon\rceil) \le \mathbb{P}\left(\frac{T_n^{(2)}}{n} < 1-\epsilon \right) \le \sum_{j=\lceil n\epsilon\rceil}^{n-1} \mathbb{E}_{\mu,\nu}(j).$$

Now we take the logarithm, divide by n, take the limit and use the fact that the divergence is non-increasing. A change of variables concludes the proof. □

5. *Convergence in law*

In this section we prove the convergence of the normalized distribution of $T_n^{(2)}$ to a non-degenerate distribution. We also present several examples and provide an application of the main result of the section.

To state the result we first need to introduce the coefficients that appear in the theorem.

Definition 5.1. Set $a_{n-1,n}^{(2)} = 0$ and for every $1 \le k \le n-2$, define:

- $a_{k,n}^{(2)} = \sum_{m=k+1}^{n-1} \sum_{\omega\notin\bigcup_{j=k}^{m-1} R_m(j)} \mu\nu(\omega)$;
- $a_k^{(2)} = \sum_{m=k+1}^{\infty} \sum_{\omega\notin\bigcup_{j=k}^{m-1} R_m(j)} \mu\nu(\omega)$.

Here we define by $R_n(k)$ a set which is a one-dimensional version of $R_n^{(2)}(k)$. Namely,

$$R_n(k) = \{x_0^{n-1} \in \chi^n : x_{n-k}^{n-1} = x_0^{k-1}\}.$$

We can now state the main result of this section.

THEOREM 5.1. *Suppose that ν has complete grammar. Then, for all $1 \leq k \leq n-1$:*
(a) $\mathbb{P}(n - T_n^{(2)} \geq k) = \mathbb{E}_{\mu,\nu}(k) + a_{k,n}^{(2)}$;
(b) $\lim_{n\to\infty} \mathbb{P}(n - T_n^{(2)} \geq k) = \mathbb{E}_{\mu,\nu}(k) + a_k^{(2)}$.

In the next examples we discuss several cases of applications of the above theorem.

Examples 5.1.
(a) The theorem does not guarantee that the limiting object $(\mathbb{E}_{\mu,\nu}(k) + a_k^{(2)})_{n\in\mathbb{N}}$ actually defines a distribution law. For instance, take μ concentrated on the unique sequence $\mathbf{x} = (11111\ldots)$. Let ν_p be a product of Bernoulli measures with success probability p. Let $0 < \lambda < 1$, and define $\nu = \lambda\mu + (1-\lambda)\nu_p$. Clearly, μ is absolutely continuous with respect to ν, which has complete grammar. It is easy to compute $\mathbb{P}(T_n^{(2)} = 1) = \nu(1^n) = \lambda + (1-\lambda)p^n$. Thus $\mathbb{P}(n - T_n^{(2)} = \infty) \geq \lambda$ and $n - T_n^{(2)}$ does not converge to a limiting distribution.
(b) Under mild conditions one gets that the limiting object is actually a distribution. For that, it is enough to give conditions in which $\mathbb{P}(n - T_n^{(2)} = \infty) = \lim_{k\to\infty} \mathbb{E}_{\mu,\nu}(k) + a_k^{(2)} = 0$. Directly from its definition, $a_k^{(2)} \leq \sum_{j=k+1}^{\infty} \mathbb{E}_{\mu,\nu}(j)$. Thus, the limiting function defines a distribution if $\sum_{j=k}^{\infty} \mathbb{E}_{\mu,\nu}(j)$ goes to zero as k diverges. It holds if $\min\{\max_{\omega\in\chi^n} \mu(\omega), \max_{\omega\in\chi^n} \nu(\omega)\}$ is summable. Notice that this is not the case in the example above.
(c) When $\mu = \nu$, we recover in the limit, the same limit distribution of the rescaled shortest return function $n - T_n$. In particular, if μ is a product measure, we recover the limit distribution obtained in [**3**].
(d) The following example shows a process that has complete grammar, and then $n - T_n^{(2)}$ converges. In contrast, the rescaled shortest return function $n - T_n$ does not converge as shown in [**2**]. This is due to the fact that the process is not β-mixing. The process $(X_n)_{n\geq 0}$ is defined over $\chi = \{0, 1\}$ in the following way. Let X_0 be uniformly chosen over $\{0, 1\}$ and independent of everything. The remaining variables are conditionally independent given X_0 and defined by

$$\mu(X_{2n} = X_0) = 1 - \epsilon = 1 - \mu(X_{2n-1} = X_0).$$

It is obvious that the process has complete grammar, has a unique invariant measure with marginal distribution of X_n, $n \geq 1$ being the uniform one. Now take $\nu = \mu$. So they satisfy the hypothesis of Theorem 5.1 and therefore $n - T_n^{(2)}$ converges.

Application. We call $\{(x_0^{n-1}, y_0^{n-1}) \in \chi^n \times \chi^n \mid T_n^{(2)}(\mathbf{x}, \mathbf{y}) = n\}$ the set of *avoiding pairs*, since only in this case do the two chosen strings not overlap. As far as we know this has never been considered in the literature. A similar quantity was actually considered, the set of *self-avoiding strings*. It is defined similarly, but using the shortest return function T_n, instead of the shortest path $T_n^{(2)}$. Namely, $\{T_n = n\}$. It is read as the set of strings which do not overlap themselves. It was first studied in [**16**], when the authors treated a problem related to cellular automata. They considered only the case of a uniform product measure. Using an argument due to Janson, the authors calculate the proportion of self-avoiding strings of length n. This result was generalized in [**3**] and previously in [**2**] to independent and β-mixing processes, respectively.

The next result follows immediately from Theorem 5.1, and gives us the probability of the set of avoiding pairs.

COROLLARY 5.1. *Under the conditions of Theorem 5.1, the measure of the avoiding pairs set is given by:*
(a) $\mathbb{P}(T_n^{(2)} = n) = 1 - \mathbb{E}_{\mu,\nu}(1) - a_{1,n}^{(2)}$;
(b) $\lim_{n\to\infty} \mathbb{P}(T_n^{(2)} = n) = 1 - \mathbb{E}_{\mu,\nu}(1) - a_1^{(2)}$.

We now proceed to the proof of the main result of this section.

Proof of Theorem 5.1. The main idea of this proof is to transform the two-dimensional problem into a one-dimensional one. By Lemma 3.1 and since ν has complete grammar, we get that

$$\{n - T_n^{(2)} \geq k\} = \bigcup_{j=k}^{n-1} R_n^{(2)}(j).$$

Decompose the right-hand side of the above equality into disjoint sets to get

$$R_n^{(2)}(k) \cup \bigcup_{j=k+1}^{n-1} R_n^{(2)}(j) \Big\backslash \bigcup_{l=k}^{m-1} R_n^{(2)}(l). \tag{11}$$

By Lemma 3.2, $\mathbb{P}(R_n^{(2)}(k)) = \mathbb{E}_{\mu,\nu}(k)$. For the second set in (11), $(\mathbf{x}, \mathbf{y}) \in R_n^{(2)}(j)$ if, and only if, $x_0^{j-1} = y_{n-j}^{n-1}$. Further, $(\mathbf{x}, \mathbf{y})$ does not belong to $\bigcup_{l=k}^{m-1} R_n^{(2)}(l)$ if, and only if $x_0^{l-1} \neq y_{n-l}^{n-1}$, for all $k \leq l \leq m - 1$. Since these last two conditions depend only on the values of $x_0, \ldots, x_{m-1}, y_{n-m}, \ldots, y_{n-1}$ we get that the probability of the rightmost set in (11) is equal to

$$\sum_{j=k+1}^{n-1} \sum_{\omega \notin \bigcup_{l=k}^{m-1} R_m(l)} \mu\nu(\omega) = a_{k,n}^{(2)}. \tag{12}$$

Since $\mathbb{E}_{\mu,\nu}(k)$ does not depend on n, the limit of the probability of the left-hand-side set in (11) when n goes to infinity only depends on its second term, which is a non-decreasing function of n. Each term is also bounded above by 1. Therefore it converges, and this concludes the proof. □

6. *Non-convergence in probability*

In the present section we show that the convergence of $n - T_n^{(2)}$ cannot be stronger than convergence in distribution. The result is stated as follows.

PROPOSITION 6.1. *Under the conditions of Theorem 5.1, $n - T_n^{(2)}$ does not converge in probability.*

Proof. It is sufficient to show that

$$\mathbb{P}(|n+1-T^{(2)}_{n+1}-(n-T^{(2)}_n)|>\epsilon)$$

does not converge to zero. Take $0<\epsilon<1$. It is obvious that

$$\{T^{(2)}_{n+1}=T^{(2)}_n\}\subset\{|n+1-T^{(2)}_{n+1}-(n-T^{(2)}_n)|>\epsilon\}.$$

Conditioning on $T^{(2)}_n=k$,

$$\mathbb{P}(T^{(2)}_{n+1}=T^{(2)}_n)=\sum_{k=1}^{\infty}\mathbb{P}(T^{(2)}_{n+1}=k\mid T^{(2)}_n=k)\mathbb{P}(T^{(2)}_n=k).$$

Since ν has complete grammar, the above sum goes just up to n. Further, to get $T^{(2)}_{n+1}=k$ whenever one has $T^{(2)}_n=k$, it is necessary and sufficient to have $y_n=x_{n-k}$, due also to the complete grammar. Thus, $\mathbb{P}(T^{(2)}_{n+1}=k\mid T^{(2)}_n=k)=\sum_{y_n\in\chi}\mu\nu(y_n)=\mathbb{E}_{\mu,\nu}(1)$ which is positive since μ is absolutely continuous respect to ν. This completes the proof. □

Acknowledgements. We thank A. Rada, B. Saussol and S. Vaienti for useful discussions. We also thank the anonymous referee for comments and corrections. This paper is part of RL's PhD thesis, developed under agreement between the University of São Paulo (CAPES and CNPq SWE-236825/2012-7 grants) and UTLN-France (with CNRS, and BREUDS FP7-PEOPLE-2012-IRSES318999 grants). This paper is part of the activities of the Project 'Statistics of extreme events and dynamics of recurrence' FAPESP Process 2014/19805-1. The paper was produced as part of the activities of FAPESP Center for Neuromathematics (grant#2013/07699-0, S. Paulo Research Foundation).

References

[**1**] M. Abadi and L. Cardeno. Renyi entropies and large deviations for the first-match function. *IEEE Trans. Inform. Theory* **61**(4) (2015), 1629–1639.

[**2**] M. Abadi, S. Gallo and E. Rada. The shortest possible return time of β-mixing processes. *IEEE Trans. Inf. Theory* to appear. Available online at http://ieeexplore.ieee.org/document/8052545/.

[**3**] M. Abadi and R. Lambert. The distribution of the short-return function. *Nolinearity* **26**(5) (2013), 1143–1162.

[**4**] M. Abadi and S. Vaienti. Large deviations for short recurrence. *Discrete Contin. Dyn. Syst.* **21** (2008), 729–747.

[**5**] V. Afraimovich, J.-R. Chazottes and B. Saussol. Point-wise dimensions for Poincaré recurrence associated with maps and special flows. *Discrete Contin. Dyn. Syst.* **9**(2) (2003), 263–280.

[**6**] C. Bauckhage, K. Kersting and B. Rastegarpanah. The Weibull as a model of shortest path distributions in random networks, avaliable online at http://snap.stanford.edu/mlg2013/submissions/mlg2013_submission_5.pdf.

[**7**] V. Blondel, J. Guillaume, J. Hendrickx and R. Jungers. Distance distribution in random graphs and application to network exploration. *Phys. Rev. E* **76** (2007), 066101.

[**8**] R. Bowen. Periodic points and measures for Axiom A diffeomorphisms. *Trans. Amer. Math. Soc.* **154** (1971), 377–397.

[**9**] T. Cover and J. Thomas. *Elements of Information Theory*, 2nd edn. Wiley, New York, 1991.

[**10**] N. Haydn and S. Vaienti. The Renyi entropy function and the large deviation of short return times. *Ergod. Th. & Dynam. Sys.* **30**(1) (2010), 159–179.

[**11**] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Math. and its Applications, 54)*. Cambridge University Press, Cambridge, 1995.

[**12**] E. Katzav, M. Nitzan, D. ben-Avraham, P. Krapivsky, R. Kühn, N. Ross and O. Biham. Analytical results for the distribution of shortest path lengths in random networks. *Europhys. Lett.* **111**(2) (2015), 26006.

[**13**] I. Kontoyiannis. Asymptotic recurrence and waiting times for stationary processes. *J. Theoret. Probab.* **11**(3) (1998), 795–811.

[**14**] K. Marton and P. Shields. Almost-sure waiting time results for weak and very weak Bernoulli processes. *Ergod. Th. & Dynam. Sys.* **15** (1995), 951–960.

[**15**] A. Nobel and A. D. Wyner. A recurrence theorem for dependent processes with applications to data compression. *IEEE Trans. Inform. Theory* **38** (1992), 1561–1564.

[**16**] A. Rocha. Substitution operators. *PhD Thesis*, Universidade Federal de Pernambuco, 2009. Avaliable online at http://toomandre.com/alunos/doutorado/andrea/tese-andrea.pdf.

[**17**] B. Saussol, S. Troubetzkoy and S. Vaienti. Recurrence, dimensions and Lyapunov exponents. *J. Stat. Phys.* **106** (2002), 623–634.

[**18**] P. Shields. Waiting times: positive and negative results on the Wyner–Ziv problem. *J. Theoret. Probab.* **6** (1993), 499–519.

[**19**] K. Sigmund. On dynamical systems with the specification property. *Trans. Amer. Math. Soc.* **190** (1974), 285–299.

[**20**] N. Sumi, P. Varandas and K. Yamamoto. Partial hyperbolicity and specification. *Proc. Amer. Math. Soc.* **144** (2016), 1161–1170.

[**21**] A. Ukkonen. Indirect estimation of shortest path distributions with small-world experiments. *Advances in Intelligent Data Analysis XIII: Proceedings of the 13th International Symposium, IDA 2014 Leuven, Belgium, October 30–November 1, 2014 (Lecture Notes in Computer Science, 8819)*. Springer, Cham, 2014, pp. 333–344.

[**22**] P. Varandas. Non-uniform specification and large deviations for weak Gibbs measures. *J. Stat. Phys.* **146**(2) (2012), 330–358.

[**23**] A. Vazquez. Polynomial growth in branching processes with diverging reproduction number. *Phys. Rev. Lett.* **96** (2006),038702.

[**24**] A. Wyner. More on recurrence and waiting times. *Ann. Appl. Probab.* **9**(3) (1999), 780–796.

[**25**] A. Wyner and J. Ziv. Some asymptotic properties of the entropy of a stationary ergodic data source with applications to data compression. *IEEE Trans. Inform. Theory* **35**(6) (1989), 1250–1258.

[**26**] K. Yamamoto. On the weaker forms of the specification property and their applications. *Proc. Amer. Math. Soc.* **137**(11) (2009), 3807–3814.