LIMITING ENTRY AND RETURN TIMES DISTRIBUTION FOR ARBITRARY NULL SETS

NICOLAI HAYDN AND SANDRO VAIENTI

Abstract. We describe an approach that allows us to deduce the limiting return times distribution for arbitrary sets to be compound Poisson distributed. We establish a relation between the limiting return times distribution and the probability of the cluster sizes, where clusters consist of the portion of points that have finite return times in the limit where random return times go to infinity. In the special case of periodic points we recover the known Pólya-Aeppli distribution which is associated with geometrically distributed cluster sizes. We apply this method to several examples the most important of which is synchronisation of coupled map lattices. For the invariant absolutely continuous measure we establish that the returns to the diagonal is compound Poisson distributed where the coefficients are given by certain integrals along the diagonal.

Contents

Date: April 1, 2020.
N Haydn, Mathematics Department, USC, Los Angeles, 90089-2532. E-mail: nhaydn@math.usc.edu.
S Vaienti, Aix Marseille Université, Université de Toulon, CNRS, CPT, UMR 7332, Marseille, France. E-mail: vaienti@cpt.univ-mrs.fr.

1. Introduction

Return times statistics have recently been studied quite extensively. For equilibrium states for Hölder continuous potentials on Axiom A systems in particular, Pitskel [22] showed that generic points have in the limit Poisson distributed return times if one uses cylinder neighbourhoods. In the same paper he also showed that this result applies only almost surely and shows that at periodic points the return times distribution has a point mass at the origin which corresponds to the periodicity of the point. It became clear later that in fact for every non-periodic point the return times are in the limit Poisson distributed while for periodic points the distribution is Pólya-Aeppli which is a Poisson distribution compounded with a geometric distribution of clusters, where the parameter for the geometric distribution is the value given by Pitskel. For ϕ-mixing systems in a symbolic setting, this dichotomy follows from [1]. For more general classes of dynamical systems with various kind of mixing properties, we showed in our paper [15] that limiting return times distributions at periodic points are compound Poissonian; moreover we derived error terms for the convergence to the limiting distribution in many other settings. The paper [19] showed that for all ψ-mixing shifts the limiting distributions of the numbers of multiple recurrencies to shrinking cylindrical neighborhoods of all points are close either to Poisson or to compound Poisson distributions. In the classical setting this dichotomy was shown in [14] using the Chen-Stein method for ϕ-mixing measures, where for cylinder sets the limiting distribution was found to be Poisson at all non-periodic points. Extension to non-uniformly hyperbolic dynamical systems are provided in [9], which establishes and discusses the connection between the laws of Return Times Statistics and Extreme Value Laws (see also the book [12] for a panorama and an account on extreme value theory and point processes applied to dynamical systems). For planar dispersing billiards the return times distribution is, in the limit, Poisson for metric balls almost everywhere w.r.t. the SRB measure: this has been proved in [10]. Convergence in distribution for the rescaled return times in planar billiard has been shown in [23] where the same authors proved that the distribution of the number of visits to a ball with vanishing radius converges to a Poisson distribution for some nonuniformly hyperbolic invertible dynamical systems which are modeled by a Gibbs-Markov-Young tower [24]. Similarly [6] established Poisson approximation for metric balls for systems modelled by a Young tower whose return-time function has a exponential tail and with one-dimensional unstable manifolds, which included the Hénon attractor. For polynomially decaying correlations this was done in [16] where also the restriction on the dimension of the unstable manifold was dropped. In a more geometric setting the limiting distribution for shrinking balls was shown in [17]. Spatio-temporal Poisson processes obtained from recording not only the successive times of visits to a set, but also the positions, have been recently studied in [25]. Another kind of extension has been proposed in [11], which studied marked point processes associated to extremal observations corresponding to exceedances of high thresholds. Finally distributions of return to different sets of cylinders have been recently considered in [20].

In the current paper we look at a more general setting which allows us to find the limiting return times distribution to an arbitrary zero measure set Γ by looking at the return times distribution of a neighbourhood $B_\rho(\Gamma)$ on a time scale suggested by Kac's lemma. For the approximating sets we then show that the return times are close to compound binomially distributions (Theorem 3), which in the limit converges to a compound Poissonian. We show this in a geometric setup that requires that the correlation functions decay at least polynomially. The slowest rate required depends on the regularity of the invariant measure.

We then apply this result to some examples which include the standard periodic point setting. It also allows us to look at coupled map lattices, where the diagonal set is invariant. The return times statistics then expresses the degree to which neighbouring points are synchronised. We will in particular get a more direct and generalisable proof of a result originally got in [8], and we will explain in Remark 10 the related improvements.

In the next section we describe the systems we want to consider and state the main result, Theorem 1. In Section 4 we connect the distribution of the return times functions to the probabilities of the cluster sizes which are the parameters that describe the limiting distribution. Section 6 consists of a very general approximation theorem that allows us to measure how close a return times distribution is to being compound binomial. Section 7 contains the proof of the main result. Section 8 has some examples including the pathological Smith example and standard periodic points. Section 9 deals with coupled map lattices, where the maps that are coupled are expanding interval maps. There we show that for the absolutely continuous invariant measure the parameters for the compound Poisson limiting distribution are given by integrals along the diagonal. In particular one sees that is this case the parameters are in general not geometrical.

2. Compound Poisson Distribution

An integer valued random variable W is compound Poisson distributed if there are i.i.d. integer valued random variables $X_j \geq 1$, $j = 1, 2, \ldots$, and an independent Poisson distributed random variable P so that $W = \sum_{j=1}^{P} X_j$. The Poisson distribution P describes the distribution of clusters whose sizes are described by the random variables X_j whose probability densities are given by values $\lambda_\ell = \mathbb{P}(X_j = \ell)$, $\ell = 1, 2, \ldots$. We then have

$$\mathbb{P}(W = k) = \sum_{\ell=1}^{k} \mathbb{P}(P = \ell)\mathbb{P}(S_\ell = k),$$

where $S_\ell = \sum_{j=1}^{\ell} X_j$ and P is Poisson distributed with parameter s, i.e. $\mathbb{P}(P = \ell) = e^{-s}s^\ell/\ell!$. By Wald's equation $\mathbb{E}(W) = s\mathbb{E}(X_j)$.

We say a probability measure $\tilde{\nu}$ on $\mathbb{N}_0$ is compound Poisson distributed with parameters $s\lambda_\ell$, $\ell = 1, 2, \ldots$, if its generating function $\varphi_{\tilde{\nu}}$ is given $\varphi_{\tilde{\nu}}(z) = \exp \int_0^\infty (z^x - 1)\, d\rho(x)$, where ρ is the measure on $\mathbb{N}$ defined by $\rho = \sum_\ell s\lambda_\ell \delta_\ell$, with δ_ℓ being the point mass at ℓ. If we put $L = \sum_\ell s\lambda_\ell$ then $L^{-1}\rho$ is a probability measure and the random variable $W = \sum_{j=1}^{P} X_j$ is compound Poisson distributed, where P is Poisson distributed with parameter L and

$X_j, j = 1, 2, \dots,$ are i.i.d. random variables with distribution $\mathbb{P}(X_j = \ell) = \lambda_\ell = L^{-1}s\lambda_\ell$, $\ell = 1, 2, \dots$.

In the special case $X_1 = 1$ and $\lambda_\ell = 0 \forall \ell \geq 2$ we recover the Poisson distribution $W = P$. The generating function is then $\varphi_W(z) = \exp(-s(1 - \varphi_X(z)))$, where $\varphi_X(z) = \sum_{\ell=1}^\infty z^\ell \lambda_\ell$ is the generating function of X_j.

An important non-trivial compound Poisson distribution is the Pólya-Aeppli distribution which happens when the X_j are geometrically distributed, that is $\lambda_\ell = \mathbb{P}(X_\ell) = (1-p)p^{\ell-1}$ for $\ell = 1, 2, \dots$, for some $p \in (0, 1)$. In this case

$$\mathbb{P}(W = k) = e^{-s} \sum_{j=1}^k p^{k-j}(1-p)^j \frac{s^j}{j!} \binom{k-1}{j-1}$$

and in particular $\mathbb{P}(W = 0) = e^{-s}$. In the case of $p = 0$ this reverts back to the straight Poisson distribution.

In our context when we count limiting returns to small sets, the Poisson distribution gives the distribution of clusters which for sets with small measure happens on a large timescale as suggested by Kac's formula. The number of returns in each cluster is given by the i.i.d. random variables X_j. These returns are on a fixed timescale and nearly independent of the size of the return set as its measure is shrunk to zero.

3. Assumptions and main results

3.1. **The counting function.** Let M be a manifold and $T : M \to M$ a C^2 local diffeomorphism with the properties described below in the assumptions. We envisage both cases of global invertible maps eventually with singularities and maps which are locally injective on a suitable partition of M. Let μ be a T-invariant Borel probability measure on M.

For a subset $U \subset M$, $\mu(U) > 0$, we define the counting function

$$\xi_U^t(x) = \sum_{n=0}^{\lfloor t/\mu(U) \rfloor} \mathbb{1}_U \circ T^n(x).$$

which tracks the number of visits a trajectory of the point $x \in M$ makes to the set U on an orbit segment of length $N = \lfloor t/\mu(U) \rfloor$, where t is a positive parameter. (We often omit the sub- and superscripts and simply use $\xi(x)$.)

3.2. **The hyperbolic structure and cylinder sets.** Let Γ^u be a collection of unstable leaves γ^u and Γ^s a collection of stable leaves γ^s. We assume that $\gamma^u \cap \gamma^s$ consists of a single point for all $(\gamma^u, \gamma^s) \in \Gamma^u \times \Gamma^s$. The map T contracts along the stable leaves (need not to be uniform) and similarly T^{-1} contracts along the unstable leaves.

For an unstable leaf γ^u denote by μ_{γ^u} the disintegration of μ to the γ^u. We assume that μ has a product like decomposition $d\mu = d\mu_{\gamma^u} d\upsilon(\gamma^u)$, where υ is a transversal measure. That is, if f is a function on M then

$$\int f(x)\, d\mu(x) = \int_{\Gamma^u} \int_{\gamma^u} f(x)\, d\mu_{\gamma^u}(x)\, d\upsilon(\gamma^u)$$

If $\gamma^u, \hat{\gamma}^u \in \Gamma^u$ are two unstable leaves then the holonomy map $\Theta : \gamma^u \to \hat{\gamma}^u$ is defined by $\Theta(x) = \hat{\gamma}^u \cap \gamma^s(x)$ for $x \in \gamma^u$, where $\gamma^u(x)$ be the local unstable leaf through x.

Let us denote by $J_n = \frac{dT^n\mu_{\gamma^u}}{d\mu_{\gamma^u}}$ the Jacobian of the map T^n with respect to the measure μ in the unstable direction.

Let γ^u be a local unstable leaf. Assume there exists $R > 0$ and for every $n \in \mathbb{N}$ finitely many $y_k \in T^n\gamma^u$ so that $T^n\gamma^u \subset \bigcup_k B_{R,\gamma^u}(y_k)$, where $B_{R,\gamma^u}(y) \subset \gamma^u$ is the embedded R-disk centered at y in the unstable leaf γ^u. Denote by $\zeta_{\varphi,k} = \varphi(B_{R,\gamma^u}(y_k))$ where $\varphi \in \mathscr{I}_n$ and $\mathscr{I}_n$ denotes the inverse branches of T^n. We call ζ an n-cylinder. In the case of piecewise expanding endomorphisms in any dimension, we will define an n-cylinder ζ_n as an element of the join partition $\mathcal{A}^n := \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$, where $\mathcal{A}$ is the initial partition into subsets of monotonicity for the map T.

3.3. **Assumptions.** We shall make two sets of assumptions, the first two will be on the map and the properties of the invariant measure per se, while Assumptions (IV), (V) and (VI) will involve the approximating sets of Γ. The sets $\mathcal{G}_n$ account for possible discontinuity sets of the map where the derivative might become singular in a controlled way.

(I) *Overlaps of cylinders:* There exists a constant L so that the number of overlaps $N_{\varphi,k} = |\{\zeta_{\varphi',k'} : \zeta_{\varphi,k} \cap \zeta_{\varphi',k'} \neq \varnothing, \varphi' \in \mathscr{I}_n\}|$ is bounded by L for all $\varphi \in \mathscr{I}_n$ and for all k and n. This follows from the fact that $N_{\varphi,k}$ equals $|\{k' : B_{R,\gamma^u}(y_k) \cap B_{R,\gamma^u}(y_{k'}) \neq \varnothing\}|$ which is uniformly bounded by some constant L. For endomorphisms the analogous requirement will be that there exists $\iota > 0$ such that for any n and any n-cylinder $\zeta_n \in \mathcal{A}^n$ we have $\mu(T^n\zeta_n) > \iota$.

(II) *Decay of correlations:* There exists a decay function $\mathcal{C}(k)$ so that

$$\left| \int_M G(H \circ T^k)\, d\mu - \mu(G)\mu(H) \right| \leq \mathcal{C}(k)\|G\|_{Lip}\|H\|_\infty \qquad \forall k \in \mathbb{N},$$

for functions H which are constant on local stable leaves γ^s of T. The functions $G : M \to \mathbb{R}$ are Lipschitz continuous w.r.t. the given metric on M. In this paper we consider the two standard cases for the decay rate:

(i) $\mathcal{C}$ decays exponentially, that is $\mathcal{C}(k) \lesssim \vartheta^k$ for some $\vartheta \in (0,1)$;

(ii) $\mathcal{C}$ decays polynomially, i.e. $\mathcal{C}(k) \lesssim k^{-p}$ for some $p > 0$.

(III) Assume there are sets $\mathcal{G}_n$ so that

(i) *Non-uniform setsize:* $\mu(\mathcal{G}_n^c) = \mathcal{O}(n^{-q})$ for some positive q.

(ii) *Distortion:* $\frac{J_n(x)}{J_n(y)} = \mathcal{O}(\omega(n))$ for all $x, y \in \zeta$, $\zeta \subset \mathcal{G}_n$ for $n \in \mathcal{N}$, where ζ are n-cylinders in unstable leaves γ^u and $\omega(n)$ is a non-decreasing sequence.

(iii) *Contraction:* There exists a $\kappa > 1$, so that diam $\zeta \leq n^{-\kappa}$ for all n-cylinders $\zeta \in \mathcal{G}_n$ and all n.

Now assume $\Gamma \subset M$ is a zero measure set that is approximated by sets $B_\rho(\Gamma) = \bigcup_{x\in\Gamma} B_\rho(x)$ for small $\rho > 0$ (in the terminology of section 3.1 $U = B_\rho(\Gamma)$). We then make the following assumptions:

(IV) *Dimension:* There exist $0 < d_0 < d_1$ such that $\rho^{d_0} \geq \mu(B_\rho(\Gamma)) \geq \rho^{d_1}$.

(V) *Unstable dimension:* There exists a u_0 so that $\mu_{\gamma^u}(B_\rho(\Gamma)) \leq C_1\rho^{u_0}$ for all $\rho > 0$ small enough and for almost all $x \in \gamma^u$, every unstable leaf γ^u.

(VI) *Annulus type condition:* Assume that for some $\eta, \beta > 0$:

$$\frac{\mu(B_{\rho+r}(\Gamma) \setminus B_{\rho-r}(\Gamma))}{\mu(B_\rho(\Gamma))} = \mathcal{O}(r^\eta \rho^{-\beta})$$

for every $r < \rho_0$ for some $\rho_0 < \rho$ (see remark below).

Here and in the following we use the notation $x_n \lesssim y_n$ for $n = 1, 2, \dots$, to mean that there exists a constant C so that $x_n < Cy_n$ for all n. As before let $T : \Omega \circlearrowleft$ and μ a T-invariant probability measure on Ω. For a subset $U \subset \Omega$ we put $I_i = \mathbb{1}_U \circ T^i$ and define

$$Z^L = Z^L_U = \sum_{i=0}^{2L} I_i$$

where L is a (large) positive integer. If $\Gamma \subset M$ is now a zero measure set, let $t > 0$ and put

$$\lambda_\ell = \lim_{L\to\infty} \lambda_\ell(L), \tag{1}$$

where

$$\lambda_\ell(L) = \lim_{\rho\to 0} \frac{\mathbb{P}(Z^L_{B_\rho(\Gamma)} = \ell)}{\mathbb{P}(Z^L_{B_\rho(\Gamma)} \geq 1)}.$$

Let us now formulate our main result.

Theorem 1. *Assume that the map $T : M \to M$ satisfies the assumptions (I)–(VI) where $\mathcal{C}(k)$ decays at least polynomially with power $p > \frac{\frac{\beta}{\eta} + d_1}{d_0 \wedge u_0'}$, where $u_0' = u_0/(1 + \kappa')$. Moreover we assume that $d_0 > \max\{\frac{d_1}{q-1}, \frac{\beta}{\kappa\eta - 1}\}$ and $\kappa u_0 > 1$. Assume $\omega(j) \lesssim j^{\kappa'}$ for some $\kappa' \in [0, \kappa u_0 - 1)$. Let $\Gamma \subset M$ be a zero measure set and λ_l the corresponding quantity as defined in (1).*

Then

$$\mathbb{P}(\xi^t_{B_\rho(\Gamma)} = k) \longrightarrow \nu(\{k\})$$

as $\rho \to 0$, where ν is the compound Poisson distribution for the parameters $s\lambda_\ell$, where $s = \alpha_1 t$ and $\frac{1}{\alpha_1} = \sum_{k=1}^\infty k\lambda_k$.

Remark 1. *In the classical case when the limiting set consists of a single point, namely $\Gamma = \{x\}$, then we recover the known results which are the two cases when x is a non-periodic point and when x is a periodic point. If x is a non-periodic point then $\lambda_1 = 1$ and $\lambda_\ell = 0$ for $\ell \geq 2$ which implies that the limiting distribution is Poissonian. Previously this was shown in* [6] *for exponentially decaying correlations and in* [16] *for polynomially decaying correlations. Another more general version is given in* [17]*. In the case when x is periodic we obtain that $\lambda_\ell = (1-p)p^{\ell-1}$ for all $\ell = 1, 2, \dots$, and where p is given by the limit $\lim_{\rho\to 0} \frac{\mu(B_\rho(x) \cap T^{-m} B_\rho(x))}{\mu(B_\rho(x))}$ if the limit exists and where m is the minimal period of x. The limiting distribution in this case is Pólya-Aeppli. Pitskel* [22] *obtained this value for equilibrium states for Axiom A systems and a more general description is found in* [15]*. See also section 8.3.*

Remark 2. *Young towers satisfy the conditions of Theorem 1 where the 'bad sets' $\mathcal{G}_n$ account for the rectangles of the partition whose return times are in the tail of the distribution. In the polynomial case one has to make a judicious choice for the cutoff. This scheme, which follows* [16], *is carried out in* [28].

Remark 3. *In* [2], *Theorem 2.5, a similar result was obtained for the extremal values distribution under some assumptions which go back to Leadbetter* [21]. *The corresponding values for λ_ℓ there are obtained by a single suitable limit rather than the double limit used in our setting.*

Remark 4. *Note that in our formulation of the theorem we require that in the decay of correlations, Assumption (II), the speed involves the Lipschitz and $\mathscr{L}^\infty$ norms respectively. This is a weaker requirement than the often times required Lipschitz and $\mathscr{L}^1$ norms which are used for related results in other places. With the $\mathscr{L}^1$ norm instead of the $\mathscr{L}^\infty$ norm for the second function the estimate $\mathcal{R}_2$ of the contribution made by short returns simplifies considerable since it immediately provides the measure of the return set as a factor instead of the factor* 1.

The proof of Theorem 1 is given in Section 7. In the following section we will express the parameters λ_ℓ in terms of the limiting return times distribution.

4. Return times

In this section we want to relate the parameters λ_k which determine the limiting probability of a k-cluster to occur to the return times distribution. To account for a more general setting, let $T : \Omega \circlearrowleft$ be a measurable map on a space Ω. For a subset $U \subset \Omega$ we define the first entry/return time τ_U by $\tau_U(x) = \min\{j \geq 1 : T^j \in U\}$. Similarly we get higher order returns by defining recursively $\tau_U^\ell(x) = \tau_U^{\ell-1} + \tau_U(T^{\tau_U^{\ell-1}}(x))$ with $\tau_U^1 = \tau_U$. We also write $\tau_U^0 = 0$ on U.

Let $U_n \subset \Omega$, $n = 1, 2, \dots$, be a nested sequence of sets and put $\Lambda = \bigcap_n U_n$. For K be a large number which later will go to infinity and put $\hat\alpha_\ell(K, U_n) = \mu_{U_n}(\tau_{U_n}^{\ell-1} \leq K)$, where μ_{U_n} is the induced measure on U_n given by $\mu_{U_n}(A) = \mu(A \cap U_n)/\mu(U_n), \forall A \subset \Omega$. Assume the limits $\hat\alpha_\ell(K) = \lim_{n\to\infty} \hat\alpha_\ell(K, U_n)$, $\ell = 1, 2, \dots$, exist for K large enough. Since $\{\tau_{U_n}^{\ell+1} \leq K\} \subset \{\tau_{U_n}^{\ell} \leq K\}$ we get that $\hat\alpha_\ell(K) \geq \hat\alpha_{\ell+1}(K)$ for all ℓ and in particular $\hat\alpha_1(K) = 1$. By monotonicity the limits $\hat\alpha_\ell = \lim_{K\to\infty} \hat\alpha_\ell(K)$ exist and satisfy $\hat\alpha_1 = 1$ and $\hat\alpha_\ell \geq \hat\alpha_{\ell+1}$ $\forall \ell$.

Now assume that moreover the limits $p_i^\ell = \lim_{n\to\infty} \mu_{U_n}(\tau_{U_n}^{\ell-1} = i)$ of the conditional size of the level sets of the ℓth return time $\tau_{U_n}^\ell$ exist for $i = 0, 1, 2, \dots$ (clearly $p_i^\ell = 0$ for $i \leq \ell - 2$). Then we can formulate the following relation.

Lemma 1. *For $\ell = 2, 3, \dots$:*
$$\hat\alpha_\ell = \sum_i p_i^\ell.$$

Proof. Let $\varepsilon > 0$, then there exists K_1 so that $|\hat\alpha_\ell - \hat\alpha_\ell(K)| < \varepsilon$ for all $K \geq K_1$. Let $K \geq K_1$, then for all small enough U one has $|\hat\alpha_\ell(K) - \mu_U(\tau_U^{\ell-1} \leq K)| < \varepsilon$. Thus $|\hat\alpha_\ell - \mu_U(\tau_U^{\ell-1} \leq K)| < 2\varepsilon$. There exists K_2 so that $\sum_{i=K+1}^\infty p_i^{\ell-1} < \varepsilon$ for all $K \geq K_2$. If we let $K \geq K_0 = K_1 \vee K_2$ then for all small enough U one has $|p_i^\ell - \mu_U(\tau_U^{\ell-1} = i)| < \varepsilon/K$.

Consequently

$$\hat\alpha_\ell=\sum_{i=1}^K\mu_U(\tau_U^{\ell-1}=i)+\mathcal{O}(2\varepsilon)=\sum_{i=1}^K p_i^{\ell-1}+\mathcal{O}(3\varepsilon)=\sum_{i=1}^\infty p_i^{\ell-1}+\mathcal{O}(4\varepsilon).$$

Now let ε go to zero. □

Now put $\alpha_\ell=\lim_{K\to\infty}\alpha_\ell(K)$, where $\alpha_\ell(K)=\lim_{n\to\infty}\mu_{U_n}(\tau_{U_n}^{\ell-1}\le K<\tau_{U_n}^\ell)$ for $\ell=1,2,\dots$. Since $\{\tau_{U_n}^\ell\le K\}\subset\{\tau_{U_n}^{\ell-1}\le K\}$ we get $\{\tau_{U_n}^{\ell-1}\le K<\tau_{U_n}^\ell\}=\{\tau_{U_n}^{\ell-1}\le K\}\setminus\{\tau_{U_n}^\ell\le K\}$. Therefore $\alpha_\ell=\hat\alpha_\ell-\hat\alpha_{\ell+1}$ which in particular implies the existence of the limits α_ℓ. Also, by the previous lemma

$$\alpha_\ell=\sum_i(p_i^{\ell-1}-p_i^\ell)$$

for $\ell=2,3,\dots$. In the special case $\ell=1$ we get in particular $\alpha_1=\lim_{K\to\infty}\lim_{n\to\infty}\mu_{U_n}(K<\tau_{U_n})$. Since $p_0^1=1$ and $p_i^1=0\;\forall\, i\ge1$ we get $\alpha_1=1-\sum_i p_i^2$.

Dropping the index n, let $I_i=\mathbb{1}_U\circ T^i$ be the characteristic function of $T^{-i}U$, then we can define the random variable $Z^L=\sum_{i=0}^{2L}I_i$ and obtain that $\lim_U\mathbb{E}(\mathbb{1}_{Z^L=\ell}|I_0)=\lim_U\mu_U(Z^L=\ell)=\alpha_\ell(L)$ ($\mathbb{E}$ and $\mathbb{P}$ are with respect to the invariant measure μ). With some abuse of notation $2L+1$ here takes the role of K previously.

Now put

$$\lambda_k(L,U)=\mathbb{P}(Z^L=k|Z^L>0)=\frac{\mathbb{P}(Z^L=k)}{\mathbb{P}(Z^L>0)}.$$

For a sequence of sets U_n for which $\mu(U_n)\to0$ as $n\to\infty$ we put $\lambda_k(L)=\lim_{n\to\infty}\lambda_k(L,U_n)$. Evidently $\lambda_k(L,U)\le\lambda_k(L',U)$ if $L\le L'$ and consequently also $\lambda_k(L)\le\lambda_k(L')$. As a result the limit $\lambda_k=\lim_{L\to\infty}\lambda_k(L)$ always exists.

Let us also define $Z^{L,+}=Z_U^{L,+}=\sum_{i=L}^{2L}I_i$ and similarly $Z^{L,-}=Z_U^{L,-}=\sum_{i=0}^{L-1}I_i$. Evidently $Z^L=Z^{L,-}+Z^{L,+}$ and moreover

$$\alpha_k=\lim_{L\to\infty}\lim_{n\to\infty}\mathbb{P}(Z_U^{L,+}=k|I_L=1)$$

which by invariance is equal to $\alpha_k=\lim_{L\to\infty}\lim_{n\to\infty}\mathbb{P}(Z_U^L=k|I_0=1)$. Let us notice that α_1 is commonly called the *extremal index*. Let us define $W^L=\sum_{i=0}^L I_i$. Then $\alpha_k=\lim_{L\to\infty}\lim_{n\to\infty}\mathbb{P}(W^L=k|I_0=1)$.

Lemma 2. *Assume that for all L large enough the limits $\hat\alpha_k(L)=\lim_{n\to\infty}\hat\alpha_k(L,U_n)$ exist along a (nested) sequence of sets U_n, $\mu(U_n)\to0$ as $n\to\infty$. Assume $\sum_{k=1}^\infty k\hat\alpha_k<\infty$ where $\hat\alpha_k=\lim_{L\to\infty}\hat\alpha_k(L)$.*

Then for every $\eta>0$ there exists an L_0 so that for all $L'>L\ge L_0$:

$$\mathbb{P}(W^{L'-L}\circ T^L>0,I_0=1)\le\eta\mu(U_n)$$

and

$$\mathbb{P}(W^L>0,I_{L'}=1)\le\eta\mu(U_n)$$

for all n large enough (depending on L,L').

Proof. (I) To prove the first estimate, let $\varepsilon>0$ and $k\ge1$. Let k_0 be so that $\sum_{k=k_0}^\infty\hat\alpha_k<\varepsilon$ and then L_0 large enough so that $\hat\alpha_k-\hat\alpha_k(L)<\varepsilon/k_0$ for all $L\ge L_0$. Then for all sufficiently large n one has $|\hat{\alpha}_k(L)-\hat{\alpha}_k(L,U_n)|<\varepsilon/k_0$ for all $k\le k_0$. Also, for n large enough we can achieve that $\sum_{k=k_0}^\infty \hat{\alpha}_k(L,U_n)=\sum_{k=k_0}^L \hat{\alpha}_k(L,U_n)\le 2\varepsilon$. From now on $U=U_n$.

Note that $\hat{\alpha}_k(L,U)=\mathbb{P}(W^L\ge k|I_0=1)$ and

$$U\cap\{W^{L'}=k\}\setminus\{W^L=k\}=U\cap T^{-L}\{W^{L'-L}>0\}\cap\{W^{L'}=k\}$$

where $U=\{I_0=1\}$. Consequently

$$\mathbb{P}(I_0=1,W^{L'-L}\circ T^L>0,W^{L'}=k)=\mu(U)(\alpha_k(L',U)-\alpha_k(L,U))$$

and therefore

$$\begin{aligned}
\mathbb{P}(W^{L'-L}\circ T^L>0,I_0=1) &= \sum_{k=1}^\infty \mathbb{P}(I_0=1,W^{L'-L}\circ T^L>0,W^{L'}=k)\\
&= \mu(U)\sum_{k=1}^\infty(\hat{\alpha}_k(L',U)-\hat{\alpha}_k(L,U))\\
&\le \mu(U)\hat{\alpha}_{k_0}(L',U)\\
&\quad +\mu(U)\sum_{k=1}^{k_0-1}\left((\hat{\alpha}_k(L',U)-\hat{\alpha}_{k+1}(L',U))-(\hat{\alpha}_k(L,U)-\hat{\alpha}_{k+1}(L,U))\right)\\
&\le \mu(U)\hat{\alpha}_{k_0}(L',U)+4\mu(U)\sum_{k=1}^{k_0}\hat{\alpha}_k(L',U)\\
&\le 5\varepsilon\mu(U)
\end{aligned}$$

since $\sum_{k=k_0}^\infty \hat{\alpha}_k(L',U)\le\varepsilon$. The first inequality of the lemma now follows if $\varepsilon=\eta/5$.

(II) To prove the second bound let $\varepsilon>0$ and $k\ge 1$. Let k_0 be so that $\sum_{k=k_0}^\infty k\hat{\alpha}_k<\varepsilon$ and then L_0 large enough so that $\hat{\alpha}_k-\hat{\alpha}_k(L)<\varepsilon/k_0$ for all $L\ge L_0$. Then for all sufficiently large n one has $|\hat{\alpha}_k(L)-\hat{\alpha}_k(L,U_n)|<\varepsilon/k_0$ for all $k\le k_0$. Moreover for n large enough we also obtain $\sum_{k=k_0}^\infty \hat{\alpha}_k(L',U_n)=\sum_{k=k_0}^L \hat{\alpha}_k(L',U_n)<2\varepsilon$. Let $U=U_n$ and notice that

$$\mathbb{P}=(W^L>0,I_{L'}=1)=\sum_{k=1}^\infty\mathbb{E}(\mathbb{1}_{W^L=k}I_{L'})=\sum_k\frac{1}{k}\mathbb{E}(\mathbb{1}_{W^L=k}W^LI_{L'})=\sum_k\frac{1}{k}\sum_{i=0}^L\mathbb{E}(\mathbb{1}_{W^L=k}I_iI_{L'})$$

and

$$\bigcup_{i=0}^L\{W^L=k,I_i=1,I_{L'}=1\}=\bigcup_{i=0}^L\bigcup_{\vec{i}\in J^k}\left(C_{\vec{i}}\cap\{I_i=I_{L'}=1\}\right),$$

where

$$J^k=\left\{\vec{i}=(i_1,i_2,\dots,i_k):0\le i_1<i_2<\dots<i_k\le l\right\}$$

and

$$C_{\vec{i}}=\left\{I_{i_j}=1\forall i=j,\dots,k,\quad I_a=0\forall a\in[0,L]\setminus\{i_j:j\}\right\}.$$

Then

$$\begin{aligned}\bigcup_{i=0}^{L}\{W^L=k\}\cap\{I_i=I_{L'}=1\} &= \bigcup_{j=1}^{k}\bigcup_{\vec{i}\in J^k}(C_{\vec{i}}\cap\{I_i=I_{L'}=1\})\\ &= \bigcup_{j=1}^{k}\bigcup_{p=0}^{L}T^{-p}\left(\bigcup_{\vec{i}\in J_p^k(j)}(C_{\vec{i}}\cap\{I_0=I_{L'-p}=1\})\right),\end{aligned}$$

where

$J_p^k(j)=\left\{\vec{i}=(i_1,\dots,i_k):-p\le i_1<\cdots<i_k\le L-p,\;\; i_j=p,\;\; I_a=0\forall a\in[-p,L-p]\setminus\{i_j:j\}\right\}$

(put $J_p^k(j)=\varnothing$ if either $p<j$ or $p>L-j$). Consequently

$$\{W^L>0,I_{L'}=1\}=\bigcup_{p=0}^{L}T^{-p}\left(\bigcup_{k=1}^{\infty}\bigcup_{j=1}^{k}\bigcup_{\vec{i}\in J_p^k(j)}(C_{\vec{i}}\cap\{I_0=I_{L'-p}=1\})\right)$$

where the triple union inside the brackets is a disjoint union. Thus

$$\begin{aligned}\mathbb{P}(W^L>0,I_{L'}=1) &\le \sum_{p=0}^{L}\mathbb{E}(I_0I_{L'-p})\\ &= \mathbb{E}(W^L\circ T^{L'-L}I_0)\\ &\le k_0\mathbb{P}(W^L\circ T^{L'-L}>0,I_0=1)+\sum_{k=k_0}^{\infty}k\mathbb{P}(W^L\circ T^{L'-L}=k,I_0=1)\\ &\le k_0 5\varepsilon\mu(U)+\sum_{k=k_0}^{\infty}k\hat{\alpha}_k(L',U)\\ &\le 7\varepsilon\mu(U)\end{aligned}$$

where we used the estimate from Part (I). Now put $\varepsilon=\eta/7$. □

Theorem 2. *Let $U_n\subset\Omega$ be a nested sequence so that $\mu(U_n)\to0$ as $n\to\infty$. Assume that the limits $\hat{\alpha}_\ell(L)=\lim_{n\to\infty}\hat{\alpha}_\ell(L,U_n)$ exist for $\ell=1,2,\dots$ and L large enough. Put $\hat{\alpha}_\ell=\lim_{L\to\infty}\hat{\alpha}_\ell(L)$ and assume $\sum_\ell\ell\hat{\alpha}_\ell<\infty$, then*

$$\lambda_k=\frac{\alpha_k-\alpha_{k+1}}{\alpha_1}$$

where $\alpha_k=\hat{\alpha}_k-\hat{\alpha}_{k+1}$. In particular the limit defining λ_k exists.

Proof. Let $\varepsilon>0$ then there exists k_0 so that $\sum_{\ell=k_0}^{\infty}\ell\hat{\alpha}_\ell<\varepsilon$. Moreover there exists L_0 so that $|\hat{\alpha}_\ell-\hat{\alpha}_\ell(L)|<\varepsilon/k_0$ for all $L\ge L_0$ and $\ell\in[1,k_0]$. For n large enough we also have $|\hat{\alpha}_\ell(L)-\hat{\alpha}_\ell(L,U_n)|<\varepsilon/k_0$. In the following we will often write U for U_n.

Let $L'>L$, then

$$\mathbb{P}(Z^{L'}=k)=\frac{1}{k}\mathbb{E}(\mathbb{1}_{Z^{L'}=k}Z^{L'})=\frac{1}{k}\sum_{i=0}^{2L'}\mathbb{E}(\mathbb{1}_{Z^{L'}=k}\mathbb{1}_{I_i=1}).$$

For $i \in [L, 2L' - L]$ put

$$D_i^{L,L'} = \left\{ \sum_{b=i+L+1}^{L'} I_b \geq 1,\ I_i = 1 \right\}.$$

By Lemma 2 $\mu(D_i^{L,L'}) = \mathcal{O}(\eta\mu(U))$ for L big enough and n large enough, where $\eta > 0$ will be chosen below. Let $k \geq 1$, then

$$\left\{W^{i+L} = k,\ I_i = 1\right\} \cap \left(D_i^{L,L'}\right)^c \subset \left\{Z^{L'} = k,\ I_i = 1\right\}$$

and also

$$\{Z^{L'} = k,\ I_i = 1\} \subset \{W^{i+L} = k,\ I_i = 1\} \cup D_i^{L,L'}.$$

These two inclusions imply

$$\mathbb{P}(Z^{L'} = k,\ I_i = 1) = \mathbb{P}\big(W^{i+L} = k,\ I_i = 1\big) + \mathcal{O}(\eta\mu(U)).$$

Put

$$R_{k,\ell}^{i,L} = \left\{ \sum_{b=i}^{i+L} I_b = k - \ell,\ W^{i-1} = \ell,\ I_i = 1 \right\}$$

for the set of k-clusters that have ℓ occurrences to the 'left' of i. Then

$$R_{k,\ell}^{i,L}(j) = R_{k,\ell}^{i,L} \cap \{I_{i-j} = 1,\ I_a = 0 \forall a = 0, \ldots, i - j - 1\}$$

denotes all those k-clusters which have ℓ occurrences to the left of i the first one of which occurs j steps to the left of i. Evidently, $R_{k,\ell}^{i,L} = \bigcup_{j=1}^{i} R_{k,\ell}^{i,L}(j)$ is a disjoint union. Let us note that the set

$$F^{i-\frac{L}{2}} = \left\{W^{i-\frac{L}{2}} > 0,\ I_i = 1\right\}$$

has by Lemma 2 measure $\mathcal{O}(\eta\mu(U))$. Then for every ℓ we obtain the inclusion

$$R_{k,0}^{i,L} \cap \left(F^{i-\frac{L}{2}}\right)^c \subset \bigcup_{j=i-\frac{L}{2}}^{i-1} T^{-j} R_{k,\ell}^{i,L}(j) \subset R_{k,0}^{i,L} \cup D_i^{L,L'} \cup F^{i-\frac{L}{2}}$$

where the union over j is a disjoint union since $T^{-j} R_{k,\ell}^{i,L}(j) \cap T^{-j'} R_{k,\ell}^{i,L}(j') = \varnothing$ if $j \neq j'$. Thus for every $\ell = 0, \ldots, k - 1$:

$$\mu\left(\bigcup_{j=i-\frac{L}{2}}^{i-1} T^{-j} R_{k,\ell}^{i,L}(j)\right) = \mu\left(R_{k,0}^{i,L}\right) + \mathcal{O}(\eta\mu(U))$$

and since the union is disjoint this implies

$$\sum_{j=i-\frac{L}{2}}^{i-1} \mu(R_{k,\ell}^{i,L}(j)) \leq \mu(R_{k,\ell}^{i,L}) \leq \sum_{j=i-\frac{L}{2}}^{i-1} \mu(R_{k,\ell}^{i,L}(j)) + \mu(F^{i-\frac{L}{2}})$$

from which we conclude that

$$\mu(R_{k,\ell}^{i,L}) = \mu(R_{k,0}^{i,L}) + \mathcal{O}(\eta\mu(U)) = \mu(R_{k,0}^{L,L}) + \mathcal{O}(\eta\mu(U))$$

where the last step is due to invariance. Therefore

$$\begin{aligned}\mathbb{P}(Z^{L'}=k) &= \frac{1}{k}\left(\sum_{i=L}^{2L'-L}\sum_{\ell=0}^{k-1}\Big(\mu(R^{i,L}_{k,\ell})+\mathcal{O}(\eta\mu(U))\Big)+\mathcal{O}(2L\mu(U))\right)\\ &= 2L'\left(1-\frac{L}{L'}\right)\Big(\mu(R^{L,L}_{k,0})+\mathcal{O}(\eta\mu(U))\Big)+\mathcal{O}(L\mu(U)),\end{aligned}$$

In a similar way let us put

$$S^{i,L}_{k,\ell}(j)=R^{i,L}_{k,\ell}\cap\{I_{i-j}=1,\ I_a=0\forall a\in(i-j,i)\}$$

for the set k-clusters which have ℓ occurrences to the left of i the last one of which occurs j steps to the left of i. As before we obtain

$$R^{i,L}_{k,\ell-1}\cap\Big(R^{i-\frac{L}{2}}\Big)^c\subset\bigcup_{j=i-\frac{L}{2}}^{i-1}T^{-j}S^{i,L}_{k,\ell}(j)\subset R^{i,L}_{k,\ell-1}\cup D^{L,L'}_i\cup F^{i-\frac{L}{2}}$$

and therefore conclude that

$$\mu(R^{i.L}_{k,\ell})=\mu(R^{i.L}_{k,\ell-1})+\mathcal{O}(\eta\mu(U)). \tag{2}$$

Since

$$\mathbb{P}(Z^{L,+}=k,I_L=1)=(1+\mathcal{O}(\varepsilon))\mu(U)\alpha_k$$

we obtain

$$\begin{aligned}\alpha_k(L,U)-\alpha_{k+1}(L,U) &= (1+\mathcal{O}(\varepsilon))\mu(U)^{-1}\big(\mathbb{P}(Z^{L,+}=k,I_L=1)-\mathbb{P}(Z^{L,+}=k+1,I_L=1)\big)\\ &= (1+\mathcal{O}(\varepsilon))\mu(U)^{-1}\sum_{\ell=0}^{\infty}\Big(\mu(R^{L,L}_{k+\ell,\ell})-\mu(R^{L,L}_{k+1+\ell,\ell})\Big)\\ &= (1+\mathcal{O}(\varepsilon))\mu(U)^{-1}\sum_{\ell=0}^{k_0}\Big(\mu(R^{L,L}_{k+\ell,\ell})-\mu(R^{L,L}_{k+1+\ell,\ell+1})+\mathcal{O}(\eta\mu(U))\Big)\\ &\quad+\mathcal{O}(\mu(U)^{-1})\sum_{\ell=k_0+1}^{\infty}\Big(\mu(R^{L,L}_{k+\ell,\ell})+\mu(R^{L,L}_{k+1+\ell,\ell})\Big)\end{aligned}$$

In order to estimate the tail sum $\sum_{\ell=k_0}^{\infty}\mu(R^{L,L}_{k+\ell,\ell})$ we first notice that

$$T^{-j}R^{L,L}_{k+\ell,\ell}(j)\cap T^{-j'}R^{L,L}_{k+\ell',\ell'}(j')=\varnothing$$

if $j=j',\ell\neq\ell'$ and also in the case when $j\neq j'$ and $|\ell'-\ell|>k$. To see the latter, assume $j'>j$ and $T^{-j}R^{L,L}_{k+\ell,\ell}(j)\cap T^{-j'}R^{L,L}_{k+\ell',\ell'}(j')\neq\varnothing$ then the occurrences in $[i,i+j)$ are identical in both sets. Moreover, since the occurrences in $[i+j,i+j')$ are identical this forces not only $\ell'\geq\ell$ but also that $\ell'-\ell\leq k$ since $T^{-j}R^{L,L}_{k+\ell,\ell}(j)$ has exactly k occurrences on $[i+j,i+k)$. (There are $k-(\ell'-\ell)$ occurrences on $[i+j',i+j+k]$ and for $T^{-j'}R^{L,L}_{k+\ell',\ell'}(j')$ there are $\ell'-\ell$ occurrences on $(i+j+k,i+j'+k]$.) If we choose an integer $k'>k$ then for every $p = 0, 1, \dots, k' - 1$ one has

$$\bigcup_{j=1}^{i} T^{-j} \bigcup_{s=\frac{k_0}{k'}}^{\infty} R^{L,L}_{k+sk'+p,sk'+p}(j) \subset T^{-L}\left\{W^{2L} \geq k_0 + k, I_0 = 1\right\}$$

where the double union on the left hand side is disjoint. Therefore

$$\sum_{s=\frac{k_0}{k'}}^{\infty} \mu(R^{L,L}_{k+sk'+p,sk'+p}) \leq \mathbb{P}(W^{2L} \geq k_0 + k, I_0 = 1) = \mu(U)\hat{\alpha}_{k_0+k}(2L, U).$$

and consequently

$$\sum_{\ell=k_0}^{\infty} \mu(R^{L,L}_{k+\ell,\ell}) \leq k'\mu(U)\hat{\alpha}_{k_0+k}(2L, U).$$

The same estimate also applies to the tail sum of $\mu(R^{L,L}_{k+1+\ell,\ell})$.

This gives us

$$\mu(R^{L,L}_{k,0}) = (1 + \mathcal{O}(\varepsilon))\mu(U)(\alpha_k(L) - \alpha_{k+1}(L)) + \mathcal{O}(k_0\eta\mu(U)) + k'\mu(U)\hat{\alpha}_{k_0+k}(2L, U).$$

If we choose $\eta = \varepsilon/k_0$, $k' = k_0 + k$ and $L' = L^\gamma$ for some $\gamma > 1$, then

$$\begin{aligned}\mathbb{P}(Z^{L^\gamma} = k) &= 2L^\gamma\mu(U)\big(\left(1 - L^{1-\gamma}\right)(1 + \mathcal{O}(\varepsilon))(\alpha_k(L, U) - \alpha_{k+1}(L, U)) \\ &\qquad + \mathcal{O}(\varepsilon) + \mathcal{O}(L^{1-\gamma}) + (k_0 + k)\hat{\alpha}_{k_0+k}(2L, U)\big)\,,\end{aligned}$$

Without loss of generality we can assume that L is large enough so that $L^{1-\gamma} < \varepsilon$. Then

$$\begin{aligned}\mathbb{P}(Z^{L^\gamma} > 0) &= \sum_{k=1}^{\infty} \mathbb{P}(Z^{L^\gamma} = k) \\ &= 2L^\gamma(1 + \mathcal{O}(\varepsilon))\mu(U)\left(\sum_{k=1}^{k_0}(\alpha_k(L, U) - \alpha_{k+1}(L, U) + \mathcal{O}(\varepsilon)) + \sum_{\ell=k_0}^{\infty} \ell\hat{\alpha}_\ell(2L, U)\right) \\ &= 2L^\gamma(1 + \mathcal{O}(\varepsilon))\mu(U)(\alpha_1(L, U) + \mathcal{O}(\varepsilon))\end{aligned}$$

where the tail sum on the RHS is estimated by 2ε. Hence

$$\mathbb{P}(Z^{L^\gamma} > 0) = 2L^\gamma(1 + \mathcal{O}(\varepsilon))\mu(U)(\alpha_1(L, U) + \mathcal{O}(\varepsilon)).$$

Combining the two estimates yields

$$\lambda_k(L^\gamma, U_n) = \frac{\mathbb{P}(Z^{L^\gamma} = k)}{\mathbb{P}(Z^{L^\gamma} > 0)} = (1 + \mathcal{O}(\varepsilon))\frac{\alpha_k(L, U_n) - \alpha_{k+1}(L, U_n) + \mathcal{O}(\varepsilon)}{\alpha_1(L, U_n) + \mathcal{O}(\varepsilon)}.$$

Letting $\varepsilon \to 0$ implies $L \to \infty$ and consequently $\mu(U_n) \to 0$ as $n \to \infty$ let us finally obtain (as $\gamma > 1$) as claimed $\lambda_k = (\alpha_k - \alpha_{k+1})/\alpha_1$. □

Remark 5. *Under the assumption of Theorem 2 the expected length of the clusters is*

$$\sum_{k=1}^{\infty} k\lambda_k = \frac{1}{\alpha_1}\sum_{k=1}^{\infty} k(\alpha_k - \alpha_{k+1}) = \frac{1}{\alpha_1}$$

which is the reciprocal of the extremal index α_1. Let us note that Smith [27] gave an example where α_1^{-1} is not the expected cluster length and which, naturally enough, does not satisfy the condition of Theorem 2.

Remark 6. *Since $\lambda_k \geq 0$ we conclude that $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \cdots$ is a decreasing sequence. It is moreover easy to see that $\lambda_k = \alpha_k \forall k$ only when both are geometrically distributed, i.e. when $\lambda_k = \alpha_k = \alpha_1(1-\alpha_1)^{k-1}$. Also notice that the condition $\sum_k k\hat{\alpha}_k < \infty$ of the theorem is equivalent to $\sum_k k^3\lambda_k < \infty$ or $\sum_k k^2\alpha_k < \infty$.*

Corollary 1. *For every $\eta >$ one has*

$$\left|\mathbb{P}(Z_i^{L,-} = k, Z_i^{L,+} = \ell - k, I_i = 1) - \mathbb{P}(Z_i^{L,-} = k', Z_i^{L,+} = \ell - k', I_i = 1)\right| \leq \eta\mu(U_n)$$

for all $0 \leq k, k' < \ell$, provided L and n are large enough.

Proof. This follows from (2) as $\mathbb{P}(Z_i^{L,-} = k, Z_i^{L,+} = \ell - k, I_i = 1) = \mu(R_{k,\ell}^{i,L})$. □

5. Entry times

Let us consider the entry time $\tau_U(x)$ where $x \in \Omega$.

Lemma 3. *Let $U_n \subset \Omega$ be a nested sequence so that $\mu(U_n) \to 0$ as $n \to \infty$. Assume that the limits $\hat{\alpha}_\ell(L) = \lim_{n\to\infty} \hat{\alpha}_\ell(L, U_n)$ exist for $\ell = 1, 2, \ldots$ and L large enough. Assume $\sum_{\ell=1}^\infty \hat{\alpha}_\ell < \infty$,*

Then

$$\lim_{L\to\infty}\lim_{n\to\infty} \frac{\mathbb{P}(\tau_{U_n} \leq L)}{L\mu(U_n)} = \alpha_1.$$

Proof. If we write again U for U_n then

$$\mathbb{P}(\tau_U \leq L) = \mu\left(\bigcup_{j=0}^L T^{-j}U\right) = L\mu(U) - \sum_{\ell=2}^L (\ell-1)\mathbb{P}(\tau_U^\ell \leq L < \tau_U^{\ell+1})$$

since every $x \in \{\tau_U^\ell \leq L < \tau_U^{\ell+1}\}$ there exist exactly ℓ entry times $1 \leq j_1 < j_2 < \cdots j_\ell \leq L$ so that $x \in T^{-j_i}U$, $i = 1, \ldots, \ell$ and $x \notin T^{-j}U$ otherwise which means that in the principal term x is counted ℓ times of which we have to remove $\ell - 1$ over counts. If we put $Z^L = \sum_{i=0}^L \mathbb{1}_U \circ T^i$, then

$$\mathbb{P}(\tau_U^\ell \leq L < \tau_U^{\ell+1}) = \mathbb{P}(Z^L = \ell) = \lambda_\ell(L,U)\mathbb{P}(Z^L \geq 1) = \lambda_\ell(L,U)\mathbb{P}(\tau_U \leq L)$$

where, as before, we put $\lambda_\ell(L,U) = \mathbb{P}(Z^L = \ell | Z^L \geq 1)$. Similarly we put $\lambda_\ell(L) = \lim_{n\to\infty}\lambda_\ell(L,U_n)$ and $\lambda_\ell = \lim_{L\to\infty}\lambda_\ell(L)$.

Let $\varepsilon > 0$, then there exists ℓ_0 so that $\sum_{\ell=\ell_0}^\infty \hat{\alpha}_\ell + \ell_0\hat{\alpha}_{\ell_0} < \varepsilon$. By monotonicity also $\sum_{\ell=\ell_0}^\infty \ell\hat{\alpha}_\ell(L) + \ell_0\hat{\alpha}_{\ell_0}(L) < \varepsilon$ for all L. Then for each L there exists $N_1(\varepsilon, L)$ so that $\sum_{\ell=\ell_0}^L \ell\hat{\alpha}_\ell(L,U_n) + \ell_0\hat{\alpha}_{\ell_0}(L,U_n) < 2\varepsilon$ for all $n \geq N_1$. We decompose λ_ℓ as follows

$$\lambda_\ell(L,U_n)\mathbb{P}(\tau_{U_n} \leq L) = \sum_{j=1}^L \mathbb{P}(Z^L = \ell, \tau_{U_n} = j)$$

where

$$\mathbb{P}(Z^L = \ell, \tau_{U_n} = j) \leq \mathbb{P}(Z^{L-j} = \ell, U_n) = \alpha_\ell(L-j, U_n)\mu(U_n).$$

Since $\alpha_\ell = \hat\alpha_\ell - \hat\alpha_{\ell+1}$ we therefore obtain

$$\begin{aligned}\sum_{\ell=\ell_0}^{L} \ell\lambda_\ell(L,U_n)\mathbb{P}(\tau_{U_n}\le L) &\le \mu(U_n)\sum_{j=1}^{L}\sum_{\ell=\ell_0}^{L}\ell\alpha_\ell(L-j,U_n)\\ &\le L\mu(U_n)\left(\sum_{\ell=\ell_0+1}^{L}\hat\alpha_\ell(L-j,U_n)+\ell_0\hat\alpha_\ell(L-j,U_n)-(L+1)\hat\alpha_L(L-j,U_n)\right)\\ &\le L\mu(U_n)\left(\sum_{\ell=\ell_0+1}^{L}\hat\alpha_\ell(L,U_n)+\ell_0\hat\alpha_\ell(L,U_n)\right)\\ &\le 2\varepsilon L\mu(U_n)\end{aligned}$$

for all $n \ge N_1$. Furthermore, there exists $L_0(\varepsilon)$ so that $|\lambda_\ell - \lambda_\ell(L)| < \varepsilon\ell_0^{-2}$ for all $\ell < \ell_0$ and $L \ge L_0$. In addition, for every $L \ge L_0$ there exists an $N_2(\varepsilon, L)$ so that $|\lambda_\ell(L) - \lambda_\ell(L, U_n)| < \varepsilon\ell_0^{-2}$ for $\ell < \ell_0$ and for all $n \ge N_2$. Therefore

$$\left|\sum_{\ell=1}^{\ell_0-1}(\ell-1)\lambda_\ell(L,U_n)-\sum_{\ell=1}^{\ell_0-1}(\ell-1)\lambda_\ell\right|<2\varepsilon$$

for all $L \ge L_0$ and all $n \ge N_2$. Combining the two estimates we obtain that for all $L \ge L_0$ and $n \ge N_0(\varepsilon, L) = N_1 \vee N_2$:

$$\begin{aligned}\frac{\mathbb{P}(\tau_{U_n}\le L)}{L\mu(U_n)}\sum_{\ell=2}^{L}(\ell-1)\lambda_\ell(L,U_n) &= \frac{\mathbb{P}(\tau_{U_n}\le L)}{L\mu(U_n)}\left(\sum_{\ell=1}^{\infty}(\ell-1)\lambda_\ell+\mathcal{O}(2\varepsilon)\right)+\mathcal{O}(2\varepsilon)\\ &= \frac{\mathbb{P}(\tau_{U_n}\le L)}{L\mu(U_n)}\left(\frac{1}{\alpha_1}-1+\mathcal{O}(2\varepsilon)\right)+\mathcal{O}(2\varepsilon).\end{aligned}$$

We finally end up with the identity

$$\frac{\mathbb{P}(\tau_{U_n}\le L)}{L\mu(U_n)}=1-\left(\frac{1}{\alpha_1}-1+\mathcal{O}(2\varepsilon)\right)\frac{\mathbb{P}(\tau_{U_n}\le L)}{L\mu(U_n)}+\mathcal{O}(2\varepsilon)$$

from which the statement of the lemma follows as we let ε go to zero which implies $n \to \infty$ and then let $L \to \infty$. □

Remark 7. *It now follows from the lemma and its proof that*

$$\lim_{L\to\infty}\lim_{n\to\infty}\frac{\mathbb{P}(\tau_{U_n}^{\ell}\le L<\tau_{U_n}^{\ell+1})}{L\mu(U_n)}=\alpha_1\lambda_\ell$$

for $\ell = 1, 2, 3, \ldots$ *. In a similar way as in the previous lemma on can show for* $\ell = 2, 3, \ldots$ *that*

$$\mathbb{P}(\tau_{U_n}^{\ell}\le L)=\sum_{k=\ell}^{L}\mathbb{P}(\tau_{U_n}^{k}\le L<\tau_{U_n}^{k+1})=\mathbb{P}(\tau_{U_n}\le L)\sum_{k=\ell}^{L}\lambda_k(L,U_n)$$

which implies as before that

$$\lim_{L\to\infty}\lim_{n\to\infty}\frac{\mathbb{P}(\tau_{U_n}^{\ell}\le L)}{L\mu(U_n)}=\alpha_\ell.$$

6. The Compound Binomial Distribution

In this section we prove an approximation theorem that provides an estimate how closely the level sets of the counting function W is approximated by a compound binomial distribution which represents the independent case. As the measure of the approximating target set $B_\rho(\Gamma)$ goes to zero, the compound binomial distribution then converges to a compound Poisson distribution.

To be more precise, the following abstract approximation theorem which establishes the distance between sums of $\{0,1\}$-valued dependent random variables X_n and a random variable that has a compound Binomial distribution is used in Section 7.1 in the proof of Theorem 1 to compare the number of occurrences in a finite time interval with the number of occurrences in the same interval for a compound binomial process.

Let Y_j be $\mathbb{N}$ valued i.i.d. random variables and denote $\lambda_\ell = \mathbb{P}(Y_j = \ell)$. Let N be a (large) positive integer, $s > 0$ a parameter and put $p = s/N$. If Q is a binomially distributed random variable with parameters (N,p), that is $\mathbb{P}(Q = k) = \binom{N}{k}p^k(1-p)^{N-k}$, then $W = \sum_{i=1}^{Q} Y_i$ is compound binomially distributed. The generating function of W is $\varphi_W(z) = (p(\varphi_{Y_1}(z) - 1) + 1)^N$, where $\varphi_{Y_1}(z) = \sum_{\ell=0}^\infty z^\ell\lambda_\ell$ is the generating function of Y_1. As N goes to infinity, Q converges to a Poisson distribution with parameter s and W converges to a compound Poisson distribution with parameters $s\lambda_\ell$. In particular $\varphi_W(z) \to \exp s(\varphi_{Y_1}(z) - 1)$. (In the following theorem we assume for simplicity's sake that N' and Δ are integers.)

Theorem 3. *Let $(X_n)_{n\in\mathbb{N}}$ be a stationary $\{0,1\}$-valued process and $W = \sum_{i=0}^N X_i$ for some (large) integer N. Let K, Δ be positive integers so that $\Delta(2K+1) < N$ and define $Z = \sum_{i=0}^{2K} X_i$ and $W_a^b = \sum_{i=a}^b X_i$ ($W = W_0^N$). Let $\tilde\nu$ be the compound binomial distribution measure where the binomial part has values $p = \mathbb{P}(Z \ge 1)$ and $N' = N/(2K+1)$ and the compound part has probabilities $\lambda_\ell = \mathbb{P}(Z = \ell)/p$. Then there exists a constant C_3, independent of K and Δ, such that*

$$|\,\mathbb{P}(W = k) - \tilde\nu(\{k\})\,| \le C_3(N'(\mathcal{R}_1 + \mathcal{R}_2) + \Delta\mathbb{P}(X_0 = 1)),$$

where

$$\mathcal{R}_1 = \sup_{\substack{0<\Delta<M\le N'\\ 0<q<N'-\Delta-1/2}} \left|\sum_{u=1}^{q-1}\left(\mathbb{P}\left(Z = u \wedge W_{\Delta(2K+1)}^{M(2K+1)} = q-u\right) - \mathbb{P}(Z = u)\mathbb{P}\left(W_{\Delta(2K+1)}^{M(2K+1)} = q-u\right)\right)\right|$$

$$\mathcal{R}_2 = \sum_{n=2}^{\Delta}\mathbb{P}(Z \ge 1 \wedge Z\circ T^{(2K+1)n} \ge 1).$$

Proof. Let us assume for simplicity's sake that N is a multiple of $2K+1$ and put $N' = N/(2K+1)$. Now put $Z_j = \sum_{i=j(2K+1)}^{(2K+1)-1} X_i = Z\circ T^{j(2K+1)}$ for $j = 0,1,\dots,N'$. Thus $V = \sum_{i=0}^N X_i = \sum_{j=0}^{N'} Z_j$. Let $(\tilde Z_j)_{j\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables taking values in $\mathbb{N}_0$ which have the same distribution as Z_j. Moreover let us put $V_k^\ell = \sum_{j=k}^\ell Z_j$ and similarly $\tilde V_k^\ell = \sum_{j=k}^\ell \tilde Z_j$. We have to estimate the following quantity:

$$\mathbb{P}(V_0^{N'} = k) - \mathbb{P}(\tilde V_0^{N'} = k) = \sum_{j=0}^{N'-1} D_j(k),$$

where

$$\begin{aligned} D_j(k) &= \mathbb{P}(\tilde V_0^{j-1} + V_j^{N'} = k) - \mathbb{P}(\tilde V_0^{j} + V_{j+1}^{N'} = k) \\ &= \sum_{\ell=0}^{k} \mathbb{P}(\tilde V_0^{j-1} = \ell)\Big(\mathbb{P}(V_j^{N'} = k-\ell) - \mathbb{P}(\tilde Z_j + V_{j+1}^{N'} = k-\ell)\Big) \end{aligned}$$

By invariance it suffices to estimate

$$\mathbb{P}(V_0^M = q) - \mathbb{P}(\tilde Z_0 + V_1^M = q) = \sum_{u=0}^{q} \mathcal{R}(u)$$

for every $M \le N'$ and q, where

$$\mathcal{R}(u) = \mathbb{P}(Z_0 = u, V_1^M = q-u) - \mathbb{P}(\tilde Z_0 = u)\mathbb{P}(V_1^M = q-u).$$

Let us first single out the terms $u = q$ and $u = 0$. For $u = 0$ we see that

$$\mathbb{P}(Z_0 = 0, V_1^M = q) = \mathbb{P}(V_1^M = q) - \mathbb{P}(Z_0 \ge 1, V_1^M = q)$$

and

$$\mathbb{P}(Z_0 = 0)\mathbb{P}(V_1^M = q) = \mathbb{P}(V_1^M = q) - \mathbb{P}(Z_0 \ge 1)\mathbb{P}(V_1^M = q).$$

Consequently

$$\begin{aligned} \mathcal{R}(0) &= \mathbb{P}(Z_0 = 0, V_1^M = q) - \mathbb{P}(\tilde Z_0 = 0)\mathbb{P}(V_1^M = q) \\ &= \mathbb{P}(Z_0 \ge 1)\mathbb{P}(V_1^M = q) - \mathbb{P}(Z_0 \ge 1, V_1^M = q) \\ &\le \sum_{u=1}^{q} \mathcal{R}(u). \end{aligned}$$

Similarly one obtains for $u = q$:

$$\mathcal{R}(q) = \mathbb{P}(Z_0 = q)\mathbb{P}(V_1^M \ge 1) - \mathbb{P}(Z_0 = q, V_1^M \ge 1).$$

This implies that

$$|\mathcal{R}| \le 4 \sum_{u=1}^{q-1} |\mathcal{R}(u)|.$$

In order to estimate $|\mathcal{R}(u)|$ for $u = 1, 2, \ldots, q-1$ let $0 \le \Delta < M$ be the length of the gap we will now introduce. Then

$$|\mathcal{R}(u)| \le \mathcal{R}_1(u) + \mathcal{R}_2(u) + \mathcal{R}_3(u),$$

where

$$\mathcal{R}_1 = \max_{\substack{\Delta < M \le N' \\ q}} \left| \sum_{u=1}^{q-1} \Big(\mathbb{P}(Z_0 = u, V_\Delta^M = q-u) - \mathbb{P}(\tilde Z_0 = u)\mathbb{P}(V_\Delta^M = q-u) \Big) \right|.$$

The other two terms $\mathcal{R}_2$ and $\mathcal{R}_3$ account for opening a 'gap'. More precisely

$$\mathcal{R}_2(u) = \Big| \mathbb{P}(Z_0 = u, V_\Delta^M = q-u) - \mathbb{P}(Z_0 = u, V_1^M = q-u) \Big|$$

and since

$$\{Z_0=u,W_1^M=q-u\}\setminus\{Z_0=u,W_\Delta^M=q-u\}\subset\{Z_0=u,W_1^{\Delta-1}\ge1\}$$

we get therefore

$$\sum_{u=1}^{q-1}\mathcal{R}_2(u)\le\mathbb{P}(Z_0\ge1,W_1^{\Delta-1}\ge1).$$

For the third term we get

$$\mathcal{R}_3(u)\;=\;\mathbb{P}(\tilde Z_0=u)\left|\mathbb{P}(V_1^M=q-u)-\mathbb{P}(V_\Delta^M=q-u)\right|.$$

To estimate $\mathcal{R}_3$ observe that $(q'=q-u)$

$$\mathbb{P}(V_1^M=q')=\mathbb{P}(Z_1\ge1,V_1^M=q')+\mathbb{P}(Z_1=0,V_1^M=q')$$

where

$$\mathbb{P}(Z_1=0,V_1^M=q')=\mathbb{P}(Z_1=0,V_2^M=q')=\mathbb{P}(V_2^M=q')-\mathbb{P}(Z_1\ge1,V_2^M=q').$$

Hence

$$\mathbb{P}(V_1^M=q')-\mathbb{P}(V_2^M=q')=\mathbb{P}(Z_1\ge1,V_1^M=q')-\mathbb{P}(Z_1\ge0,V_2^M=q')$$

which implies more generally

$$|\mathbb{P}(V_j^M=q')-\mathbb{P}(V_{j+1}^M=q')|\le\mathbb{P}(Z_j\ge1)\le(2K+1)\mathbb{P}(X_0=1)$$

for any $k=1,\dots,\Delta$. Hence

$$|\mathbb{P}(V_1^M=q')-\mathbb{P}(V_\Delta^M=q')|\le\sum_{j=1}^{\Delta-1}\mathbb{P}(Z_j\ge1)\le(2K+1)\Delta\mathbb{P}(X_0=1)$$

and thus

$$\sum_{u=1}^{q-1}\mathcal{R}_3(u)\le(2K+1)\Delta\mathbb{P}(X_0=1)\sum_{u=1}^{q-1}\mathbb{P}(Z_0=u)\le(2K+1)\Delta\mathbb{P}(X_0=1)^2$$

since $\{Z_0\ge1\}\subset\bigcup_{j=0}^{2K}\{X_j=1\}$. We now can estimate one of the gap terms:

$$\mathcal{R}_3=\sum_{u=1}^{q-1}\mathcal{R}_3(u)\le2(2K+1)\Delta\mathbb{P}(X_0=1)^2$$

for all q and M.

Finally, from the previous estimates we obtain for $k\le N$,

$$|\,\mathbb{P}(V_0^{N'}=k)-\mathbb{P}(\tilde V_0^{N'}=k)\,|\le\text{const.}N'(\mathcal{R}_1+\mathcal{R}_2+K\Delta\mathbb{P}(X_0=1)^2).$$

Since $N\mathbb{P}(X_0=1)=\mathcal{O}(1)$ and $N'=N/(2K+1)$ we obtain the RHS in the theorem.

It remains to show that $\mathbb{P}(\tilde V_0^{N'}=k)=\tilde\nu(\{k\})$. To see this put $p=\mathbb{P}(\tilde Z_1\ge1)$ and let Y_j be $\mathbb{N}$-valued i.i.d. random variables with distribution $\mathbb{P}(Y_j=\ell)=\frac1p\mathbb{P}(\tilde Z_j=\ell)=\lambda_\ell$ for $\ell=1,2,\dots$. Then $Q=|\{i\in[0,N']:\ \tilde Z_i\ne0\}|$ is binomially distributed with parameters (N',p) and consequently $\tilde V=\sum_{i=1}^QY_i$ is compound binomial. □

7. Proof of Theorem 1

In this section we bound the quantities in the assumption of Theorem 3 in the usual way by making a distinction between short interactions, i.e. those that are limited by a gap of length Δ, and long interactions which constitute the principal part. The near independence of long interactions is expressed bv the decay of correlations and gives rise to the error term $\mathcal{R}_1$. The short interactions are estimated by $\mathcal{R}_2$ and use the assumptions on limited distortion, the fact that 'cylinders' are pull-backs of uniformly sized balls and the positivity of the local dimension.

7.1. **Compound binomial approximation of the return times distribution.** To prove Theorem 1 we will employ the approximation theorem from Section 6 where we put $U = B_\rho(\Gamma)$. Let $X_i = \mathbb{1}_U \circ T^{i-1}$, then we put $N = \lfloor t/\mu(U) \rfloor$, where t is a positive parameter. Let K be an integer and put as before $V_a^b = \sum_{j=a}^b Z_j$, where the $Z_j = \sum_{i=j(2K+1)}^{(j+1)(2K+1)-1} X_i$ are stationary random variables. Then for any $2 \le \Delta \le N' = N/(2K+1)$ (for simplicity's sake we assume N is a multiple of $2K+1$)

$$\left|\mathbb{P}(V_0^{N'} = k) - \tilde{\nu}(\{k\})\right| \le C_3(N'(\mathcal{R}_1 + \mathcal{R}_2) + \Delta\mu(U)), \tag{3}$$

where

$$\mathcal{R}_1 = \sup_{\substack{0<\Delta<M\le N' \\ 0<q<N'-\Delta-1/2}} \left|\sum_{u=1}^{q-1}\left(\mathbb{P}\big(Z_0 = u \wedge V_\Delta^M = q-u\big) - \mathbb{P}(Z_0 = u)\mathbb{P}\big(V_\Delta^M = q-u\big)\right)\right|$$

$$\mathcal{R}_2 = \sum_{j=1}^{\Delta} \mathbb{P}(Z_1 \ge 1 \wedge Z_j \ge 1),$$

and $\tilde{\nu}$ is the compound binomial distribution with parameters $p = \mathbb{P}(Z_j \ge 1)$ and distribution $\frac{t}{p}\mathbb{P}(Z_j = k)$. Notice that $\mathbb{P}(V_0^{N'} = k) = 0$ for $k > N$ and also $\tilde{\nu}(\{k\}) = \mathbb{P}(\tilde{V}_0^{N'} = k) = 0$ for $k > N$.

We now proceed to estimate the error between the distribution of S and a compound binomial based on Theorem 3.

7.2. **Estimating $\mathcal{R}_1$.** Let us fix ρ for the moment and put $U = B_\rho(\Gamma)$. Fix q and u and we want to estimate the quantity

$$\mathcal{R}_1(q,u) = \left|\mathbb{P}\big(Z_0 = u, V_\Delta^M = q-u\big) - \mathbb{P}(Z_0 = u)\mathbb{P}\big(V_\Delta^M = q-u\big)\right|$$

In order to use the decay of correlations (II) to obtain an estimate for $\mathcal{R}_1(q,u)$ we approximate $\mathbb{1}_{Z_0=u}$ by Lipschitz functions from above and below as follows. Let $r > 0$ be small ($r \ll \rho$) and put $U''(r) = B_r(U)$ for the outer approximation of U and $U'(r) = (B_r(U^c))^c$ for the inner approximation. We then consider the set $\mathcal{U} = \{Z_0 = u\}$ which is a disjoint union of sets

$$\bigcap_{j=1}^{u} T^{-v_j}U \cap \bigcap_{i\in[0,2K+1]\setminus\{v_j:j\}} T^{-i}U^c$$

where $0 \le v_1 < v_2 < \cdots < v_u \le 2K+1$ the u entry times vary over all possibilities. Similarly we get its outer approximation $\mathcal{U}''(r)$ and its inner approximation $\mathcal{U}'(r)$ by using $U''(r)$ and $U'(r)$ respectively. We now consider Lipschitz continuous functions approximating $\mathbb{1}_{\mathcal{U}}$ as follows

$$\phi_r(x) = \begin{cases} 1 & \text{on } \mathcal{U} \\ 0 & \text{outside } \mathcal{U}''(r) \end{cases} \qquad \text{and} \qquad \tilde{\phi}_r(x) = \begin{cases} 1 & \text{on } \mathcal{U}'(r) \\ 0 & \text{outside } \mathcal{U} \end{cases}$$

with both linear in between. The Lipschitz norms of both ϕ_r and $\tilde{\phi}_r$ are bounded by a^{2K+1}/r where $a = \sup_{x\in\mathcal{G}} |DT(x)|$. By design $\tilde{\phi}_r \le \mathbb{1}_{Z_0=u} \le \phi_r$.

We obtain

$$\begin{aligned}
\mathbb{P}\big(Z_0 = u, V_\Delta^M = q-u\big) &- \mathbb{P}(Z_0 = u)\mathbb{P}\big(V_\Delta^M = q-u\big) \\
&\le \int_M \phi_r \ \cdot \mathbb{1}_{V_\Delta^M = q-u}\, d\mu - \int_M \mathbb{1}_{Z_0=u}\, d\mu \int_M \mathbb{1}_{V_\Delta^M = q-u}\, d\mu \\
&= X + Y
\end{aligned}$$

where

$$\begin{aligned}
X &= \left(\int_M \phi_r\, d\mu - \int_M \mathbb{1}_{Z_0=u}\, d\mu\right) \int_M \mathbb{1}_{V_\Delta^M = q-u}\, d\mu \\
Y &= \int_M \phi_r \ (\mathbb{1}_{V_\Delta^M = q-u})\, d\mu - \int_M \phi_r\, d\mu \int_M \mathbb{1}_{V_\Delta^M = q-u}\, d\mu.
\end{aligned}$$

The two terms X and Y are estimated separately. The first term is readily estimated by:

$$X \le \mathbb{P}(V_\Delta^M = q-u) \int_M (\phi_r - \mathbb{1}_{Z_0=u})\, d\mu \le \mu(\mathcal{U}''(r) \setminus \mathcal{U}(r)).$$

In order to estimate the second term Y we use the decay of correlations. For this we have to approximate $\mathbb{1}_{V_0^{M-\Delta}=q-u}$ by a function which is constant on local stable leaves. (Note that if the map is expanding then there are no stable leaves and Y is straighforwardly estimated by $\mathcal{C}(\Delta)\|\phi_r\|_{Lip}$ as $\|\mathbb{1}_{V_\delta^M = q-u}\|_\infty = 1$.) Let us define

$$\mathcal{S}_n = \bigcup_{\substack{\gamma^s \\ T^n\gamma^s \subset U}} T^n\gamma^s, \qquad \partial\mathcal{S}_n = \bigcup_{\substack{\gamma^s, T^n\gamma^s \not\subset U \\ T^n\gamma^s \cap U \ne \varnothing}} T^n\gamma^s$$

and

$$\mathscr{S}_\Delta^M = \bigcup_{n=\Delta(2K+1)}^{M(2K+1)} \mathcal{S}_n, \qquad \partial\mathscr{S}_\Delta^M = \bigcup_{n=\Delta(2K+1)}^{M(2K+1)} \partial\mathcal{S}_n.$$

The set

$$\mathscr{S}_\Delta^M(q) = \{V_0^{M-\Delta} = q-u\} \cap \mathscr{S}_\Delta^M$$

is then a union of local stable leaves. This follows from the fact that by construction $T^n y \in U$ if and only if $T^n\gamma^s(y) \subset U$. We also have $\{V_0^{M-\Delta} = q-u\} \subset \tilde{\mathscr{S}}_\Delta^M(q)$ where the set $\tilde{\mathscr{S}}_\Delta^M(k) = \mathscr{S}_\Delta^M(k) \cup \partial\mathscr{S}_\Delta^M$ is a union of local stable leaves.

Denote by ψ_Δ^M the characteristic function of the set $\mathscr{S}_\Delta^M(k)$ and by $\tilde\psi_\Delta^M$ the characteristic function for $\tilde{\mathscr{S}}_\Delta^M(k)$. Then ψ_Δ^M and $\tilde\psi_\Delta^M$ are constant on local stable leaves and satisfy

$$\psi_\Delta^M \le \mathbb{1}_{V_0^{M-\Delta}=q-u} \le \tilde\psi_\Delta^M.$$

Since $\{y : \psi_\Delta^M(y) \neq \tilde\psi_\Delta^M(y)\} \subset \partial\mathscr{S}_\Delta^M$ we need to estimate the measure of $\partial\mathscr{S}_\Delta^M$.

By the contraction property $\mathrm{diam}(T^n\gamma^s(y)) \le \delta(n) \lesssim n^{-\kappa}$ outside the set $\mathcal{G}_n^c$ and consequently

$$\bigcup_{\substack{\gamma^s\subset\mathcal{G}_n\\ T^n\gamma^s\not\subset U\\ T^n\gamma^s\cap U\neq\varnothing}} T^n\gamma^s \subset U''(\delta(n))\setminus U'(\delta(n)).$$

Therefore

$$\begin{aligned}\mu(\partial\mathscr{S}_\Delta^M) &\le \mu\left(\bigcup_{n=\Delta(2K+1)}^{M(2K+1)} T^{-n}\left(U''(\delta(n))\setminus U'(\delta(n))\right)\right) + \sum_{n=\Delta(2K+1)}^{\infty}\mu(\mathcal{G}_n^c)\\ &\le \sum_{n=\Delta(2K+1)}^{M(2K+1)}\mu(U''(\delta(n))\setminus U'(\delta(n))) + \sum_{n=\Delta(2K+1)}^{\infty}\mu(\mathcal{G}_n^c)\end{aligned}$$

where the last term is estimated by assumption (III) as follows

$$\sum_{n=\Delta(2K+1)}^{\infty}\mu(\mathcal{G}_n^c) = \mathcal{O}(1)\sum_{n=\Delta(2k+1)}^{\infty} n^{-q} = \mathcal{O}(K^{-q}\Delta^{-q+1}) = \mathcal{O}(K^{-q}\rho^{\epsilon}\mu(U))$$

where we put $\Delta \sim \rho^{-v}$ and we also assume that q satisfies $v(q-1) > d_1$ (that is $\epsilon = v(q-1) - d_1 > 0$). Now by assumption (VI):

$$\mu(\partial\mathscr{S}_\Delta^M) = \mathcal{O}(1)\sum_{n=\Delta}^{\infty}\frac{n^{-\kappa\eta}}{\rho^\beta}\mu(U) + \rho^\epsilon\mu(U) = \mathcal{O}(\rho^{v(\kappa\eta-1)-\beta} + \rho^\epsilon)\mu(U)$$

with $\delta(n) = \mathcal{O}(n^{-\kappa})$ and $\Delta\sim\rho^{-v}$ where this time we also need that $v > \frac{\beta}{\kappa\eta-1}$ which is determined in Section 7.4 below. Both constraints imply that we must have $v > \max\{\frac{d_1}{q-1}, \frac{\beta}{\kappa\eta-1}\}$. If we split $\Delta = \Delta' + \Delta''$ then, using assumption (II), we can estimate as follows:

$$\begin{aligned} Y &= \left|\int_M \phi_r\, T^{-\Delta'}(\mathbb{1}_{V_{\Delta''}^{M-\Delta'}=q-u})\,d\mu - \int_M \phi_r\,d\mu\int_M \mathbb{1}_{V_0^{M-\Delta}=q-u}\,d\mu\right|\\ &\le \mathcal{C}(\Delta')\|\phi_r\|_{Lip}\|\mathbb{1}_{\tilde{\mathscr{S}}_{\Delta''}^{M-\Delta'}}\|_{\mathscr{L}^\infty} + 2\mu(\partial\mathscr{S}_{\Delta''}^{M-\Delta'}).\end{aligned}$$

Hence

$$\begin{aligned}&\mu(\mathcal{U}\cap T^{-\Delta}\{V_0^{M-\Delta} = q-u\}) - \mu(\mathcal{U})\,\mathbb{P}(V_0^{M-\Delta} = q-u)\\ &\qquad\le\; a^{2K+1}\frac{\mathcal{C}(\Delta/2)}{r} + \mu(\mathcal{U}(r)\setminus\mathcal{U}) + \mu(\partial\mathscr{S}_\Delta^M)\end{aligned}$$

by taking $\Delta'=\Delta''=\frac{\Delta}{2}$. A similar estimate from below can be done using $\tilde\phi_\rho$. Hence

$$\begin{aligned}\mathcal{R}_1 &\le c_2\left(a^{2K+1}\frac{\mathcal{C}(\Delta/2)}{r}+\mu(\mathcal{U}''(r)\setminus\mathcal{U}'(r))\right)+\mu(\partial\mathscr{S}_\Delta^M)\\ &\lesssim a^{2K+1}\frac{\mathcal{C}(\Delta/2)}{r}+\mu(\mathcal{U}''(r)\setminus\mathcal{U}'(r))+(\rho^{v(\kappa\eta-1)-\beta}+\rho^\epsilon)\mu(U).\end{aligned}$$

In the exponential case when $\delta(n)=\mathcal{O}(\vartheta^n)$ we choose $\Delta=s|\log\rho|$ for some $s>0$ and obtain the estimate

$$\mathcal{R}_1\le c_2\left(a^{2K+1}\frac{\mathcal{C}(\Delta/2)}{r}+\mu(\mathcal{U}''(r)\setminus\mathcal{U}'(r)))\right)+\mathcal{O}(\rho^{s|\log\vartheta|-\beta}+\rho^\epsilon)\mu(U).$$

7.3. **Estimating the terms $\mathcal{R}_2$.** We will first estimate the measure of $U\cap T^{-j}U$ for positive j. Fix j and and let γ^u be an unstable local leaf through U. Let us define

$$\mathscr{C}_j(U,\gamma^u)=\{\zeta_{\varphi,j}:\zeta_{\varphi,j}\cap U\neq\varnothing,\varphi\in\mathscr{I}_j\}$$

for the cluster of j-cylinders that covers the set U. As before the sets $\zeta_{\varphi,k}$ are φ-pre-images of embedded R-balls in $T^j\gamma^u$. Then

$$\begin{aligned}\mu_{\gamma^u}(T^{-j}U\cap U) &\le \sum_{\zeta\in\mathscr{C}_j(U,\gamma^u)}\frac{\mu_{\gamma^u}(T^{-j}U\cap\zeta)}{\mu_{\gamma^u}(\zeta)}\mu_{\gamma^u}(\zeta)\\ &\le \sum_{\zeta\in\mathscr{C}_j(U,\gamma^u)}c_3\omega(j)\frac{\mu_{T^j\gamma^u}(U\cap T^j\zeta)}{\mu_{T^j\gamma^u}(T^j\zeta)}\mu_{\gamma^u}(\zeta)\end{aligned}$$

The denominator is uniformly bounded from below because $\mu_{T^j\gamma^u}(T^j\zeta)=\mu_{T^j\gamma^u}(B_{R,\gamma^u}(y_k))$ for some y_k. Thus, by assumption (I), we have:

$$\begin{aligned}\mu_{\gamma^u}(T^{-j}U\cap U) &\le c_4\omega(j)\mu_{T^j\gamma^u}(U)\sum_{\zeta\in\mathscr{C}_j(U,\gamma^u)}\mu_{\gamma^u}(\zeta)\\ &\le c_4\omega(j)\mu_{T^j\gamma^u}(U)\,L\,\mu_{\gamma^u}\left(\bigcup_{\zeta\in\mathscr{C}_j(U,\gamma^u)}\zeta\right)\end{aligned}$$

Now, since outside the set $\mathcal{G}_n^c$ one has

$$\bigcup_{\zeta\in\mathscr{C}_j(U,\gamma^u)}\zeta\subset B_{j^{-\kappa}}(U)$$

where by assumption $\mu_{\gamma^u}(B_{j^{-\kappa}}(U))=\mathcal{O}((\delta(j)+\rho)^{u_0})=\mathcal{O}((j^{-\kappa}+\rho)^{u_0})$ in the polynomial case when $\delta(j)\sim j^{-\kappa}$. Therefore

$$\mu_{\gamma^u}\left(\bigcup_{\zeta\in\mathcal{C}_j(U,\gamma^u)}\zeta\right)\lesssim\delta(j)^{u_0}+\rho^{u_0}+\mu(\mathcal{G}_j^c)\lesssim j^{-\kappa u_0}+\rho^{u_0}+j^{-q}$$

in the case when $\mu(\mathcal{G}_j^c)\le\delta'(j)\sim j^{-q}$ is decaying polynomially. Consequently

$$\mu_{\gamma^u}(T^{-j}U\cap U)\le c_5\omega(j)\mu_{T^j\gamma^u}(U)(j^{-\kappa u_0}+\rho^{u_0}+j^{-q}).$$

Since $d\mu = d\mu_{\gamma^u} dv(\gamma^u)$ we obtain

$$\mu(T^{-j}U \cap U) \leq c_6\omega(j)\mu(U)(j^{-\kappa u_0} + \rho^{u_0} + j^{-q}).$$

Next we estimate for $j \geq 2$ the quantity

$$\begin{aligned}\mathbb{P}(Z_0 \geq 1, Z_j \geq 1) &\leq \sum_{0\leq k,\ell<2K+1} \mu(T^{-k}U \cap T^{-\ell-(2K+1)j}U)\\ &= \sum_{u=(j-1)(2K+1)}^{(j+1)(2K+1)} ((2K+1) - |u - j(2K+1)|)\mu(U \cap T^{-u}U)\end{aligned}$$

and consequently obtain

$$\begin{aligned}\sum_{j=2}^{\Delta} \mathbb{P}(Z_0 \geq 1 \wedge Z_j \geq 1) &\leq (2K+1) \sum_{u=2K+1}^{(\Delta+1)(2K+1)} \mu(U \cap T^{-u}U)\\ &\leq c_7 K\mu(U) \sum_{u=2K+1}^{(\Delta+1)(2K+1)} \omega(u)(u^{-\kappa u_0} + \rho^{u_0} + u^{-q})\\ &\leq c_8 K\mu(U)\Big(K^{-\sigma} + K\Delta^{1+\kappa'}\rho^{u_0}\Big)\end{aligned}$$

since $\omega(j) = \mathcal{O}(j^{-\kappa'})$, provided $\sigma = \min\{\kappa u_0, q\} - \kappa' - 1$ is larger than 0. For the term $j = 1$ let $K' < K$ and put $Z_0' = \sum_{i=2K+1-K'}^{2K+1} X_i$ and $Z_0'' = Z_0 - Z_0'$. Then

$$\mathbb{P}(Z_0 \geq 1, Z_1 \geq 1) \leq \mathbb{P}(Z_0'' \geq 1, Z_1 \geq 1) + \mathbb{P}(Z_0' \geq 1),$$

where $\mathbb{P}(Z_0' \geq 1) \leq K'\mu(U)$. Since by the above estimates

$$\mathbb{P}(Z_0'' \geq 1, Z_1 \geq 1) \lesssim KK'^{-\sigma}\mu(U) + K^2\Delta^{1+\kappa'}\rho^{u_0}\mu(U)$$

we conclude that

$$\mathbb{P}(Z_0 \geq 1, Z_1 \geq 1) \lesssim \mu(U)(K' + KK'^{-\sigma} + K^2\Delta^{1+\kappa'}\rho^{u_0}).$$

The entire error term is now estimated by

$$\begin{aligned}N'\mathcal{R}_2 &\leq N' \sum_{j=1}^{\Delta} \mathbb{P}(Z_0 \geq 1, Z_j \geq 1)\\ &\lesssim N'\mu(U)(K^{1-\sigma} + KK'^{-\sigma} + K' + K^2\Delta^{1+\kappa'}\rho^{u_0})\\ &\lesssim t\bigg(K'^{-\sigma} + \frac{K'}{K} + K^2\Delta^{1+\kappa'}\rho^{u_0}\bigg)\\ &\lesssim t\frac{K'}{K}\end{aligned}$$

if $v(1+\kappa') + u_0 > 0$ (as $\Delta = \rho^{-v}$) as $K > K' = K^\alpha$ where we put $\alpha = \frac{1}{1+\sigma}$.

If diam ζ (ζ n-cylinders) and $\mu(\mathcal{G}_n^c)$ decay super polynomially then

$$N'\mathcal{R}_2 \lesssim \delta(K')^{u_0} + \delta'(K') + K'/K + K^2\Delta^{1+\kappa'}\rho^{u_0} \lesssim tK^{\alpha-1},$$

where diam $\zeta \leq \delta(n)$, $\mu(\mathcal{G}_n^c) \leq \delta'(n)$ are super polynomial.

In the exponential case ($\delta(n),\delta'(n)=\mathcal{O}(\vartheta^n)$) one has

$$N'\mathcal{R}_2\lesssim \vartheta^{(u_0\wedge 1)K'}+K'/K+K^2\Delta^{1+\kappa'}\rho^{u_0}.$$

7.4. **The total error.** For the total error we now put $r=\rho^w$ and as above $\Delta=\rho^{-v}$ where $v<d_0$ since $\Delta\lll N$ and $N\ge\rho^{-d_0}$. Moreover $K'=K^\alpha$ for $\alpha=1/(1+\sigma)$ and $\mathcal{C}(\Delta)=\mathcal{O}(\Delta^{-p})=\mathcal{O}(\rho^{pv})$ and we get (in the polynomial case)

$$\begin{aligned}
&|\mathbb{P}(W=k)-\tilde\nu(\{k\})|\\
&\quad\lesssim\ N'\left(a^{2K+1}\frac{\mathcal{C}(\Delta)}{r}+\mu(\mathcal{U}''(r)\setminus\mathcal{U}'(r))\right)+\frac{t}{K^{\sigma'}}+\frac{t}{K}(\rho^{v(\kappa\eta-1)-\beta}+\rho^{\epsilon})+\frac{K'}{K}\\
&\quad\lesssim\ a^{2K+1}\rho^{vp-w-d_1}+\rho^{w\eta-\beta}+K^{\alpha-1}
\end{aligned}$$

as $N'\mu(U)=\frac{s}{2K+1}$ and $s=N'\mathbb{P}(Z^K\ge1)$. Put $u_0'=u_0/(1+\kappa')$ and we can now choose $v<d_0\wedge u_0'$ arbitrarily close to $d_0\wedge u_0$ and then require $vp-w-d_1>0$, $w\eta-\beta>0$ and $v(\kappa\eta-1)-\beta>0$. We can choose $w>\frac{\beta}{\eta}$ arbitrarily close to $\frac{\beta}{\eta}$ and can satisfy all requirements if $p>\frac{\frac{\beta}{\eta}+d_1}{d_0\wedge u_0'}$ in the case when $\mathcal{C}$ decays polynomially with power p, i.e. $\mathcal{C}(k)\sim k^{-p}$.

In the exponential case (diam $\zeta=\mathcal{O}(\vartheta^n)$ for n cylinders ζ and $\mathcal{C}(\Delta)\sim\vartheta^{\Delta}$) we obtain with $\Delta=s|\log\rho|$ for s large enough

$$|\mathbb{P}(W=k)-\tilde\nu(\{k\})|\lesssim a^{2K+1}\rho^{s|\log\vartheta|-w-d_1}+\rho^{w\eta-\beta}+K^{\alpha-1},$$

where $\varepsilon\in(0,u_0)$.

7.5. **Convergence to the compound Poisson distribution.** First observe that for $t>0$ we take $N=t/\mu(U)$ and since by Lemma 3 $N'\alpha_1\mu(U)=s$ this implies that $s=\alpha_1 t$. We will have to do a double limit of first letting ρ go to zero and then to let K go to infinity. If $\rho\to0$ then $\mu(U)\to0$ which implies that $N'\to\infty$ and that the compound binomial distribution $\tilde\nu$ converges to the compound Poisson distribution $\tilde\nu_K$ for the parameters $t\lambda_\ell(K)$. Thus for every K:

$$\mathbb{P}(W=k)\longrightarrow\tilde\nu_K(\{k\})+\mathcal{O}(tK^{-\sigma'}).$$

Now let $K\to\infty$. Then $\lambda_\ell(K)\to\lambda_\ell$ for all $\ell=1,2,\dots$ and $\tilde\nu_K$ converges to the compound Poisson distribution ν for the parameters $s\lambda_\ell=\alpha_1 t\lambda_\ell$. Finally we obtain

$$\mathbb{P}(W=k)\longrightarrow\nu(\{k\})$$

as $\rho\to0$. This concludes the proof of Theorem 1. □

8. Examples

8.1. **A non-uniformly expanding map.** On the torus $\mathbb{T}=[0,1)\times[0,1)$ we consider the affine map T given by the matrix $A=\begin{pmatrix}1&1\\0&a\end{pmatrix}$ for some integer $a\ge2$. This is a partially hyperbolic map since A has one eigenvalue equal to 1 and is uniformly expanding in the y-direction. Horizontal lines are mapped to horizontal lines and in particular the line $\Gamma=\{(x,0):x\in[0,1)\}$ is an invariant set which entirely consists of fixed points. Since in estimating the error terms $\mathcal{R}_2$ involves terms of the form $U\cap T^{-n}U$ we only need to consider the uniformly expanding y-direction when verifying the assumption (III)(iii). This means the vertical diameter of n-cylinders ζ contracts exponentially like a^{-n}.

The Lebesgue measure μ is invariant. To see this notice that T has a inverse branches whose Jacobians all have determinant $\frac{1}{a}$. The neighbourhoods U of Γ are $B_\rho(\Gamma)$. In Assumptions (IV) and (V) we thus have $d_0 = d_1 = u_0 = 1$ and in the "annulus condition" (VI) one can take $\eta = \beta = 1$.

Although this map does not have good decay of correlation we can still apply our method because the return sets $B_\rho(\Gamma)$ are of very special form since A maps horizontal lines $y \times [0,1)$ to horizontal lines $y' \times [0,1)$ ($y' = ay \mod 1$) and in vertical direction is uniformly expanding by factor a.

The limiting return times are in the limit compound Poisson distributed. It is straightforward to determine that

$$\hat{\alpha}_{k+1} = \lim_{\rho\to 0} \mu_{B_\rho(\Gamma)}\left(T^{-1}B_\rho(\Gamma) \cap T^{-2}B_\rho(\Gamma) \cap \cdots \cap T^{-k}B_\rho(\Gamma)\right) = \left(\frac{1}{a}\right)^k,$$

$k = 1, 2, \ldots,$ since $\mu\left(\bigcap_{j=0}^k T^{-j}B_\rho(\Gamma)\right) = a^{-k}\rho$. Consequently $\alpha_k = \hat{\alpha}_k - \hat{\alpha}_{k+1} = \left(1 - \frac{1}{a}\right)\left(\frac{1}{a}\right)^{k-1}$ and by Theorem 2 $\lambda_k = (1 - \frac{1}{a})(\frac{1}{a})^{k-1}$, $k = 1, 2, \ldots,$ which shows that the return times to a strip neighbourhood of Γ is in the limit Pólya-Aeppli distributed. (The extremal index is $\alpha_1 = 1 - \hat{\alpha}_2 = 1 - \frac{1}{a}$.)

8.2. **Regenerative processes.** Here we give two examples, one which exhibits some pathology and which was also recently used in [3] and another one to show that nearly all compound Poisson distributions can be achieved.

8.2.1. *Smith example.* To emphasise the regularity condition made in Theorem 2 we look at an example by Smith [27] which was also recently used in [2, 3] to exhibit some pathology.

Let Y_j for $j \in \mathbb{Z}$ be i.i.d. $\mathbb{N}$-valued random variables and denote $\gamma_k = \mathbb{P}(Y_j = k)$ its probability density. For each $k \in \mathbb{N}$, put $p_k = 1 - \frac{1}{k}$ and $q_k = \frac{1}{k}$. Then we define the regenerative process $X_j, j \in \mathbb{Z}$, as follows: the sequence of $\vec{X} = (\ldots, X_{-1}, X_0, X_1, X_2, \ldots)$ is parsed into blocks of lengths $\zeta_i \in \mathbb{N}$ so that the sequence of integers N_i satisfy $N_{i+1} = N_i + \zeta_i$. Then $X_{N_i} = k$ with probability γ_k and $\mathbb{P}(\zeta_i = 1) = q_k$ and $\mathbb{P}(\zeta_i = k) = q_k$. If $\zeta_i = k$ then we put $X_{N_i+\ell} = k$ for $\ell = 1, 2, \ldots, k$. That means every time the symbol k is chosen (with probability γ_k) then appears a block of only that one symbol with probability p_k or as a block of length $k+1$ of k times repeated symbol k with probability q_k.

The sets $U_m = \{\vec{X} : X_0 > m\}$ form a nested sequence within the space $\Omega = \{\vec{X}\}$ which carries the left shift transform $\sigma : \Omega \circlearrowleft$. Moreover there exists a σ-invariant probability measure μ for which $\mu(\{k\}) = \gamma_k$. To find $\hat{\alpha}_k(L)$ for (large) L we let $m > L$. For $\vec{X} \in \Omega$, let i be so that $N_i(\vec{X}) \le 0 < N_{i+1}(\vec{X})$. Then $\zeta_i = N_{i+1} - N_i$ is the length of the block containing X_0. Let $\varepsilon > 0$, then for all k large enough we have

$$\mathbb{P}(\zeta_i = 1 | X_0 = k) = \frac{p_k}{p_k + (k+1)q_k} \in \left(\frac{1}{2} - \varepsilon, \frac{1}{2}\right)$$

as $\mathbb{E}(\zeta_i|X_0) = 2$ for all k. Similarly

$$\mathbb{P}(\zeta_i = k+1|X_0 = k) = \frac{(k+1)q_k}{p_k + (k+1)q_k} \in \left(\frac{1}{2}, \frac{1}{2} + \varepsilon\right)$$

for all k large enough. In particular, for all m large enough,

$$\mathbb{P}(\zeta_i = 1|U_m) \in \left(\frac{1}{2} - \varepsilon, \frac{1}{2}\right), \qquad \mathbb{P}(\zeta_i > 1|U_m) \in \left(\frac{1}{2}, \frac{1}{2} + \varepsilon\right).$$

Therefore ($\zeta_1 > 1$ here means $\zeta_i > m$)

$$\begin{aligned}\left|\mathbb{P}_{U_m}\left(\tau_{U_m}^{k-1} > L\right) - \frac{1}{2}\right| &\leq \mathbb{P}(\zeta_i = 1|U_m)\mu(U_m) + \left|\mathbb{P}(\zeta_i = 1|U_m) - \frac{1}{2}\right| \\ &\quad + \mathbb{P}(\zeta_i > 1|U_m)\frac{L}{m} + \left|\mathbb{P}(\zeta_i > 1|U_m) - \frac{1}{2}\right| \\ &\leq 4\varepsilon\end{aligned}$$

for all m large enough so that in particular also $L/m < \varepsilon$ and $\mu(U_m) < \varepsilon$. The first term on the RHS comes from the events that re-enter U_m after exiting and the third term accounts for the probability that the block of length ζ_i does not cover the entire interval $(0, L]$. Consequently

$$\hat{\alpha}_k(L) = \lim_{m\to\infty} \mathbb{P}_{U_m}\left(\tau_{U_m}^{k-1} > L\right) = \frac{1}{2}$$

for all $k \geq 2$ and for all L (trivially $\hat{\alpha}_1 = 1$). Consequently $\hat{\alpha}_k = \frac{1}{2}$ for all $k = 1, 2, \dots$. Moreover we also obtain that $\alpha_1 = \frac{1}{2}$ and $\alpha_k = 0$ for all $k \geq 2$.

Since the condition of Theorem 2 is not satisfied, we cannot use it to obtain the probabilities λ_k for the k-clusters. We can however proceed more directly by noting that

$$\mathbb{P}(Z^L > 0) = (2L+1)\mu(U_m)(1 - \mathcal{O}^*(1/m)) - \mathcal{O}(\mu(U_m)^2 g(L, \mu(U_m))),$$

where $Z^L = \sum_{j=-L}^{L} \mathbb{1}_{U_m} \circ \sigma^j$ and $g(L, \mu(U_m))$ is a function which is bounded and stays bounded as $\mu(U_m) \to 0$ ($\mathcal{O}^*$ expresses that the implied constant is 1, i.e. $x = \mathcal{O}^*(\epsilon)$ if $|x| < \epsilon$) Similarly we get that

$$\mathbb{P}(Z^L = 1) = (2L+1)\mu(U_m) + \mathcal{O}(\mu(U_m)^2 g'(L, \mu(U_m)))$$

where g' is like g. Also

$$\mathbb{P}(Z^L > 1) = \mathcal{O}(\mu(U_m))$$

where the implied constant depends on L. Consequently

$$\lambda_k(L) = \lim_{n\to\infty} \frac{\mathbb{P}(Z^L = k)}{\mathbb{P}(Z^L > 0)} = \mathcal{O}(1/L) \to 0$$

as $L \to \infty$ and therefore $\lambda_k = 0$ for all $k \geq 2$. For $k = 1$ we obtain

$$\lambda_1(L) = \lim_{n\to\infty} \frac{\mathbb{P}(Z^L = 1)}{\mathbb{P}(Z^L > 0)} = 1.$$

This does not square with the statement of Theorem 2 since the we have masses that are wandering off to infinity.

8.2.2. *Arbitrary parameters.* We use an example which is similar to Smith's to show that any sequence of parameters λ_k can be realised as long as the expected value is finite. As above let Y be an $\mathbb{N}$-valued random variable with probability distribution $\gamma_k = \mathbb{P}(Y = k)$. Let $\lambda_k, k = 1, 2, \dots$, be a sequence of parameter values so that $\sum_{k=1}^\infty \lambda_k = 1$ and $\sum_{k=1}^\infty k\lambda_k < \infty$. As above we define the regenerative process $X_j, j \in \mathbb{Z}$ by parsing the sequence of $\vec{X} = (\dots, X_{-1}, X_0, X_1, \dots)$ into blocks of lengths $\zeta_i \in \mathbb{N}$ so that the sequence of integers N_i which indicates the heads of runs satisfy $N_{i+1} = N_i + \zeta_i$. Then $X_{N_i} = k$ with probability γ_k and $\mathbb{P}(\zeta_i = j) = \lambda_j$. That means that blocks of the symbol k which are of length j are chosen with the given probability λ_j. Put $\Omega = \{\vec{X}\}$.

As before, let $U_m = \{\vec{X} \in \Omega : X_0 > m\}$. For $\vec{X} \in \Omega$ let i be so that $N_i \le 0 < N_{i+1}$. Then X_0 belongs to a block of length $\zeta_i = N_{i+1} - N_i$. This implies

$$\mathbb{P}(\zeta_i = \ell) = \frac{\ell\lambda_\ell}{\sum_{s=1}^\infty s\lambda_s}$$

Also $\mathbb{P}(X_0 = X_1 = \cdots X_{k-1} \ne X_k | \zeta_i = \ell) = 1/\ell$ and consequently for $k < m$:

$$\alpha_k(L, U_m) = \frac{\sum_{\ell=k}^\infty \lambda_\ell}{\sum_{s=1}^\infty s\lambda_s} + \mathcal{O}(L\mu(U_m))$$

where the error terms expresses the likelyhood for entering the set U_m after the ζ_i-block of being inside U_m. Taking a limit $m \to \infty$ we obtain

$$\alpha_k = \lim_{L\to\infty} \alpha_k(L) = \frac{\sum_{\ell=k}^\infty \lambda_\ell}{\sum_{s=1}^\infty s\lambda_s}.$$

In particular if $k = 1$ we get $\alpha_1 = 1/\sum_{s=1}^\infty s\lambda_s = 1/\mathbb{E}(\zeta_i)$ as $\sum_{\ell=1}^\infty \lambda_\ell = 1$. This is the relation to be expected in general, where the extremal index α_1 is the reciprocal of the expected value of the cluster length.

8.3. **Periodic points.** For a set $U \subset \Omega$ we write $\tau(U) = \inf_{y\in U} \tau_U(y)$ for the *period* of U. In other words, $U \cap T^{-j}U = \varnothing$ for $j = 1, \dots, \tau(U) - 1$ and $U \cap T^{-\tau(U)}U \ne \varnothing$. Let us now consider a sequence of nested sets $U_n \subset \Omega$ so that $U_{n+1} \subset U_n \forall n$ and $\bigcap_n U_n = \{x\}$ a single point x. Then we have the following simple result which is independent of the topology or an invariant measure on Ω.

Lemma 4. *Let $U_n \subset \Omega$ be so that $U_{n+1} \subset U_n \forall n$ and $\bigcap_n U_n = \{x\}$ for some $x \in \Omega$. Then the sequence $\tau(U_n)$, $n = 1, 2, \dots$ is bounded if and only if x is a periodic point.*

Proof. If we put $\tau_n = \tau(U_n)$ then $\tau_{n+1} \ge \tau_n$ for all n. Thus either $\tau_n \to \infty$ or τ_n has a finite limit τ_∞. Assume $\tau_n \to \tau_\infty < \infty$. Then $\tau_n = \tau_\infty$ for all $n \ge N$, for some N, and thus $U_n \cap T^{-\tau_\infty}U_n \ne \varnothing$ for all $n \ge N$. Since the intersections $U_n \cap T^{-\tau_\infty}U_n$ are nested, i.e. $U_{n+1} \cap T^{-\tau_\infty}U_{n+1} \subset U_n \cap T^{-\tau_\infty}U_n$ we get

$$\varnothing \ne \bigcap_{n\ge N} (U_n \cap T^{-\tau_\infty}U_n) = \bigcap_{n\ge N} U_n \cap \bigcap_{n\ge N} T^{-\tau_\infty}U_n = \{x\} \cap \{T^{-\tau_\infty}x\}$$

which implies that $x = T^{\tau_\infty}x$ is a periodic point. Conversely, if x is periodic then clearly the τ_n are bounded by its period. □

Let us now compute the values λ_ℓ. Assume x is a periodic point with minimal period m, then $p_i^\ell = \mu_{U_n}(\tau_{U_n}^{\ell-1} = i) = 0$ for $i < m$ and $m = \tau(U_n)$ if n is large enough. For n large enough one has $\tau(U_n) = \tau_\infty = m$ and therefore $U_n \cap \{\tau_{U_n} = m\} = U_n \cap T^{-m}U_n$.

Assume the limit $p = p_m^2 = \lim_{n\to\infty} \frac{\mu(U_n\cap T^{-m}U_n)}{\mu(U_n)}$ exists, then one also has more generally

$$p_{(\ell-1)m}^\ell = \lim_{n\to\infty} \frac{\mu(\bigcap_{j=1}^{\ell-1} T^{-jm}U_n)}{\mu(U_n)} = p^{\ell-1}.$$

All other values of p_i^ℓ are zero, that is $p_i^\ell = 0$ if $i \neq (\ell-1)m$. Thus $\hat{\alpha}_\ell = p_{(\ell-1)m}^\ell = p^{\ell-1}$ and consequently

$$\alpha_\ell = \hat{\alpha}_\ell - \hat{\alpha}_{\ell+1} = (1-p)p^{\ell-1}$$

which is a geometric distribution. This implies that the random variable W is in the limit Pólya-Aeppli distributed with the parameters $\lambda_k = (1-p)p^{k-1}$.

In particular the extremal index here is $\alpha_1 = 1 - \hat{\alpha}_2 = 1 - p$.

Remark 8. *The extremal index can be explicitly evaluated in some cases. Two examples are: For one-dimensional maps T of Rychlik type with potential ϕ (with zero pressure) and equilibrium state μ_ϕ, if x is a periodic point of prime period m, then we get Pitskel's value $\alpha_1 = 1 - e^{\sum_{k=0}^{m-1}\phi(T^kx)}$, see* [22, 15, 9]. *In the notation adopted above that is $p = e^{\sum_{k=0}^{m-1}\phi(T^kx)}$ and $\lambda_k = (1-p)p^{k-1}$. For piecewise multidimensional expanding maps T considered in* [26], *if ξ is again a periodic point of prime period m, then $\alpha_1 = 1 - |\det D(T^{-m})(\xi)|$, i.e. $p = |\det D(T^{-m})(\xi)|$, see Corollary 4 in* [9] *and also* [4].

If x is a non-periodic point, then $\tau_n = \tau(U_n) \to \infty$ as $n \to \infty$ which implies that $\mathbb{P}(\tau_{U_n} \leq L) = 0$ for all n large enough (i.e. when $\tau_n > L$), and therefore for all $k \geq 2$ $\hat{\alpha}_k(L) \leq \lim_n \mathbb{P}_{U_n}(\tau_{U_n} \leq L) = 0$. That is $\alpha_1 = \lambda_1 = 1$ and $\alpha_k = \lambda_k = 0 \forall k \geq 2$. The limiting return times distribution is therefore a regular Poisson distribution.

9. Coupled map lattice

Let T be a piecewise continuous map on the unit interval $[0,1]$. We want to consider a map $\hat{T}$ on $\Omega = [0,1]^n$ for some integer n which is given by

$$\hat{T}(\vec{x})_i = (1-\gamma)T(x_i) + \gamma \sum_{j=1}^n M_{i,j}T(x_j) \qquad \forall\, i = 1, 2, \dots, n, \tag{4}$$

for $\vec{x} \in \Omega$, where M is an $n \times n$ stochastic matrix and $\gamma \in [0,1]$ is a coupling constant. For $\gamma = 0$ we just get the product of n copies of T. We assume that T is a piece-wise expanding map of the unit interval onto itself, with a finite number of branches, say q, and which T is assumed to be of class C^2 on the interiors of the domains of injectivity $I_1, \dots, I_q$, and extended by continuity to the boundaries. Whenever the coupling constant $\gamma = 0$ the map $\hat{T}$ is the direct product of T with itself; therefore $\hat{T}$ could be seen as a coupled map lattice (CML). Let us denote by $U_k, k = 1, \dots, q^n$, the domains of local injectivity of $\hat{T}$. By the previous assumptions on T, there exist open sets $W_k \supset U_k$ such that $\hat{T}|_{W_k}$ is a C^2 diffeomorphism (on the image). We will require that

$$s := \sup_k \sup_{\vec{x}\in\hat{T}(W_k)} ||D\hat{T}|_{W_k}^{-1}(\vec{x})|| < b < 1,$$

where $b := \sup_i \sup_{x\in T(A_i)} |DT|_{I_i}^{-1}(x)|$, and $||\cdot||$ stands for the euclidean norm. We will write *dist* for the distance with respect to this norm. We will suppose that the map $\hat{T}$ preserve an absolutely continuous invariant measure μ which is moreover mixing. Recall that $\mathrm{osc}(h,A) := \mathrm{Esup}_{\vec{x}\in A} h(\overline{x}) - \mathrm{Einf}_{\overline{x}\in A} h(\vec{x})$ for any measurable set A: see the proof of Lemma 5 for a more detailed definition. Finally Leb is the Lebsegue measure on Ω.

Let

$$S_\nu := \{\vec{x} \in I^n : |x_i - x_j| \le \nu \forall i,j\} \tag{5}$$

be the ν-neighbourhood of the diagonal Δ. Then $\hat{\alpha}_{k+1}(L, S_\nu) = \mathbb{P}(\tau^k_{S_\nu} \le L | S_\nu)$. The value $\hat{\alpha}_{k+1}$ is the limiting probability of staying in the neighborhood of the diagonal until time k and as the strip S_ν collapses to the diagonal Δ.

Theorem 4. *Let $\hat{T} : \Omega \to \Omega$ be a coupled map lattice over the uniformly expanding map $T : [0,1] \circlearrowleft$ and assume that the hypersurfaces of discontinuities are piecewise $C^{1+\alpha}$ and intersections with the diagonal Δ are transversal. Moreover suppose the stochastic matrix M has constant columns, that is $M_{i,j} = p_j$ for a probability vector $\vec{p} = (p_1, \dots, p_n)$ and assume the map $\hat{T}$ satisfies Assumption (0) for any $\gamma \in [0,1]$.*

Finally suppose that the density h of the invariant absolutely continuous probability measures μ satisfies

$$\sup_{0<\varepsilon\le\varepsilon_0} \frac{1}{\varepsilon} \int_0^1 osc(h, B_\varepsilon((x)^n)\, dx < \infty,$$

where $(x)^n \in \Delta$ is the point on the diagonal all of whose coordinates are equal to $x \in [0,1]$.

Then

$$\hat{\alpha}_{k+1} = \lim_{L\to\infty} \lim_{\nu\to 0} \hat{\alpha}_{k+1}(L, S_\nu) = \frac{1}{(1-\gamma)^{k(n-1)} \int_I h((x)^n)\, dx} \int_I \frac{h((x)^n)}{|DT^k(x)|^{n-1}}\, dx$$

and the limiting return times to the diagonal Δ are compound Poisson distributed with parameters $t\lambda_k$ where $\lambda_k = \frac{1}{1-\hat{\alpha}_2}(\hat{\alpha}_{k-1} - 2\hat{\alpha}_k + \hat{\alpha}_{k+1})$ and $t > 0$ is real.

Remark 9. *If $|DT|$ is constant, as for instance for the doubling map, then we obtain $\hat{\alpha}_{k+1} = ((1-\gamma)|DT|)^{-k(n-1)}$. This implies that the probabilities $\alpha_k = \hat{\alpha}_k - \hat{\alpha}_{k+1}$ are geometric and, by extension, also*

$$\lambda_k = ((1-\gamma)|DT|)^{-(k-1)(n-1)} \left(1 - ((1-\gamma)|DT|)^{-(n-1)}\right).$$

This means that the cluster sizes are geometrically distributed and therefore the limiting return times distribution is Pólya-Aeppli. If $|DT|$ is non-constant then in general we cannot expect the probabilities α_k and λ_k to be geometric, which implies that in the generic case, the limiting return times distribution is not Pólya-Aeppli. This should clarify a remark made in [8]*, Section 6.*

Remark 10. *Theorem 4 is a generalization of Propositions 5.5 and 5.6 in* [8] *where the latter were shown for $k = 1$. But there are two more substantial differences:*
(i) In [8] *the proof was based on the transfer operator which we avoid here. Naturally, the present argument can be be extended to situations where the use of the transfer operator would not be practical.*
(ii) In [8] *we introduced the conditions (P01, P02) in order to compute the limit in*

Lemma 6 (see below). The present proof avoids those assumptions and replaces them we the rather natural requirements that the hypersurfaces of discontinuities are piecewise $C^{1+\alpha}$ and intersections with the diagonal Δ are transversal.

Remark 11. *We now give an example of a map verifying Assumption (I). Suppose the map T is defined on the unit circle as $T(x) = ax \mod 1$, with $a \in \mathbb{Z}$. Then, by using the quantities M and B introduced in the proof of Theorem 4, it is easy to see that $\hat{T}^k(\vec{x}) = B^k(a^k x_1 \mod 1, \cdots, a^k x_n \mod 1)^T$, and therefore the images of the k-cylinders will be the whole space.*

For $k = 1$ a proof already appeared in [8]; the proof which we give here is considerably simpler and easily adaptable to other coupled map lattices. In particular, instead of using the transfer operator to determine the measure of S_ν^k (below) we use the tangent map of the coupled map in the neighbourhood of the diagonal.

Let us put

$$S_\nu^k = \bigcap_{\ell=0}^{k} \hat{T}^{-\ell} S_\nu = \left\{ \vec{x} \in \Omega : |(\hat{T}^\ell(\vec{x}))_i - (\hat{T}^\ell(\vec{x}))_j| < \nu, \ell = 0, \dots, k \right\}$$

as the set of points in S_ν which for $k-1$ iterates of $\hat{T}$ stay in the S_ν-neighbourhood of the diagonal Δ. We proceed in two steps.

Lemma 5. *Under the assumption of Theorem 4 we get*

$$\hat{\beta}_{k+1} := \lim_{\nu \to 0} \frac{\mu(S_\nu^k)}{\mu(S_\nu)} = \frac{1}{(1-\gamma)^{k(n-1)} \int_I h((x)^n)\, dx} \int_I \frac{h((x)^n)}{|DT^k(x)|^{n-1}}\, dx.$$

Proof. The density h of μ is the (unique) eigenfunction of the transfer operator acting on the space of quasi-Hölder functions, see [18] and especially [26]. For all functions h on Ω we define a semi-norm $|h|_\alpha$ which, given two real numbers $\varepsilon_0 > 0$ and $0 < \alpha \le 1$, writes

$$|h|_\alpha := \sup_{0<\varepsilon\le\varepsilon_0} \frac{1}{\varepsilon^\alpha} \int \operatorname{osc}(h, B_\varepsilon(\vec{x}))\, d\mathrm{Leb}(\vec{x}).$$

We say that $h \in V_\alpha(\Omega)$ if $|h|_\alpha < \infty$. Although the value of $|h|_\alpha$ depends on ε_0, the space $V_\alpha(\Omega)$ does not. Moreover the value of ε_0 can be chosen in order to satisfy a few geometric constraints, like distortion, and to guarantee the Lasota-Yorke inequality on the Banach space $\mathcal{B} = (V_\alpha(\Omega), ||\cdot||_\alpha)$, where the norm $||\cdot||_\alpha$ is defined as $||h||_\alpha := |h|_\alpha + ||h||_1$. It has been shown [26] that $\mathcal{B}$ can continuously be injected into $\mathscr{L}^\infty$ since $||h||_\infty \le C_H ||h||_\alpha$, where $C_H = \frac{\max(1,\varepsilon_0^\alpha)}{Y_n \varepsilon_0^n}$, being Y_n the volume of the unit ball in $\mathbb{R}^n$. The density in the neighborhood of the diagonal Δ is controlled by the assumption

$$h_D := \sup_{0<\varepsilon\le\varepsilon_0} \frac{1}{\varepsilon} \int_0^1 \operatorname{osc}(h, B_\varepsilon((x)^n)\, dx < \infty,$$

where $(x)^n \in \Delta$. This means that we compute the oscillation in balls moving along the diagonal. By decreasing the radius ε the oscillation decreases; this plus Fatou Lemma implies that

$$\lim_{\varepsilon \to 0} \operatorname{osc}(h, B_\varepsilon((x)^n) = 0,$$

for Lebesgue almost all $x \in [0,1]$, which in turns implies that h *is almost everywhere continuous along the diagonal.* Consequently, if x_1 is chosen almost everywhere in $[0,1]$ and the vector $(y_2,\dots,y_n)$ is chosen almost everywhere (with respect to the Lebesgue measure on $\mathbb{R}^{n-1}$) in a ball of radius $\nu < \varepsilon_0$ around the point $(x_1)^n$, we have $|h(x_1,y_2,\dots,y_n) - h((x_1)^n)| \le \operatorname{osc}(h, B_\nu((x_1)^n)))$ and therefore

$$\int |h(x_1,y_2,\dots,y_n) - h((x_1)^n)|\,dx_1 \le \int \operatorname{osc}(h, B_\varepsilon((x_1)^n))\,dx_1 \le \nu\, h_D,$$

which goes to 0 when ν tends to zero.

For the neighbourhood $S_\nu := \{\vec{x}\in I^n : |x_i - x_j| \le \nu \forall i,j\}$ of the diagonal Δ, we now want to compute the limit $\hat{\beta}_{k+1} = \lim_{\nu\to 0}\frac{\mu(S_\nu^k)}{\mu(S_\nu)}$, where as before $S_\nu^k = \bigcap_{j=0}^k S_\nu$, which measures the limiting probability of staying in the neighborhood of the diagonal until time k and as the strip S_ν collapses to the diagonal Δ. We begin to observe that the derivative $D\hat{T}$ has the form

$$D\hat{T} = ((1-\gamma)\mathrm{id} + \gamma M)\, D\mathbb{T},$$

or $D\hat{T} = B \cdot D\mathbb{T}$, where $D\mathbb{T}(T)$ is the diagonal $n\times n$ matrix with diagonal entries $DT(x_1), DT(x_2),\dots,DT(x_n)$ and $B = (1-\gamma)\mathrm{id} + \gamma M$.

Let $\vec{u} = n^{-\frac12}(1,1,\dots,1)$ be the unit vector that spans the diagonal Δ. For a point $\vec{x}\in\Omega$ put $\vec{v}$ for the vector in $\mathbb{R}^n$ with components $v_j = x_j - x_0$ where $x_0 \in [0,1]$ is arbitrary. Then

$$d(\vec{x},\Delta) = \left(|\vec{v}|^2 - (\vec{v}\cdot\vec{u})^2\right)^{\frac12}$$

is distance of $\vec{x}$ from the diagonal.

For $x_0\in[0,1]$ denote by $(x_0)^n \in \Delta$ the point on the diagonal all of whose coordinates are equal to x_0. Notice that $\hat{T}$ leaves the diagonal invariant as $\hat{T}((x_0)^n) = (T(x_0))^n$. If $\vec{v}$ and $(x_0)^n$ lie in the same region of continuity of $\hat{T}^\ell$ then

$$d(\hat{T}^\ell((x_0)^n), \hat{T}^\ell(\vec{x})) = D\hat{T}^\ell((x_0)^n)\vec{v} + \mathcal{O}(|\vec{v}|^2),$$

where as before $\vec{v} = \vec{x} - (x_0)^n$ and $D\hat{T}^\ell((x_0)^n) = DT^\ell(x_0)B^n$. Consequently

$$d(\hat{T}^\ell(\vec{x}),\Delta) = \left(|\vec{v}(\ell)|^2 - (\vec{v}(\ell)\cdot\vec{u})^2\right)^{\frac12},$$

where ($\vec{v} = \vec{v}(0)$)

$$\vec{v}(\ell) = \hat{T}^\ell(\vec{x}) - \hat{T}^\ell((x_0)^n) = DT^\ell(x_0)B^\ell\vec{v} + \mathcal{O}(|\vec{v}|^2).$$

Using the linearisation of $\hat{T}$, the set S_ν^k is approximated by

$$\tilde{S}_\nu^k = \left\{\vec{x}\in\Omega : DT^\ell(x_0)\left(\left|B^\ell\vec{v}\right|^2 - \left(B^\ell\vec{v}\cdot\vec{u}\right)^2\right)^{\frac12} \le \nu, \vec{v} = \vec{x} - (x_0)^n\right\}.$$

Let us consider the special case when M has constant columns, that is $M_{i,j} = p_j$, where $\vec{p} = (p_1,p_2,\dots,p_n)$ is a probability vector. Then $M^\ell = M$ for $\ell = 1,2,\dots$ and

$$B^\ell = ((1-\gamma)\mathrm{id} + \gamma M)^\ell = (1-\gamma)^\ell\mathrm{id} + \left(1-(1-\gamma)^\ell\right)M$$

which yields

$$B^\ell\vec{v} = (1-\gamma)^\ell\vec{v} + \left(1-(1-\gamma)^\ell\right)\sqrt{n}(\vec{v}\cdot\vec{p})\vec{u},$$

as $M\vec{v}=\sqrt{n}(\vec{v}\cdot\vec{p})\vec{u}$. Thus

$$|B^\ell\vec{v}|^2-\left((B^\ell\vec{v})\cdot\vec{u}\right)^2=(1-\gamma)^{2\ell}\left(|\vec{v}|^2-V^2\right),$$

where $V=\sum_{j=1}^n v_j$. If we can choose $x_0=\frac{1}{n}\sum_{j=1}^n x_j$ then $V=0$ and the distance $\left(|B^\ell\vec{v}|^2-\left((B^\ell\vec{v})\cdot\vec{u}\right)^2\right)^{\frac{1}{2}}$ is equal to $(1-\gamma)^\ell|\vec{v}|$. For this the points $\vec{x}$ and $(x_0)^n$ have to lie in the same connected partition element of continuity for $\hat{T}$.

Since $M_{i,j}=p_j\forall i,j$ and if we choose $x_0=\frac{1}{n}\sum_{j=1}^n x_j$ we obtain $B^\ell\vec{v}\cdot\vec{u}=0$ and

$$\left|B^\ell\vec{v}\right|=(1-\gamma)^\ell|\vec{v}|=(1-\gamma)^\ell d(\vec{x},\Delta).$$

Consequently

$$|\vec{v}(\ell)|=(1-\gamma)^\ell|DT^\ell(x_0)|\cdot|\vec{v}|$$

and $d(\hat{T}^\ell\vec{x},\Delta)=(1-\gamma)^\ell|DT^\ell(x_0)|d(\vec{x},\Delta)+o(d(\vec{v},\Delta))$. Therefore in linear approximation

$$\tilde{S}_\nu^k=\left\{\vec{x}\in\Omega: d(\vec{x},\Delta)\le\frac{\nu}{DT^\ell(x_1)(1-\gamma)^\ell}, \ell=0,1,\dots,k\right\}$$

and since T is expanding only the term $\ell=k$ is relevant.

Denote by $\mathcal{D}^k$ the set of discontinuity points for $\hat{T}^\ell$ for $\ell=1,\dots,k$. We assume that $\mathcal{D}^k$ is a union of piecewise smooth hyper surfaces which intersect the diagonal Δ transversally. Then $\mathcal{D}^k\cap\Delta=\{(y_1)^n,(y_2)^n,\dots,(y_m)^n\}$ consists of finitely many points $(y_j)^n\in\Delta$. For each j denote by $\varphi_j=\angle(\Delta,\mathcal{D}^k)$ the angle between Δ and $\mathcal{D}^k$ at the point of intersection $(y_j)^n$. Clearly the angles φ_j are bounded away from 0 and we can put $r=2\nu(\cot\varphi+1)$ where $\varphi=\min_j\varphi_j$. If we put $\Delta_\nu^k=\Delta\setminus\bigcup_j B_r((y_j)^n)$ then for all ν small enough $B_\nu(\Delta_\nu^k)\cap\mathcal{D}^k=\varnothing$.

In order to compute $\mu(S_\nu^k)$ and $\mu(S_\nu)$ put

$$S_\nu^k(x_1)=\left\{(x_2,x_3,\dots,x_n)\in[0,1]^{n-1}:|T^\ell(x_1)-T^\ell(x_j)|\le\nu, j=2,\dots,n,\ \ \ell=1,\dots,k\right\}.$$

Then $S_\nu^\ell=\bigcup_{x_1\in[0,1]}\{x_1\}\times S_\nu^\ell(x_1)$ for $\ell=1,\dots,k$. In the same fashion we can look at the linear appproximation and put

$$\tilde{S}_\nu^k(x_1)=\left\{(x_2,x_3,\dots,x_n)\in[0,1]^{n-1}:|DT^\ell(x_1)|\cdot|x_1-x_j|\le\nu, j=2,\dots,n,\ \ \ell=1,\dots,k\right\}.$$

By the C^2-regularity of the maps one obtains

$$\int_{S_\nu^k(x_1)}dx_2\cdots dx_n=(1+\mathcal{O}(\nu))\int_{\tilde{S}_\nu^k(x_1)}dx_2\cdots dx_n=(1+\mathcal{O}(\nu))\left(\frac{2\nu}{(1-\gamma)|T^k(x_1)|}\right)^{n-1}$$

As we concluded above, we obtain by regularity of the density h that

$$\mu(S_\nu)=\int_{S_\nu}h(\vec{x})\,d\vec{x}=(1+o(1))\int_{S_\nu}h((x_1)^n)\,d\vec{x}$$

where the second integral is

$$\int_{S_\nu}h((x_1)^n)\,d\vec{x}=\int_{[0,1]}\int_{S_\nu^0(x_1)}h((x_1)^n)\,dx_2\cdots dx_n\,dx_1=\int_{[0,1]}h((x_1)^n)(2\nu)^{n-1}\,dx_1$$

as $\int_{S_\nu^0(x_1)}dx_2\cdots dx_n=(2\nu)^{n-1}$.

Similarly we obtain

$$\begin{aligned}\mu(S_\nu^k) &= (1+o(1))\int_{S_\nu^k} h((x_1)^n)\,d\vec{x}\\ &= (1+o(1))\int_{[0,1]}\int_{S_\nu^k(x_1)} h((x_1)^n)\,dx_2\cdots dx_n\,dx_1\\ &= (1+o(1))\int_{[0,1]} h((x_1)^n)\left(\frac{2\nu}{(1-\gamma)|T^k(x_1)|}\right)^{n-1}\,dx_1.\end{aligned}$$

Finally

$$\hat{\beta}_{k+1} = \lim_{\nu\to 0}\frac{\mu(S_\nu^k)}{\mu(S_\nu)} = \frac{\int_I \frac{h((x)^n)}{|DT^k(x)|^{n-1}}\,dx}{(1-\gamma)^{k(n-1)}\int_I h((x)^n)\,dx}.$$

□

The second ingredient to Theorem 4 is the following lemma which establishes that all returns to S_ν within a cluster are of first order which makes Δ look like a fixed point. That is $\hat{\beta}_k = \hat{\alpha}_K$:

Lemma 6. *Under the assumptions of Theorem 4*

$$\hat{\alpha}_{k+1} = \lim_{\nu\to 0}\frac{\mu(S_\nu^k)}{\mu(S_\nu)}.$$

Proof. We follow the proof of Proposition 5.3 in [8] adapted to our setting. We begin to consider again the set $\Delta_\nu^k = \Delta \setminus \bigcup_j B_r((y_j)^n)$. The ν-neighborhood of Δ, Δ_ν^k, will be a subset of S_ν with empty intersection with the discontinuity surfaces $\mathcal{D}^k$ of the maps $\hat{T}^\ell$ for $\ell = 1,\dots,k$. We put $G_1(\nu) := \bigcup_j B_r((y_j)^n)$. For reasons which will be clear in a moment, we now remove from the ν-neighborhood of the diagonal, another set. Consider the intersection points $\{(z_1)^n,(z_2)^n,\dots,(z_l)^n\}$ of Δ with the images of the discontinuity surfaces $\mathcal{D}$ of $\hat{T}$ only, and as we did previously we introduce the set $G_2(\nu) := \bigcup_i B_{2r}((z_i)^n)$, where we double the radius to allow an upcoming construction. Notice that with the choice of r given above, we have that $\mu(G_1(\nu)) = o(\mu(S_\nu)$, and $\mu(G_2(\nu)) = o(\mu(S_\nu)$, when $\nu \to 0$.

Let us take a point $x \in \Delta_\nu^k$ and a neighborhood $\mathcal{O}(x)$ such that $\mathcal{O}(x)\cap\Lambda \neq \varnothing$, and $\mathcal{O}(x)\cap(\mathcal{D}^k\cup\hat{T}^{-1}(\mathcal{D}^k)\cup\cdots\hat{T}^{-k}(\mathcal{D}^k)) = \varnothing$. With these assumptions, $\hat{T}^\ell$ for $\ell = 1,\dots,k$ are open maps on $\mathcal{O}(x)$. In particular, $\hat{T}^k(\mathcal{O}(x))$ will be included in the interior of one of the U_l and it will intersect Δ by the forward invariance of the latter. We now suppose that $\hat{T}^k(x)$ is in S_ν and we prove that $\hat{T}^{k-1}(x)$ is in S_ν too. Let us call D_* the domain of the function $\hat{T}_*^{-1}$, namely the inverse branch of the map sending $\hat{T}^{k-1}(x)$ to $\hat{T}^k(x)$. If the distance between $\hat{T}^k(x)$ and any point $z \in \hat{T}^k(\mathcal{O}(x))\cap\Delta$, such that the segment $[\hat{T}^k(x), z]$ is included in D_*, is less than ν, we have done since $\text{dist}(\hat{T}_*^{-1}(z), \hat{T}_*^{-1}(\hat{T}^k(x)) = \text{dist}(\tilde{z}, T^{\hat{k}-1}(x)) \le \lambda\nu$, where $\tilde{z} = \hat{T}_*^{-1}(z) \in \Delta$. Notice that such a point $z \in \Delta$ should not be necessarily in $\hat{T}^k(\mathcal{O}(x))$, provided the segment $[\hat{T}^k(x), z] \in D_*$ and $\text{dist}(z, \hat{T}^k(x)) \le \nu$. What could prevent the latter conditions to happen is the presence of the boundaries of the domains of definition of the preimages of $\hat{T}$, which are the images of $\mathcal{D}$. We should therefore avoid that $\hat{T}^k(x)$ lands in the set $G_2(\nu)$, which, with the choice of doubling the radius r, is enough large to allow the point $\hat{T}^k(x) \in G_2(\nu)^c$ to be joined to Δ with a segment included in D_*. We have therefore to discard those points $x \in S_\nu$ which are in $\hat{T}^{-k}G_2(\nu)$ and, by invariance, the measure of those point is bounded from above by $\mu(G_2(\nu))$. We now iterate backward the process to guarantee that $T^{k-2}(x)$ is in S_ν too. At this regard we must avoid again that $T^{k-2}(x) \in G_2(\nu)$, which means we have to remove a new portion of points of measure $\mu(G_2(\nu))$ from S_ν; at the end we will have k times of this measure of order $o(\nu)$. In conclusion the points which are not in $\bigcup_{l=2}^{k} \hat{T}^{-l}G_2(\nu) \cap S_\nu \cap G_1(\nu)$ gives zero contribution to the quantity $\mu(S_\nu^k)$, while the measure of the remaining points divided by $\mu(S_\nu)$ goes to zero for ν tending to zero. □

Proof of Theorem 4. Let μ be the absolutely continuous invariant measure on Ω. By Lemma 6 the values of $\hat{\alpha}_k$ are given by the expression in the statement of the theorem. The parameters $t\lambda_k$ are then given by Theorem 2 since the assumption $\sum_k k\hat{\alpha}_k < \infty$ is satisfied by uniform expansiveness which implies that $\hat{\alpha}_k$ decay exponentially fast.

In order to apply Theorem 1 it remains to verify Assumptions (I)–(VI). Assumption (IV) is satisfied for any $d_0 < 1 < d_1$ arbitrarily close to $n-1$. Since the unstable manifold is all of Ω, Assumption (V) is satisfied for any $u_0 < n-1$ arbitrarily close to $n-1$. Similarly, Assumption (VI) is satisfied with $\beta = \eta = 1$. Assumption (III) is satisfied as T is uniformly expanding (III-i) is trivially satisfied with $q = \infty$. (III-ii) follows from the regularity of the map and (III-iii) is satisfied since the contraction is in fact exponential. Assumption (II) is satisfied by a result of Saussol [26] where the the decay of correlations is shown for functions of bounded variation vs $\mathscr{L}^1$. Since characteristic functions have bounded variation we can take $\phi = \tilde{\phi} = \mathbb{1}_U = \mathbb{1}_{B_\rho(\Gamma)}$ in Section 7.2 and since functions that are bounded in the supremum norm (as characteristic functions are) are automatically in $\mathscr{L}^1$ the assumption is fulfilled. □

In the special case when the coupling constant γ is equal to zero, then μ is the product measure of the absolutely continuous T-invariant measure $\hat{\mu}$ on the interval $I = [0, 1]$, that is $\mu = \hat{\mu} \times \hat{\mu} \times \cdots \times \hat{\mu}$, n times. Consequently the density h on the diagonal Δ is equal to $\hat{h}^n$, where $\hat{h} = \frac{d\hat{\mu}}{dx}$. Then we conclude as follows:

Corollary 2. *Let $\Omega = I^n$ and $T : \Omega \circlearrowleft$ be the n-fold product of a uniformly expanding map $T : I \circlearrowleft$ with a.c.i.m $\hat{\mu}$ with density $\hat{h}$. Then*

$$\hat{\alpha}_{k+1} = \frac{1}{\int_I \hat{h}^n(x)\, dx} \int_I \frac{\hat{h}^n(x)}{|DT^k(x)|^{n-1}}\, dx$$

and in particular

$$\begin{aligned} \lambda_k &= \frac{1}{\alpha_1}(\hat{\alpha}_k - 2\hat{\alpha}_{k+1} + \hat{\alpha}_{k+2}) \\ &= \frac{\alpha_1^{-1}}{\int_I \hat{h}^n(x)\, dx} \int_I \hat{h}^n(x) \left(\frac{1}{|DT^k(x)|^{n-1}} - \frac{2}{|DT^{k+1}(x)|^{n-1}} + \frac{1}{|DT^{k+2}(x)|^{n-1}} \right) dx, \end{aligned}$$

where

$$\alpha_1 = 1 - \hat{\alpha}_2 = \left(\int_I \hat{h}^n(x)\, dx \right)^{-1} \int_I \hat{h}^n(x) \left(1 - |DT(x)|^{-(n-1)}\right) dx$$

is the extremal index.

For $n = 2$ these formulas were derived by Coelho and Collet [7], Theorem 1.

By using the theory developed in the present article, the paper [5], section 6.1, considered a Markov map of the interval for which the density h is piecewise constant and the quantities $\hat{\alpha}_k$ were computed rigorously, see also our upcoming paper [13]. The interesting point is that the λ_k do not follow a geometric distribution; for the statistics of the number of visits, we got a compound Poisson distribution which is not Pólya-Aeppli.

10. Aknowledgements

SV thanks the Laboratoire International Associé LIA LYSM, the INdAM (Italy), the UMI-CNRS 3483, Laboratoire Fibonacci (Pisa) where this work has been completed under a CNRS delegation and the *Centro de Giorgi* in Pisa for various supports.
NH thanks the University of Toulon and the Simons Foundation (award ID 526571) for support.
The authors thanks the anonimous referees whose comments and suggestion helped them to improve the paper.

References

[1] M Abadi: Hitting, returning and the short correlation function; Bull. Braz. Math. Soc. **37(4)** (2006), 1–17.

[2] M Abadi, A C M Freitas and J M Freitas: Dynamical Counterexamples Regarding the Extremal Index and the Mean of the Limiting Cluster Size Distribution; preprint available at `https://arxiv.org/pdf/1808.02970.pdf`.

[3] M Abadi, A C M Freitas and J M Freitas: Clustering Indices and Decay of Correlations in Non-Markovian Models; preprint available at `https://arxiv.org/pdf/1810.03216.pdf`.

[4] V S Afraimovich and L A Bunimovich: Which hole is leaking the most: a topological approach to study open systems; Nonlinearity **23** (2010), 643–656.

[5] Th Caby, D Faranda, S Vaienti, P Yiou: On the Computation of the Extremal Index for Time Series, Journal of Statistical Physics, in press.

[6] J-R Chazottes and P Collet: Poisson approximation for the number of visits to balls in nonuniformly hyperbolic dynamical systems; Ergod. Th. & Dynam. Sys. **33** (2013), 49–80.

[7] Z Coelho and P Collet: Asymptotic limit law for the close approach of two trajectories in expanding maps of the circle; Prob. Theory Relat. Fields **99** (1994), 237–250.

[8] D Faranda, H Ghoudi, P Guiraud and S Vaienti: Extreme value theory for synchronization of coupled map lattices; Nonlinearity **31(7)** (2018), 3326–3358.

[9] A C M Freitas, J M Freitas, M Todd: The compound Poisson limit ruling periodic extreme behaviour of non-uniformly hyperbolic dynamics; Comm. Math. Phys. **321(2)** (2013), 483–52.

[10] J M Freitas, N Haydn, and M Nicol: Convergence of rare event point processes to the Poisson process for planar billiards, Nonlinearity; **27(7)** (2014), 1669–1687

[11] A C M Freitas, J M Freitas, M Magalhães: Convergence of Marked Point Processes of Excesses for Dynamical Systems; J. Eur. Math. Soc. (JEMS), **20(9)** (2018), 2131–2179.

[12] D Faranda, A C Moreira Freitas, J Milhazes Freitas, M Holland, T Kuna, V Lucarini, M Nicol, M Todd, S Vaienti; *Extremes and Recurrence in Dynamical Systems*, Wiley, New York, 2016.

[13] S Gallo, N Haydn, S Vaienti, in preparation.

[14] N Haydn and Y Psiloyenis: Return times distribution for Markov towers with decay of correlations; Nonlinearity **27(6)** (2014), 1323–1349.

[15] N Haydn and S Vaienti: The distribution of return times near periodic orbits; Probability Theory and Related Fields **144** (2009), 517–542.

[16] N Haydn and K Wasilewska: Limiting distribution and error terms for the number of visits to balls in non-uniformly hyperbolic dynamical systems; Discr. Cont. Dynam. Syst. **36(5)** (2016), 2585–2611.

[17] N Haydn and F Yang: A Derivation of the Poisson Law for Returns of Smooth Maps with Certain Geometrical Properties; Contemporary Mathematics Proceedings in memoriam Chernov 2017
[18] G Keller: Generalized bounded variation and applications to piecewise monotonic transformations; Z. Wahr. verw. Geb.
[19] Y Kifer, A Rapaport: Poisson and compound Poisson approximations in conventional and nonconventional setups, Probab. Th. Related Fields, **160**, (2014), 797–831.
[20] Y Kifer, F Yang: Geometric law for numbers of returns until a hazard under ϕ-mixing, arXiv:1812.09927
[21] M R Leadbetter: Extremes and local dependence in stationary sequences; Z. Wahrsch. verw. Gebiete, **65(2)** (1983), 291–306.
[22] B Pitskel: Poisson law for Markov chains; Ergod. Th. & Dynam. Syst. **11** (1991), 501–513.
[23] B Saussol, F Pène: Back to balls in billiards; Communications in mathematical physics **293(3)** (2010),837–866.
[24] B Saussol, F Pène: Poisson law for some nonuniformly hyperbolic dynamical systems with polynomial rate of mixing; Ergodic Theory Dynam. Systems, **36(8)** (2016), 2602–2626.
[25] B Saussol, F Pène: Spatio-temporal Poisson processes for visits to small sets; available at `arxiv.org/abs/1803.06865`.
[26] B Saussol: Absolutely continuous invariant measures for multidimensional expanding maps; Israel J. Math. **116** (2000), 223–248.
[27] R L Smith: A counterexample concerning the extremal index; Adv. Appl. Probab. **20(3)** (1988), 681–683.
[28] F Yang: Rare event process and entry times distribution for arbitrary null sets on compact manifolds; preprint 2019 available at `https://arxiv.org/abs/1905.09956`