Entry Times Distribution for Mixing Systems

N. Haydn & F. Yang

Journal of Statistical Physics
1

ISSN 0022-4715

J Stat Phys
DOI 10.1007/s10955-016-1487-y

J Stat Phys
DOI 10.1007/s10955-016-1487-y

Entry Times Distribution for Mixing Systems

N. Haydn[1] · **F. Yang**[1]

Received: 24 October 2015 / Accepted: 18 February 2016

Abstract We consider the return times dynamics to Bowen balls for continuous maps on metric spaces which have invariant probability measures with certain mixing properties. These mixing properties are satisfied for instance by systems that allow Young tower constructions. We show that the higher order return times to Bowen balls are in the limit Poisson distributed. We also provide a general result for the asymptotic behavior of the recurrence time for Bowen balls for ergodic systems and those with specification.

Keywords Return times distribution · Dynamic balls · Bowen balls · Recurrence times

1 Introduction

Recently there has been a great interest in the statistics of return times to small sets and their limiting distributions as the target sets shrink to a point and the observation time is scaled accordingly as suggested by Kac's theorem. Lacroix and Kupsa [19,20] have shown that the shrinking of the target sets has to be done in a dynamical or geometric regular way as their examples show that otherwise any limiting distribution could be achieved. The first dynamical result is due to Doeblin [11] who showed that for the Gauss map higher order returns to cylinder-like neighbourhood of the origin are Poissonian distributed in the limit. In main stream dynamics, Pitskel [22] was the first one to consider the limiting distribution for Axiom A systems and showed in 1990 that for cylinders the return times are Poissonian in the limit and, by an approximation argument, also for metric balls for hyperbolic maps on two dimensional torii. In the successive years, a sequence of results then established that returns to cylinder sets in the limit become Poissonian under increasingly general conditions

✉ N. Haydn
nhaydn@usc.edu

F. Yang
yang617@usc.edu

1 Department of Mathematics, University of Southern California, Los Angeles, CA 90089-2532, USA

(see e.g. [2,4,9,10,12–14,17,18,26]. Similar results have recently been proven for geometric balls (see e.g. [8,15,21]). For dynamical balls, which are the metric equivalent of cylinder sets and which are used in the construction of equlibrium states and the formulation of entropy for continuous maps on metric spaces, much less is known. According to a result of Varandas [24] the exponential growth rate of the recurrence time equals the entropy. Previously, Brin and Katok [7] have proven a Shannon-McMillan-Breiman type theorem for Bowen balls. This paper builds on [16] where limiting distributions of entry and return times were determined together and rates of convergence were given. The principal assumption is that the given invariant probability measure is ϕ-mixing or α-mixing. Although this seems restrictive, all systems that allow a Young tower construction [27,28] do satisfy the α-mixing property.

In the next section we give the main results. In Sect. 2.5 we prove Theorem 3 which states that for ergodic, positive entropy systems the minimal recurrence time grows at least linearly. In Sect. 4 we prove a general result on the higher order return distributions for α-mixing systems, where the return sets can be unions of cylinders over a countably infinite alphabet. For that purpose we use the Chen-Stein method of which we give a short sketch at the beginning of the section. This result is then used in Sect. 5 to prove the first two main theorems which in fact follow from the more general Theorem 5.

2 Main Results

Let X be a compact metric space and $T : X \to X$ a continuous mapping. We equip X with the Borel sigma-algebra and assume that there is a T-invariant probability measure μ which we assume to be ergodic. We denote by $h = h(\mu)$ its measure theoretic entropy which is assumed to be positive.

For $A \subset X$ we then denote by $W_{A,m}(x)$ the number of visits of the orbit $\{T(x), T^2(x), \ldots, T^m(x)\}$ (for some $m \in \mathbb{N}$) to the set A, i.e.

$$W_{A,m}(x) = \sum_{j=1}^{m} \chi_A(T^j(x))$$

where χ_A is the characteristic function of the set A, i.e. $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise. The purpose of this paper is to get results on the distributions of $W_{A,m}$ in the case when the return set A is a Bowen ball and the cutoff values m for the length of the orbits are scaled by the measures of the return set, that is $m = [t/\mu(A)]$ for a positive parameter t.

The function $\tau_A(x) = \min\{j \geq 1 : T^j x \in A\}$ is the *entry time* if $x \in X$ and called the *return time* if $x \in A$. Clearly $W_{A,m}(x) = 0$ if the entry/return time $\tau_A(x)$ is larger than m.

2.1 Bowen Balls

Let X be a metric space and $T' : X \to X$ a continuous map. For $\epsilon > 0$ and $n \in \mathbb{N}$ one defines the (ϵ, n)-*Bowen ball* (or *dynamical balls*) by:

$$B_{\epsilon,n}(x) = \left\{ y : \sup_{0 \leq k < n} d(T^k x, T^k y) < \epsilon \right\}.$$

Bowen balls have the property that they capture the local dynamics in metric spaces and are used to define entropy, pressure and prove the existence of equilibrium states for given potential functions (see e.g. [25]). In many ways Bowen balls play on metric spaces the rôle that cylinder sets play in symbolic systems. For instance, according to Brin and Katok [7] one has the metric analogue of the theorem of Shannon–McMillan–Breimann:

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}|\log\mu(B_{\epsilon,n}(x))| = h(\mu)$$

for μ-almost every x provided μ is ergodic and the entropy $h(\mu)$ is positive. Similarly, Varandas [24] provided us with the metric equivalent of Ornstein–Weiss' formula for the *recurrence time* $R_{\epsilon,n}(x) = \min\{j \geq 1 : T^j x \in B_{\epsilon,n}(x)\}$, according to which

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\frac{1}{n}\log R_{\epsilon,n}(x) = h(\mu)$$

μ almost everywhere for ergodic μ and $h(\mu) > 0$. In a previous paper [16] we adressed the distribution of the first entry and return times and showed that $\mathbb{P}(\tau_{B_{\epsilon,n}(x)} > t/\mu(B_{\epsilon,n}(x))) \to e^{-t}$ as $n \to \infty$ and then $\epsilon \to 0$ for almost every x. We also provided rates of convergence. Here we consider higher order returns, that is we are interested in the distribution of the random variable $W_{A,m}$ for $m = [t/\mu(A)]$, $t > 0$ a parameter, and in particular if $\mu(A) \to 0$ along a sequence of suitable sets.

2.2 Mixing Properties

Let $\mathcal{A}$ be a finite measurable partition of X and denote by $\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$ the nth join. Its elements are referred to as n-cylinders. We assume that $\mathcal{A}$ is generating, i.e. that $\mathcal{A}^\infty = \bigvee_{j=0}^{\infty} T^{-j}\mathcal{A}$ consists of singletons. For a set $Y \subset X$ we shall use the notation $A_n(Y) = \bigcup_{A\in\mathcal{A}^n,\, A\cap Y\neq\varnothing} A$ as the smallest union of n-cylinders that approximates Y from the outside. In particular $A_n(x)$ denotes the n-cylinder that contains x.

We shall require that the measure μ have some mixing property with respect to this partition $\mathcal{A}$. We consider two situations:

(i) We say that μ is ϕ-*mixing* if

$$|\mu(A\cap T^{-n-k}B) - \mu(A)\mu(B)| \leq \phi(k)\mu(B)$$

for all $A \in \sigma(\mathcal{A}^n)$, $B \in \sigma(\bigcup_{j=1}^{\infty}\mathcal{A}^j)$, where $\phi(k)$ is a decreasing function that converges to 0.

(ii) We say that μ is α-*mixing* if

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

$\forall A \in \sigma(\mathcal{A}^n)$, $B \in \sigma(\bigcup_{j=1}^{\infty}\mathcal{A}^j)$, for some decreasing function $\alpha(k)$ that converges to 0.

2.3 Regularity of the Measure

In order to carry through an approximation scheme that involves cylider sets, we have to assume that the invariant measure has some regularity. Denote by $B(x,\epsilon)$ the metric ball in X with centre x and radius ϵ. Then for $x \in X$, $0 < \delta < \epsilon$ define the function

$$\psi(\epsilon,\delta,x) = \frac{\mu(B(x,\epsilon+\delta)) - \mu(B(x,\epsilon-\delta))}{\mu(B(x,\epsilon))}$$

as in [16]. In other words, ψ measures the proportion of the measure of the annulus $B(x,\epsilon+\delta)\setminus B(x,\epsilon-\delta)$ to the ball $B(x,\epsilon)$.

Example If X is the unit interval and T a map on X, then if μ is an absolutely continuous measure, one has $\psi(\epsilon,\delta,x) \le c_1\frac{\delta}{\epsilon}$ for some constant c_1 and all $x \in X$. More generally, let X be a compact Riemann manifold and T an expanding map. Then if the invariant measure μ is equivalent to the Riemann measure one again has $\psi(\epsilon,\delta,x) = \mathcal{O}(\frac{\delta}{\epsilon})$ for all $x \in X$. This follows from the fact that $\mu(B(x,\epsilon)) \sim \epsilon^d$, where d is the dimension of the manifold.

Remark Also let us note that sometimes the following annulus condition is used (see e.g. [21]):

$$\frac{\mu(B(x,\epsilon+\delta)\setminus B(x,\epsilon-\delta))}{\mu(B(x,\epsilon))} \le c_0\frac{\delta^g}{\epsilon^h}$$

for some g, h and a constant c_0. In this case $\psi = c_0\delta^g\epsilon^{-h}$. In Theorem 1 the requirement would be that h is positive and g can be arbitrary. In Theorem 2 $h = \xi$ and again g can be arbitrary.

2.4 Main Results on the Distribution

We can now formulate our main results on the limiting distribution of $W_{A,m}$ when A are Bowen balls and m follows the traditional Kac scaling. For a partition $\mathcal{A}$ we denote by $\mathrm{diam}(\mathcal{A}) = \sup_{A\in\mathcal{A}} \mathrm{diam}(A)$ the diameter of $\mathcal{A}$, where $\mathrm{diam}(A) = \sup_{x,y\in A} d(x,y)$.

Theorem 1 *Assume that the invariant measure μ is ϕ-mixing where $\phi(n) = \mathcal{O}(\frac{1}{n^{2+\kappa}})$ and diam$(\mathcal{A}^n) = \mathcal{O}(\gamma^{n^\xi})$ for some $\gamma < 1$, $\xi \le 1$ and $\kappa > 0$. Moreover assume that μ satisfies the following regularity condition*

$$\psi(\epsilon,\delta,x) \le \frac{C_\epsilon}{|\log\delta|^\zeta}$$

for some $\zeta > 1/\xi$, $C_\epsilon > 0$ independent of x. Let $t > 0$ and put $m = \frac{t}{\mu(B_{\epsilon,n}(x))}$.

Then there exists $\epsilon_0 > 0$ so that for every $\epsilon < \epsilon_0$ we have

$$\lim_{n\to\infty} \mathbb{P}\big(W_{B_{\epsilon,n}(x),m} = k\big) = e^{-t}\frac{t^k}{k!}$$

almost surely.

If the measure has better regularity then we can relax the condition on the diameter of cylinders and obtain the following statement:

Theorem 2 *Assume that there exist constants $\alpha,\kappa,\xi > 0$ satisfying $\alpha\xi > 1$, such that diam$(\mathcal{A}^n) = \mathcal{O}(n^{-\alpha})$, $\phi(n) = \mathcal{O}(n^{-(2+\kappa)})$ and*

$$\psi(\epsilon,\delta,x) \le C_\epsilon\delta^\xi$$

for some constant C_ϵ independent of δ and x.

Then there exists $\epsilon_0 > 0$ so that for every $\epsilon < \epsilon_0$ and $t > 0$ we have

$$\lim_{n\to\infty} \mathbb{P}\big(W_{B_{\epsilon,n}(x),m} = k\big) = e^{-t}\frac{t^k}{k!}$$

almost surely, where $m = \frac{t}{\mu(B_{\epsilon,n}(x))}$.

The proof of these two theorems is in Sect. 5. In the proof of these theorems we need some estimate on the minimum return time of points in $B_{\epsilon,n}(x)$, which is in the next section.

2.5 Recurrence Time for Dynamical Balls

For a set $A \subset X$ we have the first hitting time of a point x given by $\tau_A(x) = \min\{k > 0 : T^k(x) \in A\}$. The *period* of the set A is then given by

$$\tau(A) = \min\{k > 0 : T^{-k}(A) \cap A \neq \emptyset\}$$

which evidently equals $\tau(A) = \min_{x\in A} \tau_A(x)$. We can now formulate our second main result which is well known for cylinder sets [23].

Theorem 3 *Assume that μ is ergodic with entropy $h(\mu) > 0$.*

(i) Then for almost every $x \in X$

$$\lim_{\epsilon\to 0} \liminf_{n\to\infty} \frac{\tau(B_{\epsilon,n}(x))}{n} \geq 1.$$

(ii) If, moroever, the map T has specification, then

$$\limsup_{n\to\infty} \frac{\tau(B_{\epsilon,n}(x))}{n} \leq 1$$

for all ϵ small enough.

Let us recall that a map $T : X \to X$ has *specification* if for every $\epsilon > 0$ there exists a separation time $K(\epsilon)$ so that any two (in fact arbitrarily many) orbit segments $T^j x$, $j = 0, 1, \dots, n_x$ and $T^j y$, $j = 0, 1, \dots, n_y$ can be ϵ-shadowed by an actual orbit, that is there exist a point $z \in X$ and $m \leq K$ such that $d(T^j z, T^j x) < \epsilon$ for $j = 0, 1, \dots, n_x$ and $d(T^{n_x+1+m+j} z, T^j y) < \epsilon$ for $j = 0, 1, \dots, n_y$.

3 Proof of Theorem 3

In order to prove the lower bound (i) we need the following lemma.

Lemma 1 *[7] Let $\mathcal{A}$ be a finite generating partition with $\mu(\partial\mathcal{A}) = 0$. Then for all $\delta > 0$ there exist $N > 0$ and a set D_N with $\mu(D_N) > 1 - \delta$, such that*

$$\left|\{A \in \mathcal{A}^n : A \cap B_{\epsilon,n}(x) \neq \emptyset\}\right| \leq e^{\delta n} \qquad \forall\, x \in D_N$$

for all ϵ small enough and $n \geq N$.

Proof For all $\epsilon > 0$, define $U_\epsilon(\mathcal{A}) = \bigcup_{A\in\mathcal{A}} U_\epsilon(A)$ where

$$U_\epsilon(A) = A \cap B(X \setminus A, \epsilon)$$

(where we write $B(Y, \epsilon) = \bigcup_{x\in Y} B(x, \epsilon)$). Since $\bigcap_{\epsilon>0} U_\epsilon(\mathcal{A}) = \partial\mathcal{A}$, we have $\lim_{\epsilon\to 0} \mu(U_\epsilon(\mathcal{A})) = 0$, and thus for every $\beta > 0$, there exists ϵ_0 small enough so that

$$\mu(U_\epsilon(\mathcal{A})) < \beta/2, \quad \text{for all} \quad \epsilon < \epsilon_0.$$

By the Birkhoff ergodic theorem,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \chi_{U_\epsilon(\mathcal{A})}(T^k(x)) < \beta/2, \quad \text{for a.e} \quad x \in X.$$

By Egorov's theorem there exists N_1 so that the set D_N defined by

$$D_N = \left\{ x \in X : \frac{1}{n} \sum_{k=0}^{n-1} \chi_{U_\epsilon(\mathcal{A})}(T^k(x)) < \beta \quad \forall n \geq N \right\}$$

satisfies

$$\mu(D_N) > 1 - \delta, \quad \text{for all} \quad N > N_1.$$

Every n-cylinder $A_n(x)$ is identified by the n-word $x_0 x_1 \cdots x_{n-1}$ where $x_k \in \mathcal{A}$. We call this word the $(\mathcal{A}, n)$-name of $A_n(x)$. For all $y \in B_{\epsilon,n}(x)$ and $0 \leq k \leq n-1$, either $T^k(y) \in A_1(T^k(x))$, or $T^k(x) \in U_\epsilon(\mathcal{A})$. Now let us note that for all x in D_N, the frequency of the latter possibility (i.e. $T^k(x) \in U_\epsilon(\mathcal{A})$) is less than β. In other words, $d_n^H(x, y) < \beta$ for all $y \in B_{\epsilon,n}(x)$, $x \in D_N$ and $n > N$, where d_n^H is the Hamming distance given by $d_n^H(x, y) = \frac{1}{n} \sum_{k=0}^{n-1}(1 - \delta_{x_k, y_k})$ with δ denoting the Kronecker symbol.

If we denote $\mathcal{C}_{\beta,n}(x) = \{y \in X : d_n^H(x, y) < \beta\}$ the cluster of n-cylinders centred at x then

$$B_{\epsilon,n}(x) \subset A_n(B_{\epsilon,n}(x)) \subset \mathcal{C}_{\beta,n}(x), \quad \forall x \in D_N, n > N.$$

Since $d_n^H(x, y) = 0$ if the points x, y lie in the same element of $\mathcal{A}^n$, $\mathcal{C}_{\beta,n}(x)$ is a union of at most λ_n elements in $\mathcal{A}^n$, where λ_n can be estimated by

$$\lambda_n \leq \sum_{m=0}^{[n\beta]} |\mathcal{A}|^m \binom{n}{m}. \tag{1}$$

Using Stirling's formula, it is easy to show that

$$\limsup_{n\to\infty} \frac{\log \lambda_n}{n} \leq \beta \log |\mathcal{A}| - \beta \log \beta - (1 - \beta) \log(1 - \beta).$$

The right-hand-side converges to 0 as β approaches 0. For any given $\delta > 0$ we can take β small enough such that $\lambda_n \leq e^{\delta n}$ for all $n \geq N$ for some large enough N. In particular

$$\left|\{A \in \mathcal{A}^n, A \cap B_{\epsilon,n}(x) \neq \emptyset\}\right| \leq e^{\delta n}.$$

□

Proof of Theorem 3. Let $\delta > 0$ and D_N, N as in Lemma 1. Then for all $x \in D_N$ we have $B_{\epsilon,n}(x) \subset \mathcal{C}_{\beta,n}(x) = \{y \in X : d_n^H(x, y) < \beta\}$ with $\beta > 0$ being chosen below. Hence

$$\tau(B_{\epsilon,n}(x)) \geq \tau(\mathcal{C}_{\beta,n}(x)).$$

For arbitrary $\eta < 1$, fix $\zeta < \frac{1-\eta}{8} h$ ($h = h(\mu)$ the entropy of μ) small and let $E_N = \{x \in X : e^{-(h+\zeta)n} \leq \mu(\mathcal{A}^n(x)) \leq e^{-(h-\zeta)n}$ for all $n \geq N\}$. By the Theorem of Shannon-McMillan-Breiman, we can take N large such that $\mu(E_N) \geq 1 - \delta$. Set $G_N = D_N \cap E_N$, we have $\mu(G_N) \geq 1 - 2\delta$. For a large enough constant c_1 (depending on N) we achieve that $c_1^{-1} e^{-(h+\zeta)n} \leq \mu(\mathcal{A}^n(x)) \leq c_1 e^{-(h-\zeta)n}$ hold for all $n > 0$.

Define

$$B_n = \{x \in G_N : \tau(B_{\epsilon,n}(x)) < \eta n\}$$

and

$$\tilde{B}_n = \{x \in G_N : \tau(\mathcal{C}_{\beta,n}(x)) < \eta n\}.$$

Clearly $B_n \subset \tilde{B}_n$ for all $n \leq N$.

If we put $R^n(k) = \{x \in G_N : \tau(\mathcal{C}_{\beta,n}(x)) = k\}$ $(k \leq [\eta n])$ then $\tilde{B}_n = \bigcup_{k=1}^{\eta n} R^n(k)$ (disjoint union). To be precise, if a point x lies in $R^n(k)$ then $T^j\mathcal{C}_{\beta,n}(x) \cap \mathcal{C}_{\beta,n}(x) = \varnothing$ for $j = 1, \ldots, k-1$ and there exists some $y \in \mathcal{C}_{\beta,n}(x)$ such that $T^k(y) \in \mathcal{C}_{\beta,n}(x)$. Hence we have $d_n^H(y, T^k y) \leq d_n^H(y, x) + d_n^H(x, T^k y) \leq 2\beta$. Set

$$\tilde{R}^n(k) = \{y \in G_N : d_n^H(y, T^k y) \leq 2\beta\}$$

and we obtain

$$R^n(k) \subset \{x \in X : \text{ there exist } y \in \tilde{R}^n(k) \text{ such that } d_n^H(x, y) \leq \beta\}. \tag{2}$$

First we estimate $\mu(\tilde{R}^n(k))$. For every $y \in \tilde{R}^n(k)$, let

$$A_n(y) = (y_1 \ldots y_k y_{k+1} \ldots y_{2k} \ldots y_{mk+1} \ldots y_n)$$

with $y_i \in \mathcal{A}$, $m = [\frac{n}{k}]$, then

$$A_n(T^k y) = (y_{k+1} \ldots y_{2k} y_{2k+1} \ldots y_{3k} \ldots y_{(m+1)k+1} \ldots y_{n+k}).$$

Let $g_i = \sum_{j=ik+1}^{(i+1)k}(1 - \delta_{y_j, y_{j+k}})$ for $i = 1, 2, \ldots, m$, where $\delta_{a,b}$ is the standard Kronecker symbol. That is g_i is the number of coordinates on which $y_{ik+1} \ldots y_{(i+1)k}$ and $y_{(i+1)k+1} \ldots y_{(i+2)k}$ differ. Obviously $g_i \leq k$ and also $\sum_{i=1}^m g_i \leq 2\beta n$ as $y \in \tilde{R}^n(k)$.

For given $(g_1, g_2, \ldots, g_m)$ and given k-word $y_1 y_2 \cdots y_k$, the total number of n-cylinders $A_n(y)$ that lie in the given $A_k(y) = (y_1 y_2 \cdots y_k)$ and for which $y \in \tilde{R}^n(k)$ is bounded from above by

$$\begin{aligned} a_{n,y_1,\ldots,y_k,g_1,\ldots,g_m} &\leq \binom{k}{g_1}|\mathcal{A}|^{g_1}\binom{k}{g_2}|\mathcal{A}|^{g_2}\cdots\binom{k}{h_m}|\mathcal{A}|^{g_m} \\ &\leq \binom{n}{2\beta n}|\mathcal{A}|^{2\beta n}. \end{aligned}$$

To simplify notation, we abbreviate the LHS to a_n. By Stirling's formula

$$\frac{\log a_n}{n} \leq 2\beta \log|\mathcal{A}| - (1-2\beta)\log(1-2\beta) - 2\beta\log 2\beta \to 0$$

as $\beta \to 0^+$. We can take $\beta > 0$ small such that $a_n \leq e^{\delta n}$ where $\delta > 0$ is as above.

Denote by $b_{n,k}$ the total number of such possible $(g_1, \ldots, g_m) \in \{1, 2, \ldots, k\}^m$. Then

$$b_{n,k} = \sum_{j=0}^{[2\beta n]}\binom{j+m-1}{m-1} = \binom{[2\beta n]+m}{m} = \binom{[2\beta n]+\frac{n}{k}}{\frac{n}{k}}$$

which, again by Stirling's formula, can be bound as follows:

$$\frac{\log b_{n,k}}{n} \leq f\Big(2\beta + \frac{1}{k}\Big) - f\Big(\frac{1}{k}\Big) - f\Big(2\beta\Big),$$

where we put $f(x) = x\log x$. Since $f(x) \to 0$ as $x \to 0$ and $f(x)$ is uniformly continuous on $(0, 2]$, we have $\lim_{\beta\to 0}\frac{\log b_{n,k}}{n} = 0$ and in particular

$$b_{n,k} \leq e^{\delta n}$$

if we only take $\beta > 0$ small enough.

We are now able to estimate the measure of the set $\tilde{R}^n(k)$ as follows, where the sum over $g_1, \ldots, g_m$ is as above,

$$\begin{aligned}\mu(\tilde{R}^n(k)) &\leq \sum_{y\in\tilde{R}^n(k)} \mu(A_n(y))\\ &\leq \sum_{A_k(y),y\in\tilde{R}^n(k)} \sum_{g_1,\ldots,g_m} a_n c_1 e^{-(h-\zeta)n}\\ &\leq \sum_{A_k(y),y\in\tilde{R}^n(k)} b_{n,k} c_1 e^{-(h-\zeta-\delta)n}\\ &\leq \sum_{A_k(y),y\in\tilde{R}^n(k)} c_1 e^{-(h-\zeta-2\delta)n}.\end{aligned}$$

For $y \in \tilde{R}^n(k) \subset G_N$, we have $c_1^{-1}e^{-(h+\zeta)k} \leq \mu(A_k(y))$, hence $1 \leq c_1 e^{(h+\zeta)k}\mu(A_k(y))$. Therefore

$$\begin{aligned}\mu(\tilde{R}^n(k)) \leq &\sum_{A_k(y),y\in\tilde{R}^n(k)} c_1^2 e^{-(h-\zeta-2\delta)n} e^{(h+\zeta)k}\mu(A_k(y))\\ \leq &c_1^2 e^{-(h-\zeta-2\delta)n+(h+\zeta)k}.\end{aligned}$$

Since $c_1^{-1}e^{-(h+\zeta)n} \leq \mu(A_n(x)) \leq c_1 e^{-(h-\zeta)n}$ for every n-cylinder in G_N, $\tilde{R}^n(k)$ can be covered by at most $c_1^3 e^{-(h-\zeta-2\delta)n+(h+\zeta)k+(h+\zeta)n}$ many n-cylinders. Since according to (2) $R^n(k)$ is contained in the β-neighbourhood of $\tilde{R}^n(k)$ (under the Hamming metric d_n^H), and every β-neighbourhood of an n-cylinder contains at most $\lambda_n < e^{\delta n}$ many n-cylinders according to (1), the total number of n-cylinders that intersects $R^n(k)$ is bounded from above by

$$\lambda_n c_1^3 e^{-(h-\zeta-2\delta)n+(h+\zeta)k+(h+\zeta)n} \leq c_1^3 e^{(2\zeta+3\delta)n+(h+\zeta)k}.$$

Therefore,

$$\mu(R^n(k)) \leq c_1^3 e^{(2\zeta+3\delta)n+(h+\zeta)k} c_1 e^{-(h-\zeta)n} \leq c_1^4 e^{(-h+3\zeta+3\delta)n+(h+\zeta)k}.$$

Summing over k, we obtain with those estimates the following bound:

$$\begin{aligned}\mu(\tilde{B}_n) \leq &\sum_{k=1}^{\eta n} \mu(R^n(k))\\ \leq &\sum_{k=1}^{\eta n} c_1^4 e^{(-h+3\zeta+3\delta)n+(h+\zeta)k}\\ \leq &c_2 e^{(-h+3\zeta+3\delta)n+(h+\zeta)\eta n}\\ \leq &c_2 e^{(-(1-\eta)h+4\zeta+3\delta)n}.\end{aligned}$$

Since $B_n \subset \tilde{B}_n$ for all $n \geq N$, we have

$$\sum_n \mu(B_n) \leq N + \sum_{n>N} \mu(\tilde{B}_n) \leq N + \sum_{n>N} c_2 e^{(-(1-\eta)h+4\zeta+3\delta)n}.$$

We can choose $\delta < \frac{1-\eta}{8}$ and $\zeta < \frac{1-\eta}{8}h$, hence

$$-(1-\eta)h + 4\zeta + 3\delta \leq -\frac{1-\eta}{8}h < 0.$$

Therefore $\sum_{n=1}^{\infty}\mu(B_n)<\infty$. By the Borel–Contelli lemma, for almost every $x\in G_N$ we have $\liminf_n \frac{\tau(B_{\epsilon,n}(x))}{n}\geq\eta$. Since $\eta<1$ is arbitrary, the lower bound (i) of the theorem follows.

In order to get the upper bound (ii) for a map with specification let $K(\epsilon)$ be the separation time. Then there exists a point $z\in B_{\epsilon,n}(x)$ and an $m\leq K$ so that $T^{n+m}\in B_{\epsilon,n}(x)$. Hence $\tau(B_{\epsilon,n}(x))\leq n+K$ and consequently $\lim_{n\to\infty}\frac{1}{n}\tau(B_{\epsilon,n}(x))\leq 1$. □

4 α-Mixing Systems have Poisson Distributed Return Times for Unions of Cylinders

This section is on the return times to sets that are unions of cylinders, where the underlying partition $\mathcal{A}$ is allowed to be countably infinite. Recall that $W_{A,m}(x)$ is the number of visits of the orbit $\{T(x),T^2(x),\ldots,T^m(x)\}$ to the set A, i.e.

$$W_{A,m}(x)=\sum_{j=1}^{m}\chi_A(T^j(x)).$$

We then have the following result.

Theorem 4 *Let μ be α-mixing w.r.t. a finite or countably infinite partition $\mathcal{A}$ and let $A\in\sigma(\mathcal{A}^n)$. As before, let $\tau(A)$ be the period of A. For any $t>0$, let $m=\frac{t}{\mu(A)}$ and denote by ν_t the Poisson measure on $\mathbb{N}_0$ with parameter t. Then there exists a constant C_1 so that for every set $E\subset\mathbb{N}_0$*

$$|\mathbb{P}(W_{A,m}\in E)-\nu_t(E)|\leq C_1\min_{\tau(A)<\Delta<m}\left(\frac{\alpha(\Delta)}{\mu(A)}+\Delta\mu(A)+\mathbb{P}_A(\tau_A\leq\Delta)\right)(t+\log m).$$

For similar result see [2,4,5] where the Poisson distribution for ϕ-mixing measures was shown for single cylinders centred at a generic point. Let us note that for unions of cylinders it was shown in [3] that the hitting and first return times are exponentially distributed for α-mixing measures and in [14] that higher order returns are Poissonian in the limit for ϕ-mixing systems. To prove this theorem we use the Chen-Stein method similar to [14] where it was laid out in more detail than we do here although we shall proceed to give a summary of the procedure.

Let ν be a probability measure on $\mathbb{N}_0$ (equipped with the power σ-algebra $\mathscr{B}_{\mathbb{N}_0}$). If we denote by $\mathcal{F}$ the set of all real-valued functions on $\mathbb{N}_0$, then the Stein operator $\mathcal{S}:\mathcal{F}\to\mathcal{F}$ is defined by

$$\mathcal{S}f(k)=tf(k+1)-kf(k),\quad\forall k\in\mathbb{N}_0. \tag{3}$$

Denote by ν_t the Poisson-distribution measure with mean t, i.e. $\mathbb{P}_{\nu_t}(\{k\})=\frac{e^{-t}t^k}{k!}$ $\forall k\in\mathbb{N}_0$ then the Stein equation

$$\mathcal{S}f=h-\int_{\mathbb{N}_0}h\,d\nu_t \tag{4}$$

has a solution f for each ν_t-integrable $h \in \mathcal{F}$ (see [6]). The solution f is unique except for $f(0)$, which can be chosen arbitrarily.[1] In particular, if $h : \mathbb{N}_0 \to \mathbb{R}$ is bounded then so is the associated Stein solution f. A probability measure ν on $(\mathbb{N}_0, \mathscr{B}_{\mathbb{N}_0})$ is Poisson (with parameter t) if and only if [6] $\int_{\mathbb{N}_0} \mathcal{S} f \, d\nu = 0$ for all bounded functions $f : \mathbb{N}_0 \to \mathbb{R}$. The total variation distance of a probability measure ν from the Poisson distribution ν_t can then be estimated as follows:

$$|\nu(E) - \nu_t(E)| = \left| \int_{\mathbb{N}_0} \mathcal{S} f \, d\nu \right| = \left| \int_{\mathbb{N}_0} (tf(k+1) - kf(k)) \, d\nu \right| \tag{5}$$

where $E \subset \mathbb{N}_0$ and f is the Stein solution that corresponds to the indicator function χ_E. The following lemma on the function f associated to characteristic functions was proven in [14].

Lemma 2 *For the Poisson distribution ν_t with parameter t, the Stein solution of the Stein equation (4) that corresponds to the indicator function $h = \chi_E$, with $E \subset \mathbb{N}_0$, satisfies*

$$\left| f_{\chi_E}(k) \right| \le \begin{cases} 1 & \text{if} \quad k \le t \\ \frac{2+t}{k} & \text{if} \quad k > t \ . \end{cases} \tag{6}$$

In particular

$$\sum_{k=1}^{m} \left| f_{\chi_E}(k) \right| \le \begin{cases} m & \text{if} \quad m \le t \\ t + (2+t) \log \frac{m}{t} & \text{if} \quad m > t \ . \end{cases} \tag{7}$$

4.1 Return Times Distribution

Proof of Theorem 4. The Poisson parameter t is the expected value of $W_{A,m}$ which implies $t = \sum_{i=1}^{m} \mu\left(\chi_A T^i\right) = m\mu(A)$, where $\mu(T^{-i}A) = \mu(A)$ by invariance. If $h = \chi_E$ with $E \subset \mathbb{N}_0$ an arbitrary subset of the positive integers, then we obtain from (5) and (3)

$$\begin{aligned} |\nu(\mathcal{S}f)| = |\nu(h) - \nu_t(h)| &= \left|\mathbb{P}(W_{A,m} \in E) - \nu_t(E)\right| \\ &= \left|\mathbb{E}\left(tf(W_{A,m}+1) - W_{A,m} f(W_{A,m})\right)\right| . \end{aligned}$$

Hence we can proceed as follows:

$$\begin{aligned} \left|\mathbb{P}(W_{A,m} \in E) - \nu_t(E)\right| &= \left| t\mathbb{E} f(W_{A,m}+1) - \mathbb{E}\left(\sum_{i=1}^{m} I_i f(W_{A,m}) \right) \right| \\ &= \left| \sum_{i=1}^{m} p_i \mathbb{E} f(W_{A,m}+1) - \sum_{i=1}^{m} p_i \mathbb{E}(f(W_{A,m})|I_i = 1) \right| \\ &\le \sum_{i=1}^{m} p_i \left| \sum_{a=0}^{m} f(a+1)\mathbb{P}(W_{A,m} = a) - \sum_{a=0}^{m} f(a)\mathbb{P}(W_{A,m} = a|I_i = 1) \right| \\ &\le \sum_{i=1}^{m} p_i \sum_{a=0}^{m} f(a+1)\epsilon_{a,i}, \end{aligned} \tag{8}$$

[1] f can be computed recursively from the Stein equation:

$$f(k) = \frac{(k-1)!}{t^k} \sum_{i=0}^{k-1} (h(i) - \mu_0(h)) \frac{t^i}{i!} = -\frac{(k-1)!}{t^k} \sum_{i=k}^{\infty} (h(i) - \mu_0(h)) \frac{t^i}{i!}, \quad \forall k \in \mathbb{N}.$$

where we put $I_i(x) = \chi_A T^i(x)$ for the characteristic function of the set $T^{-i}A$ and

$$\epsilon_{a,i} = \left|\mathbb{P}(W_{A,m} = a) - \mathbb{P}(W_{A,m} = a + 1 | I_i = 1)\right|. \tag{9}$$

The function f above is the solution of the Stein equation (4) that corresponds to the indicator function $h = \chi_E$ in the Stein method and has been bounded in Lemma 2.

In order to estimate the error term $\epsilon_{a,i}$ put

$$W^i_{A,m} = W_{A,m} - \chi_A \circ T^i = \sum_{\substack{1\le j\le m\\ j\ne i}} \chi_A \circ T^j$$

(punctured sum). Then

$$\begin{aligned}\epsilon_{a,i} &= \left|\mathbb{P}(W_{A,m} = a) - \frac{\mathbb{P}\left(\left\{W^i_{A,m} = a\right\} \cap T^{-i}A\right)}{\mu(A)}\right| \\ &\le \left|\mathbb{P}(W_{A,m} = a) - \mathbb{P}(W^i_{A,m} = a)\right| + \frac{\xi_a}{\mu(A)}\end{aligned}$$

where $\xi_a = \max_i \left|\mathbb{P}(\{W^i_{A,m} = a\} \cap T^{-i}A) - \mathbb{P}(W^i_{A,m} = a)\mu(A)\right|$ is zero if all I_i are independent of each other. The first term is estimated by

$$\left|\mathbb{P}(W_{A,m} = a) - \mathbb{P}(W^i_{A,m} = a)\right| \le \mathbb{P}(I_i = 1) = \mu(A).$$

For the second term, which contains ξ_a, we proceed as follows.

Let $\Delta \ll m$ be a positive integer (the halfwidth of the gap) and put for every $i \in (0, m]$

$$W^{i,-}_{A,m} = \sum_{j=1}^{i-(\Delta+1)} \chi_A \circ T^j, \qquad W^{i,+}_{A,m} = \sum_{j=i+\Delta+1}^{m} \chi_A \circ T^j,$$

$$U^{i,-}_m = \sum_{j=i-\Delta}^{i-1} \chi_A \circ T^j, \qquad U^{i,+}_m = \sum_{j=i+1}^{i+\Delta} \chi_A \circ T^j,$$

with the obvious modifications if $i < \Delta$ or $i > m-\Delta$. With these partial sums we distinguish between the hits that occur near the i^{th} iteration, namely $U^{i,-}_m$ and $U^{i,+}_m$, and the hits that occur away from the i^{th} iteration, namely $W^{i,-}_{A,m}$ and $W^{i,+}_{A,m}$. Let us put $\tilde{W}^i_{A,m} = W^i_{A,m} - U^i_m = W^{i,-}_{A,m} + W^{i,+}_{A,m}$ for the total sum minus the $2\Delta+1$ terms in the gap surrounding the coordinate i. The gap allows us to use the mixing property in the terms $W^{i,\pm}_{A,m}$ and its size will be determined later when we optimise the error term.

Note that for $a \in \mathbb{N}_0$

$$\begin{aligned}\mathbb{P}\big(\{W_{A,m} = a+1\} \cap T^{-i}A\big) &= \mathbb{P}\big(\{W^i_{A,m} = a\} \cap T^{-i}A\big) \\ &= \sum_{\substack{\mathbf{a}=(a^-,a^{0,-},a^{0,+},a^+)\\ \text{s.t } |\mathbf{a}|=a}} \mathbb{P}\big(\{W^{i,\pm}_{A,m} = a^{\pm}\} \\ &\qquad \cap \{U^{i,\pm}_m = a^{0,\pm}\} \cap T^{-i}A\big).\end{aligned}$$

We split the following sum into three terms

$$\sum_a |f(a+1)| \cdot \left|\mathbb{P}\left(\left\{W^i_{A,m} = a\right\} \cap T^{-i}A\right) - \mathbb{P}\left(W^i_{A,m} = a\right)\mu(A)\right| \le R_1 + R_2 + R_3$$

and will estimate the three terms

$$R_1 = \sum_a |f(a+1)| \cdot \left| \mathbb{P}\left(\left\{W^i_{A,m} = a\right\} \cap T^{-i}A\right) - \mathbb{P}\left(\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A\right)\right|$$
$$R_2 = \sum_a |f(a+1)| \cdot \left| \mathbb{P}\left(\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}^i_{A,m} = a\right)\mathbb{P}\left(I_i = 1\right)\right|$$
$$R_3 = \sum_a |f(a+1)| \cdot \left| \mathbb{P}\left(\tilde{W}^i_{A,m} = a\right) - \mathbb{P}\left(W^i_{A,m} = a\right)\right| \mu(A)$$

separately.

Estimate of R_1: Observe that

$$\{W^i_{A,m} = a\} \cap T^{-i}A \subset \left(\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A\right) \cup \left(\left\{U^i_m > 0\right\} \cap T^{-i}A\right)$$
$$\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A \subset \left(\left\{W^i_{A,m} = a\right\} \cap T^{-i}A\right) \cup \left(\left\{U^i_m > 0\right\} \cap T^{-i}A\right).$$

Since $U^i_m = U^{i,+}_m + U^{i,-}_m > 0$ implies that either $U^{i,+}_m > 0$ or $U^{i,-}_m > 0$ we get $\left|\mathbb{P}(\left\{W^i_{A,m} = a\right\} \cap T^{-i}A) - \mathbb{P}(\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A)\right| \le \mathbb{P}(\left\{U^i_m > 0\right\} \cap T^{-i}A) \le b^-_i + b^+_i$ where

$$b^-_i = \mathbb{P}(\left\{U^{i,-}_m > 0\right\} \cap T^{-i}A) \quad \text{and} \quad b^+_i = \mathbb{P}(\left\{U^{i,+}_m > 0\right\} \cap T^{-i}A).$$

For b^+_i we obtain the estimate

$$b^+_i = \mathbb{P}(\left\{U^{i,+}_m > 0\right\} \cap T^{-i}A) = \mathbb{P}(U^{i,+}_m > 0 | I_i = 1)\mu(A) = \mathbb{P}_A(\tau_A \le \Delta)\mu(A)$$

and since in [14] it was shown that $b^-_i = b^+_i$ we obtain

$$R_1 \le c_2 \mathbb{P}_A(\tau_A \le \Delta)\mu(A) \sum_a |f(a+1)| \le c_3 \mathbb{P}_A(\tau_A \le \Delta)\mu(A)(t + \log m)$$

for some c_3 where we used Lemma 2 to estimate the sum over a.

Estimate of R_3: In order to show that short returns are negligible note that

$$\{W^i_{A,m} = a\} \subset \left\{\tilde{W}^i_{A,m} = a\right\} \cup \{U^i_m > 0\}$$
$$\{\tilde{W}^i_{A,m} = a\} \subset \left\{W^i_{A,m} = a\right\} \cup \{U^i_m > 0\}$$

which yields

$$\left| \mathbb{P}\left(\tilde{W}^i_{A,m} = a\right) - \mathbb{P}\left(W^i_{A,m} = a\right)\right| \le \mathbb{P}\left(U^i_m > 0\right) \le 2\mathbb{P}\left(\bigcup_{k=1}^{\Delta} \{I_{i+k} = 1\}\right) \le 2\Delta\mu(A),$$

and therefore

$$R_3 \le 2\Delta\mu(A)^2 \sum_a |f(a+1)| \le c_4 \Delta\mu(A)^2(t + \log m).$$

Estimate of R_2: This is the principal term and the speed of mixing now becomes relevant. Recall that $\tilde{W}^i_{A,m}(x) = W^{i,-}_{A,m}(x) + W^{i,+}_{A,m}(x)$ and we want to estimate

$$R_2 \le \sum_a |f(a+1)| \left| \mathbb{P}\left(\left\{\tilde{W}^i_{A,m} = a\right\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}^i_{A,m} = a\right)\mu(A)\right|$$
$$\le \sum_{a^-,a^+} |f(a^- + a^+ + 1)| \left(\mathbb{P}\left(\left\{\tilde{W}^{i,\pm}_m = a^\pm\right\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}^{i,\pm}_m = a^\pm\right)\mu(A)\right)\epsilon_{a^-,a^+},$$

where $\epsilon_{a^-,a^+} = \text{sgn}\left(\mathbb{P}\left(\{\tilde{W}^i_{A,m} = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}^i_{A,m} = a\right)\mu(A)\right)$. If we put

$$\mathcal{W}^+(a^-) = \bigcup_{a^+:\epsilon_{a^-,a^+}=+1} \{\tilde{W}^{i,+}_m = a^+\}, \qquad \mathcal{W}^-(a^-) = \bigcup_{a^+:\epsilon_{a^-,a^+}=-1} \{\tilde{W}^{i,+}_m = a^+\},$$

both disjoint unions, then

$$\begin{aligned} R_2 \le \sum_a |\varphi(a)| \Big| & \mathbb{P}\left(\left\{\tilde{W}^{i,-}_m = a^-\right\} \cap \mathcal{W}^+(a^-) \cap T^{-i}A\right) \\ & -\mathbb{P}\left(\tilde{W}^{i,-}_m = a^-\right)\mu\left(\mathcal{W}^+(a^-)\right)\mu(A)\Big| \\ + \sum_a |\varphi(a)| \Big| & \mathbb{P}\left(\left\{\tilde{W}^{i,+}_m = a^+\right\} \cap \mathcal{W}^-(a^+) \cap T^{-i}A\right) \\ & -\mathbb{P}\left(\tilde{W}^{i,+}_m = a^+\right)\mu\left(\mathcal{W}^-(a^+)\right)\mu(A)\Big| \end{aligned}$$

where $\varphi(a) = \sup_{a'>a}|f(a')|$ satisfies by Lemma 2 $\varphi(a) \le \min(1, \frac{t}{a})$. We have to estimate the two mixing terms, the first of which is for $a^- \ge 0$:

$$\begin{aligned} &\left|\mathbb{P}\left(\left\{\tilde{W}^{i,-}_m = a^-\right\} \cap \mathcal{W}^+(a^-) \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}^{i,-}_m = a^-\right)\mu\left(\mathcal{W}^+(a^-)\right)\mu(A)\right| \\ &\quad \le R_{2,1} + R_{2,2} + R_{2,3} \end{aligned}$$

where

$$\begin{aligned} R_{2,1} &= \left|\mathbb{P}\left(\left\{\tilde{W}^{i,-}_m = a^-\right\} \cap \mathcal{W}^+(a^-) \cap T^{-i}A\right) - \mathbb{P}\left(\left\{\tilde{W}^{i,-}_m = a^-\right\} \cap T^{-i}A\right)\mu\left(\mathcal{W}^+(a^-)\right)\right| \\ R_{2,2} &= \left|\mathbb{P}\left(\left\{W^{i,-}_{A,m} = a^-\right\} \cap T^{-i}A\right) - \mathbb{P}\left(W^{i,-}_{A,m} = a^-\right)\mu(A)\right|\mu\left(\mathcal{W}^+(a^-)\right) \\ R_{2,3} &= \left|\mathbb{P}\left(W^{i,-}_{A,m} = a^-\right)\mu\left(\mathcal{W}^+(a^-)\right) - \mathbb{P}\left(\left\{\tilde{W}^{i,-}_m = a^-\right\} \cap \mathcal{W}^+(a^-)\right)\right|\mu(A). \end{aligned}$$

We now bound the three terms separately: Due to the mixing property we get for the first term the estimate

$$R_{2,1} \le \alpha(\Delta).$$

Similarly for the second term

$$R_{2,2} \le \alpha(\Delta)\mu\left(\mathcal{W}^+(a^-)\right),$$

while the third term is estimated by

$$R_{2,3} \le \alpha(2\Delta)\mu(A).$$

Combining these estimates and considering that the second term in the above estimate of R_2 is estimated in the same manner we obtain

$$R_2 \le c_4\alpha(\Delta)\sum_a \varphi(a) \le c_6\alpha(\Delta)(t + \log m)$$

for some constant c_4.

Finally, putting together the error terms R_1, R_2 and R_3 yields

$$\begin{aligned}&\left|\mathbb{P}(W_{A,m}\in E)-\nu_t(E)\right|\\&\le\sum_{i=1}^{m}p_i\left(\sum_{a=0}^{m}|f(a+1)|\mu(A)+c_7\left(\frac{\alpha(\Delta)}{\mu(A)}+2\Delta\mu(A)+\mathbb{P}_A(\tau_A\le\Delta)\right)(t+\log m)\right)\\&\le c_8\left(\mu(A)+\frac{\alpha(\Delta)}{\mu(A)}+\Delta\mu(A)+\mathbb{P}_A(\tau_A\le\Delta)\}\right)(t+\log m)\end{aligned}$$

for some c_8 independent of A. □

5 Poisson Distributed Return Times for Bowen Balls

In this section we will prove Theorem 1. Recall that

$$\psi(\epsilon,\delta,x)=\frac{\mu(B(x,\epsilon+\delta)\setminus B(x,\epsilon-\delta))}{\mu(B(x,\epsilon))}$$

is the proportion of the measure of the annulus $B(x,\epsilon+\delta)\setminus B(x,\epsilon-\delta)$ to the ball $B(x,\epsilon)$. Put $\tau_A^k(x)=\tau_A\circ T^{\tau_A^{k-1}}$ for the kth return of x to the set A:

$$\tau_A^k(x)=\min\{k>\tau_A^{k-1}(x):T^k(x)\in A\}$$

where $\tau_A^1=\tau_A$.

We will prove the following more general theorem and then deduce Theorems 1 and 2.

Theorem 5 *Let μ be a ϕ-mixing T-invariant ergodic measure with positive entropy. Let $\gamma_n=diam(\mathcal{A}^n)$. Assume that there exist $\epsilon_0>0$ and an increasing sequence $\{N(n)\}_{n=1}^{\infty}$ satisfying $n<N(n)<\frac{1}{4}\mu(B_{\epsilon,n}(x))^{-1}$ and sequences $\vartheta_n(\epsilon)\to0$ as $n\to\infty$ ($\forall\,\epsilon<\epsilon_0$) such that*

$$\psi(\epsilon,\gamma_{N(n)-k},T^kx)\le\vartheta_n(\epsilon)\cdot\frac{\mu(B_{\epsilon,n}(x))}{n}\tag{10}$$

for all $\epsilon<\epsilon_0$, $x\in X$, $0\le k<n$.

Then for all $t>0$ one has

$$\lim_{n\to\infty}\mathbb{P}\left(W_{B_{\epsilon,n}(x),m}=k\right)=e^{-t}\frac{t^k}{k!},$$

where $m=\frac{t}{\mu(B_{\epsilon,n}(x))}$.

The idea of the proof is to use cluster of cylinders sets to approximate Bowen balls. For this purpose, for some integer $N(n)\gg n$, define

$$\widetilde{B}_{\epsilon,n}(x)=\bigcup_{A\in\mathcal{A}^{N(n)},A\subset B_{\epsilon,n}(x)}A$$

the union of all $N(n)$-cylinders contained in $B_{\epsilon,n}(x)$. If we put

$$\widetilde{\partial B}_{\epsilon,n}(x)=\bigcup_{A\in\mathcal{A}^{N(n)},A\cap\partial B_{\epsilon,n}(x)\ne\emptyset}A$$

as the union of all cylinders which intersect the boundary of $B_{\epsilon,n}(x)$, then

$$B_{\epsilon,n}(x)\setminus\widetilde{B}_{\epsilon,n}(x)\subset\widetilde{\partial B}_{\epsilon,n}(x).$$

The next lemma (c.f. [16]) allows us to estimate the difference between $\widetilde{B}_{\epsilon,n}(x)$ and $B_{\epsilon,n}(x)$.

Lemma 3 *Under the hypothesis of Theorem 5 we have*

$$\mu(\partial\widetilde{B}_{\epsilon,n}(x)) \leq \vartheta_n(\epsilon)\mu(B_{\epsilon,n}(x))$$

and in particular, $\mu(B_{\epsilon,n}(x))/\mu(\widetilde{B}_{\epsilon,n}(x)) = \mathcal{O}(1)$.

Proof Since T is continuous, $\partial B_{\epsilon,n}(x) \subset \bigcup_{k=0}^{n-1} T^{-k}\partial B(T^k x, \epsilon)$. Hence if $A_{N(n)} \cap \partial B_{\epsilon,n}(x) \neq \emptyset$ for some $N(n)$-cylinder $A_{N(n)}$, then $A_{N(n)-k}(T^k y) \cap \partial B(T^k x, \epsilon) \neq \emptyset$ for some $0 \leq k \leq n-1$ and $y \in A^{N(n)}$. Since $\operatorname{diam}(A_{N(n)-k}(T^k y)) \leq \gamma_{N(n)-k}$ we obtain

$$\begin{aligned}
\partial\widetilde{B}_{\epsilon,n}(x) &\subset \bigcup_{k=0}^{n-1} T^{-k}(B(\partial B(T^k x, \epsilon), \gamma_{N(n)-n})) \\
&\subset \bigcup_{k} T^{-k}(B(T^k x, \epsilon + \gamma_{N(n)-k}) \setminus B(T^k x, \epsilon + \gamma_{N(n)-k})),
\end{aligned}$$

and consequently we can bound the measure of this set as follows:

$$\begin{aligned}
\mu(\partial\widetilde{B}_{\epsilon,n}(x)) &\leq n \cdot \sup_{0\leq k\leq n-1} \mu(B(T^k x, \epsilon + \gamma_{N(n)-k}) \setminus B(T^k x, \epsilon + \gamma_{N(n)-k})) \\
&= n \cdot \sup_{0\leq k\leq n-1} \{\psi(\epsilon, \gamma_{N(n)-k}, T^k x) \cdot \mu(B(T^k x, \epsilon))\} \\
&\leq n \cdot \sup_{0\leq k\leq n-1} \{\psi(\epsilon, \gamma_{N(n)-k}, T^k x)\} \\
&\leq \vartheta_n(\epsilon)\mu(B_{\epsilon,n}(x))
\end{aligned}$$

where we used the assumption to bound the measure of the annuli. In particular $\mu(\partial\widetilde{B}_{\epsilon,n}(x))/\mu(\widetilde{B}_{\epsilon,n}(x)) = \vartheta_n \to 0$ and therefore $\mu(B_{\epsilon,n}(x))/\mu(\widetilde{B}_{\epsilon,n}(x)) = \mathcal{O}(1)$. □

Next we show that the limiting distribution for the hitting times of $B_{\epsilon,n}(x)$ can be approximated by the distribution of $\widetilde{B}_{\epsilon,n}(x)$. To simplify notation we write $B = B_{\epsilon,n}(x)$ and $\widetilde{B} = \widetilde{B}_{\epsilon,n}(x)$. For $t > 0$ we put $m = \frac{t}{\mu(B)}$ and $\tilde{m} = \frac{t}{\mu(\widetilde{B})}$ and write for simplicity's sake

$$\Theta_{B,m}(k) = \mathbb{P}\Big(W_{B,m} = k\Big), \quad \Theta_{\widetilde{B},\tilde{m}}(k) = \mathbb{P}\left(W_{\widetilde{B},\tilde{m}} = k\right)$$

and others similarly. The following approximation lemma does note depend on the mixing property.

Lemma 4 *For all* $t \geq 0$ *we have*

$$\left|\Theta_{B,m}(k) - \Theta_{\widetilde{B},\tilde{m}}(k)\right| \leq 2t \cdot \vartheta_n(\epsilon) \to 0$$

as $n \to \infty$.

Proof By the triangle inequality

$$\begin{aligned}
\left|\Theta_{B,m}(k) - \Theta_{\widetilde{B},\tilde{m}}(k)\right| &\leq \left|\Theta_{B,m}(k) - \Theta_{\widetilde{B},m}(k)\right| + \left|\Theta_{\widetilde{B},m}(k) - \Theta_{\widetilde{B},\tilde{m}}(k)\right| \\
&= I + II.
\end{aligned}$$

In order to estimate the first term note that $\widetilde{B} \subset B$ which implies $W_{B,m} \geq W_{\widetilde{B},m}$. Consequently

$$I \leq \mathbb{P}(W_{B\setminus\tilde{B},m} > 0) \leq \mathbb{P}(\tau_{B\setminus\tilde{B}} < m) \leq m\mu(B \setminus \tilde{B}).$$

For the second term we proceed as follows:

$$II = \mathbb{P}(\{W_{\tilde{B},m} = k\} \cap \{W_{\tilde{B},\tilde{m}} > k\}) \le \mu(\tilde{B})(\tilde{m} - m) = m\mu(B \setminus \tilde{B}).$$

Combining the estimates for I and II yields by Lemma 3

$$\begin{aligned}\left|\Theta_{B,m}(k) - \Theta_{\widetilde{B},\tilde{m}}(k)\right| &\le 2m \cdot \mu(B \setminus \tilde{B})\\ &\le 2m \cdot \mu(\widetilde{\partial B}_{\epsilon,n}(x))\\ &\le 2m \cdot \vartheta_n(\epsilon)\mu(B)\\ &= 2t\vartheta_n(\epsilon) \to 0.\end{aligned}$$

□

Before we prove Theorem 5 let us consider the case of α-mixing measures. As noted in [16] generalised SRB measures for systems that allow a Young tower construction as in [27,28] are α-mixing and thus are prime examples to which the following proposition can be applied. (Generalised Sinai–Ruelle–Bowen measures are on the unstable leaves absolutely continuous with respect to a given reference measure which in the traditional setting is taken to be the Lebesgue measure.) We though have to make an assumption on the short retun times.

Proposition 1 *Let μ be an α-mixing measure where $\alpha(k)$ decreases exponentially fast to* 0. *Let $\Delta_n = a\left|\log \mu(B_{\epsilon,n}(x))\right|$ where $a > 0$ is so that $\frac{\alpha(\Delta_n)}{\mu(B_{\epsilon,n}(x))}\Delta_n \to 0$ as $n \to \infty$. If $\mathbb{P}_{B_{\epsilon,n}(x)}(\tau_{B_{\epsilon,n}(x)} \le \Delta_n)\Delta_n \to 0$ then*

$$\lim_{n\to\infty} \mathbb{P}\left(W_{B_{\epsilon,n}(x),m} = k\right) = e^{-t}\frac{t^k}{k!}.$$

Proof By Lemma 4 it is sufficient to prove that

$$\lim_{n\to\infty} \Theta_{\widetilde{B},\tilde{m}}(k) = e^{-t}\frac{t^k}{k!}.$$

The result then follows from Theorem 4 with $m = 1/\mu(B_{\epsilon,n}(x))$. □

Let us now prove Theorem 5 where the ϕ-mixing property is used to control the short return times up to Δ.

Proof of Theorem 5. Again, by Lemma 4 it is enough to show that $\Theta_{\widetilde{B},\tilde{m}}(k) \to e^{-t}\frac{t^k}{k!}$ as $n \to \infty$. We apply Theorem 1 of [14] to the set $\widetilde{B}_{\epsilon,n}(x) \in \sigma(\mathcal{A}^{N(n)})$ and obtain (for some c_1)

$$\left|\Theta_{\widetilde{B},\tilde{m}}(k) - e^{-t}\frac{t^k}{k!}\right| \le c_1 t(t\vee 1)\inf_{\Delta>0}\left\{\Delta\mu(\tilde{B}) + \sum_{j=\tau(\tilde{B})}^{\Delta}\delta_{\widetilde{B}}(j) + \frac{\phi(\Delta)}{\mu(\tilde{B})}\right\}|\log\mu(\tilde{B})|,$$

where $\delta_{\widetilde{B}}(j) = \min_{1\le\omega\le j\wedge N(n)}\{\mu(A_\omega(\widetilde{B})) + \phi(j-\omega)\}$ and, as before, $A_\omega(\widetilde{B}) = \bigcup_{A\in\mathcal{A}^\omega, A\cap\widetilde{B}\ne\emptyset} A$. Let $\eta' \in (\frac{1}{2+\kappa}, 1)$ so that the gaps $\Delta = \mu(\widetilde{B})^{-\eta'}$ are larger than $N(n)$. Then

$$\left|\Theta_{\widetilde{B},\tilde{m}}(k) - e^{-t}\frac{t^k}{k!}\right| \le c_1 t(t\vee 1)\left(\mu(\tilde{B})^{1-\eta'} + \sum_{j=\tau(\widetilde{B})}^{\Delta}\delta_{\widetilde{B}}(j) + \mu(\tilde{B})^{-1+\eta'(2+\kappa)}\right)|\log\mu(\tilde{B})|.$$

Since $\mu(\tilde{B}) = \mathcal{O}(\mu(B_{\epsilon,n}(x)))$ we conclude by [7] $|\log \mu(\tilde{B})| = \mathcal{O}(n)$ and it thus remains to show that $\sum_{j=\tau(\tilde{B})}^{\Delta} \delta_{\tilde{B}}(j) = o(\frac{1}{n})$.

Since $\tilde{B} \subset B$ we get $A_\omega(\tilde{B}) \subset A_\omega(B)$ and therefore $A_\omega(B) \subset \widetilde{B}_{\epsilon,n}(x) \cup \widetilde{\partial B}_{\epsilon,n}(x)$ for all $\omega \geq N(n)$. As in Lemma 1 let us put

$$D_{N_0} = \left\{ x \in X : \frac{1}{n} \sum_{k=0}^{n-1} \chi_{U_\epsilon(\mathcal{A})}(T^k(x)) < \beta \quad \forall n \geq N_0 \right\}.$$

Then $A_\omega(B_{\epsilon,n}(x)) \subset A_n(B_{\epsilon,n}(x)) \subset \mathcal{C}_{\beta,n}(x)$ for all $x \in D_{N_0}$ and $\omega \geq n \geq N_0$.

By Theorem 3, we can take N_0 large enough such that the set

$$\left\{ x : \tau(B_{\epsilon,n}(x)) > \frac{n}{2} \quad \forall n > N_0 \right\}$$

has measure arbitrarily close to 1. Since $\widetilde{B}_{\epsilon,n}(x) \subset B_{\epsilon,n}(x)$ we conclude that

$$E_{N_0} = \left\{ x : \tau(\widetilde{B}_{\epsilon,n}(x)) > \frac{n}{2} \quad \forall n > N_0 \right\}$$

also has measure arbitrarily close to 1 for N_0 large enough. For $x \in G_{N_0} = D_{N_0} \cap E_{N_0}$, and all $n > 4N_0$ we then split the following sum into three parts:

$$\begin{aligned}
\sum_{j=\tau(\tilde{B})}^{\Delta} \delta_{\tilde{B}}(j) &= \sum_{j=\tau(\tilde{B})}^{\Delta} \min_{1 \leq \omega \leq j \wedge N(n)} \{\mu(A_\omega(\tilde{B})) + \phi(j-\omega)\} \\
&\leq \sum_{j=n/2}^{\Delta} \min_{1 \leq \omega \leq j \wedge N(n)} \{\mu(A_\omega(B)) + \phi(j-\omega)\} \\
&\leq \sum_{j=n/2}^{2n-1} \min_{1 \leq \omega \leq j} \{\mu(A_\omega(B)) + \phi(j-\omega)\} + \sum_{j=2n}^{N(n)} \min_{1 \leq \omega \leq j} \{\mu(A_\omega(B)) + \phi(j-\omega)\} \\
&\quad + \sum_{j=N(n)+1}^{\Delta} \min_{1 \leq \omega \leq N(n)} \{\mu(A_\omega(B)) + \phi(j-\omega)\} \\
&= I + II + III.
\end{aligned}$$

Since μ is ϕ-mixing there exists a $\pi < 1$ so that $\mu(A_m(x)) < \pi^m$ for all x and m large enough [1]. We now assume that $\beta > 0$ is small enough so that the size λ_m of the (β, m)-clusters $\mathcal{C}_{\beta,m}(x)$ satisfies $\lambda_m < \pi^{-\frac{m}{2}}$ for all $x \in D_{N_0}$ (see (1)). Thus

$$\mu(\mathcal{C}_{\beta,m}(x)) \leq \lambda_m \pi^m < \pi^{\frac{m}{2}} \tag{11}$$

for all m large enough and $x \in D_{N_0}$. We now estimate the three parts on the right hand side above using the mixing property as follows:

(I) For the term I, we also take $\omega = \frac{j}{2}$. Since $B_{\epsilon,n}(x) \subset B_{\epsilon,\frac{n}{4}}(x)$ and $\frac{j}{2} \geq \frac{n}{4} \geq N_0$ we have

$$A_{\frac{j}{2}}(B_{\epsilon,n}(x)) \subset A_{\frac{j}{2}}\left(B_{\epsilon,\frac{n}{4}}(x)\right) \subset A_{\frac{n}{4}}\left(B_{\epsilon,\frac{n}{4}}(x)\right) \subset \mathcal{C}_{\beta,\frac{n}{4}}(x).$$

The bound (11) then yields

$$I \leq \sum_{j=n/2}^{2n-1} \mu\left(A_{\frac{j}{2}}(B)\right) + \phi\left(\frac{j}{2}\right) \leq 2n\mu\left(\mathcal{C}_{\beta,\frac{n}{4}}(x)\right) + \frac{c_2}{n^{1+\kappa}} = o\left(\frac{1}{n}\right).$$

(II) For the second term we take $\omega = \frac{j}{2}$ and obtain

$$II \leq \sum_{j=2n}^{N(n)} \left(\mu(A_{j/2}(B)) + \phi\left(\frac{j}{2}\right)\right) \leq \sum_{j=2n}^{N(n)} \mu\left(\mathcal{C}_{\beta,\frac{j}{2}}(x)\right) + \frac{c_1}{n^{1+\kappa}}$$

since $\frac{j}{2} \geq n > N_0$. By (11) we conclude that $II = o(\frac{1}{n})$.

(III) For the third term III we take $\omega = \frac{N(n)}{2}$. Lemma 3 shows that $\mu(A_{N(n)/2}(B)) = \mathcal{O}(1)\mu(B)$. We obtain

$$\begin{aligned} III &\leq \sum_{j=N(n)+1}^{\Delta} \left(\mu(A_{N(n)/2}(B)) + \phi\left(j - \frac{N(n)}{2}\right)\right) \\ &\leq \sum_{j=N(n)+1}^{\Delta} \mathcal{O}(1)\mu(\tilde{B}) + \frac{c_2}{N(n)^{1+\kappa}} \\ &= \mathcal{O}(1)\Delta\mu(\tilde{B}) + \frac{c_2}{N(n)^{1+\kappa}} \\ &= o(\frac{1}{n}). \end{aligned}$$

The three estimates combined prove Theorem 5. □

To prove Theorems 1 and 2 we need to verify that (10) is satisfied.

Proof of Theorem 1. Under the hypothesis of Theorem 1 we take $\eta \in (\frac{1}{\xi\zeta}, 1)$ and put $N(n) = \mu(B_{\epsilon,n}(x))^{-\eta}$. This yields

$$\begin{aligned} \frac{n \cdot \psi(\epsilon, \gamma_{N(n)-k}, T^k x)}{\mu(B_{\epsilon,n}(x))} &\leq \frac{nC_\epsilon}{N(n)^{\xi\cdot\zeta}|\log\gamma|^\zeta \mu(B_{\epsilon,n}(x))} \\ &= C'_\epsilon n\mu(B_{\epsilon,n}(x))^{\eta\xi\zeta-1} \to 0 \end{aligned}$$

since $\eta\xi\zeta > 1$. Consequently $\vartheta_n(\epsilon) \to 0$ as $n \to \infty$ for every small enough $\epsilon > 0$ and the statement of Theorem 1 now follows from Theorem 5 □

Proof of Theorem 2. Similarly the choice of $\eta \in (\frac{1}{\alpha\xi}, 1)$ and $N(n) = \mu(B_{\epsilon,n}(x))^{-\eta}$ yield

$$\frac{n \cdot \psi(\epsilon, \gamma_{N(n)-k}, T^k x)}{\mu(B_{\epsilon,n}(x))} \leq \frac{nC_\epsilon N(n)^{-\alpha\xi}}{\mu(B_{\epsilon,n}(x))} = C'_\epsilon n\mu(B_{\epsilon,n}(x))^{\eta\alpha\xi-1} \to 0$$

since $\eta\alpha\xi > 1$. Again $\vartheta_n(\epsilon) \to 0$ as $n \to \infty$ for every small enough $\epsilon > 0$ and the theorem follows from Theorem 5. □

References

1. Abadi, M.: Exponential approximation for hitting times in mixing stochastic processes. Math. Phys. Electron. J. **7**(2), 1–19 (2001)
2. Abadim, M.: Poisson approximations via Chen-Stein for non-Markov processes. In: Sidoravicius, V., Vares, M.E. (eds.) In and Out of Equilibrium, vol. 2, pp. 1–19 (2008)
3. Abadi, M., Saussol, B.: Hitting and returning into rare events for all alpha-mixing processes. Stoch. Proc. Appl. **121**, 314–323 (2011)
4. Abadi, M., Vergne, N.: Poisson approximation for search of rare words in DNA sequences. ALEA-Lat. Am. J. Prob. Math. Stat. **4**, 233–244 (2007)

5. Abadi, M., Vergne, N.: Sharp errors for point-wise Poisson approximations in mixing processes. Nonlinearity **21**, 2871–2885 (2008)
6. Barbour, A.D., Chen, L.H.Y.: An Introduction to Stein's Method. Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore **4**, (2005)
7. Brin, M., Katok, A.: On local entropy. In: Geometric Dynamics. Lecture Notes in Mathematics, vol. 1007. Springer, Berlin
8. Chazottes, J.-R., Collet, P.: Poisson approximation for the number of visits to balls in nonuniformly hyperbolic dynamical systems. Ergod. Theory Dyn. Syst. **33**, 49–80 (2013)
9. Denker, M.: Remarks on weak limit laws for fractal sets. Progress in Probability **37**, 167–178 (1995)
10. Denker, M., Gordin, M., Sharova, A.: A Poisson limit theorem for toral automorphisms. Ill. J. Math. **48**(1), 1–20 (2004)
11. Doeblin, W.: Remarques sur la théorie métrique des fraction continues. Compos. Math. **7**, 353–371 (1940)
12. Haydn, N.T.A.: Entry and return times distribution. Dyn. Syst. Int. J. **28**(3), 333–353 (2013)
13. Haydn, N., Lacroix, Y., Vaienti, S.: Hitting and return times in ergodic dynamical systems. Ann. Probab. **33**, 2043–2050 (2005)
14. Haydn, N., Psiloyenis, Y.: Return times distribution for Markov towers with decay of correlations. Nonlinearity **27**, 1323–1349 (2014)
15. Haydn, N., Wasilewska, K.: Limiting distribution and error terms for the number of visits to balls in non-uniformly hyperbolic dynamical systems. Discret. Cont. Dyn. Syst. **36**(5), 2586–2611 (2016)
16. Haydn, N., Yang, F.: Entry times distribution for dynamical balls on metric spaces. http://arxiv.org/abs/1410.8640
17. Hirata, M.: Poisson law for the dynamical systems with the "self-mixing" conditions. In: Dynamical Systems and Chaos, vol. 1. Worlds Sci. Publishing, River Edge (1995)
18. Kifer, Y., Rapaport, A.: Poisson and compound Poisson approximations in a nonconventional setup; preprint 2012. http://arxiv.org/abs/1211.5238
19. Kupsa, M., Lacroix, Y.: Asymptotics for hitting times. Ann. of Probab. **33**(3), 610–614 (2005)
20. Lacroix, Y.: Possible limit laws for entrance times of an ergodic aperiodic dynamical system. Israel J. Math. **132**, 253–263 (2002)
21. Pène, F., Saussol, B.: Poisson law for some nonuniformly hyperbolic dynamical systems with polynomial rate of mixing. Université de Bretagne Occidentale (preprint)
22. Pitskel, B.: Poisson law for Markov chains. Ergod. Theory Dyn. Syst. **11**, 501–513 (1991)
23. Saussol, B., Troubetzkoy, S., Vaienti, S.: Recurrence, dimensions and Lyapunov exponents. J. Stat. Phys. **106**, 623–634 (2002)
24. Varandas, P.: Entropy and Poincaré recurrence from a geometrical viewpoint. Nonlinearity **22**(10), 2365–2375 (2009)
25. Walters, P.: An Introduction to Ergodic Theory. Springer, Berlin (1981)
26. Wang, H., Tang, M., Wang, R.: A Poisson limit theorem for a strongly ergodic non-homogeneous Markov chain. J. Math. Anal. Appl. **277**, 722–730 (2003)
27. Young, L.-S.: Statistical properties of dynamical systems with some hyperbolicity. Ann. Math. **7**, 585–650 (1998)
28. Young, L.-S.: Recurrence time and rate of mixing. Israel J. Math. **110**, 153–188 (1999)