Return times distribution for Markov towers with decay of correlations

2014 Nonlinearity 27 1323

(http://iopscience.iop.org/0951-7715/27/6/1323)

IOP Publishing | London Mathematical Society Nonlinearity

Nonlinearity **27** (2014) 1323–1349 doi:10.1088/0951-7715/27/6/1323

Return times distribution for Markov towers with decay of correlations

N T A Haydn and Y Psiloyenis

Mathematics Department, USC, Los Angeles, 90089-1113, USA

E-mail: nhaydn@math.usc.edu and psiloyen@usc.edu

Received 26 May 2013, revised 3 April 2014
Accepted for publication 3 April 2014
Published 16 May 2014

Recommended by C Liverani

Abstract

In this paper we prove two results. First we show that dynamical systems with a ϕ-mixing measure have in the limit Poisson distributed return times almost everywhere. We use the Chen–Stein method to also obtain rates of convergence. Our theorem improves on previous results by allowing for infinite partitions and dropping the requirement that the invariant measure has finite entropy with respect to the given partition. As has been shown elsewhere, the limiting distribution at periodic points is not Poissonian (but compound Poissonian). Here we show that for all non-periodic points the return times are in the limit Poisson distributed. In the second part we prove that Lai-Sang Young's Markov towers have Poisson distributed return times if the correlations decay at least polynomially with a power larger than 1.

Keywords: return times statistics, mixing measures, decay of correlations
Mathematics Subject Classification: 37A25, 37A50

1. Introduction

Beginning with the Poincaré recurrence theorem, one of the main interests in studying deterministic dynamical systems has been to show that the orbit of a typical point is on large timescales statistically regularly distributed and orbit segments that are sufficiently separated are close to independently distributed. In this paper we follow in this tradition and show that for invariant measures that are ϕ-mixing with respect to a possibly countably infinite partition the return times are in the limit Poisson distributed.

Interest in such questions go back to the 1940s when Doeblin [15] studied the Gauss map and its invariant measure. Later, in the 1970s Harris studied return times for Markov processes and then around 1990 the interest of the return times statistics became a central topic in dynamics. Using symbolic dynamics, Pitskel [32] proved that for Axiom A maps the return times are in the limit Poisson distributed with respect to equilibrium states for Hölder continuous potentials. Hirata [23] has a similar result using the Laplace transform which he then generalized later in [24]. Galves and Schmitt [17] then came up with a technique to get results for the first entry or return time which they applied to ψ-mixing systems and where they also for the first time provided error estimates. This method was then considerably extended by Abadi [2–4] to ϕ-mixing systems. Using a combinatorial argument improved error estimates were given in [6] for the first entry and return times of ϕ-mixing processes. For α-mixing systems, the limiting entry and return time distribution was established in [5]. A combinatorial argument was used in [7, 8] to show that the limiting distribution is Poissonian for ϕ-mixing measures if one takes the limit along a nested sequence of cylinders. In [30] multiple return times were shown to be Poisson distributed for a class of intermittent systems. Recently Kifer has proven limiting results for simultaneous returns to cylinder sets, first [26] an almost sure result using the Chen–Stein method and then [27] a complete classification with error terms. Let us note that in [14] the Chen–Stein method was used to get the Poisson limiting distribution for toral automorphisms where the limit is taken along sequences of ball-like sets.

Typically when entry times are Poisson distributed then so are the return times. In fact, for arbitrary return or entry time distribution there is a formula [20] that allows one to translate the entry time distribution into the return time distribution and vice versa.

For attractors on manifolds (with 1D unstable direction) which have a representation by Young towers with exponentially decaying correlations, Chazottes and Collet [13] have shown that the entry times are Poisson distributed for the SRB measure. Here the return sets are balls although the technique involves approximations by unions of cylinder sets. Wasilewska [35] extended this result to quite arbitrary measures on Young towers with polynomially decaying correlations. There, also, the return sets are balls B_ρ which are approximated by unions of cylinders. There the error terms decay with a negative power of $|\log\rho|$. In particular, for attractors this result applies to SRB measures with polynomially decaying correlations. See also [22]. For an overview of distribution results of return times also see [19].

In this paper we consider maps that are ϕ-mixing with respect to an invariant measure and a partition which can be finite or countably infinite. The purpose of the paper is three fold: (i) we develop a more direct approach to the method of Chen–Stein to obtain distribution results on return times, (ii) the Poisson law we obtain is applicable to unions of cylinders rather than single cylinder neighbourhoods, and (iii) we allow for infinite partitions and do not require the entropy to be finite. Unlike the moment method which requires the measure to have the stronger ψ-mixing property, the method of Chen–Stein requires us to only look at 'two fold' mixing sets and this is what makes it accessible to ϕ-mixing measures. We also obtain rates of convergence. Since we show the limiting distribution for unions of cylinders whose total measures are required to decay at some rate, this approach can be used to obtain limiting distribution results for metric balls in a metric space setting (theorem 3). Naturally we have to keep away from return sets that 'look' periodic. At periodic points the limiting distribution cannot be Poisson but is, as was shown in [21], compound Poisson distributed. In corollary 1 we deduce that at all non-periodic points return times are in the limit Poissonian.

In section 2 we set up the Chen–Stein method and then prove the main technical result proposition 2. A similar method is used to prove theorem 5. Most of the results of sections 2 and 3 (in particular theorem 1 and lemma 2) also appeared in [33].

In the second part (section 4) of the paper we then look at Young towers and show that return and entry times are in the limit Poisson distributed although we do not necessarily have the ϕ-mixing property for those systems. Since the invariant measure on a Young tower typically is not ϕ-mixing (although it is α-mixing), more delicate estimates are required in order to obtain the limiting Poisson distribution along sequences of sets which are unions of cylinders. This allows us to show that if the correlations decay at least polynomially with a (possibly fractional) power larger than 1, then the limiting return times distributions are Poissonian provided a short return time condition is satisfied. For the precise statements see section 4 and in particular theorem 4.

Let us note that it is crucial to select the return set to be some 'regular' set-like cylinders as Kupsa and Lacroix [28, 29] have shown that any limiting distribution can be realized if one choses the return sets appropriately. Also let us note that Kupsa has constructed an example of a symbolic system over three elements which has positive entropy and whose first entry time is not exponentially (with parameter one) distributed almost everywhere. This emphasizes that despite the plethora of existing results on the distribution of entry times, we cannot expect positive entropy systems to generically have Poisson distributed returns in the limit.

2. Distribution for ϕ-mixing systems

Let T be a map on Ω and μ a T-invariant probability measure on Ω. Let $\mathcal{A}$ be a finite or countably infinite measurable partition on Ω. We put $\mathcal{A}^n$ for its nth join $\bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$. We assume that the partition $\mathcal{A}$ is generating (i.e. the atoms of $\mathcal{A}^\infty$ consist of single points).

Throughout the paper we will assume that μ is (right) *ϕ-mixing*, that is there exists a decreasing sequence $\phi(k) \to 0$ (as $k \to \infty$) so that

$$\left| \frac{\mu(A \cap T^{-n-k}(B))}{\mu(B)} - \mu(A) \right| \leqslant \phi(k)$$

for all $A \in \mathcal{A}^n$, $B \in \sigma(\bigcup_{\ell\geqslant 1} \mathcal{A}^\ell)$ ($\mu(B) > 0$) and for all n, k (see, e.g., [16]). Let us note that there exists $\Lambda > 0$ so that for any $n \in \mathbb{N}$ and $A \in \mathcal{A}^n$ one has $\mu(A) \leqslant K\mathrm{e}^{-\Lambda n}$ for some constant K. For a proof of this fact see Abadi [2] whose proof for finite alphabet carries over to infinite alphabet without any change.

For a set $A \subset \Omega$ the *hitting time* $\tau_A : \Omega \to \mathbb{N} \cup \{\infty\}$ is a random variable defined on the entire set Ω as follows:

$$\tau_A(x) = \inf\left\{k \geqslant 1 \colon T^k(x) \in A\right\}$$

($\tau_A(x) = \infty$ if $T^k x \notin A$ $\forall k \in \mathbb{N}$). If we narrow down the domain of τ_A to the set A then τ_A is called the *return time* or first-return time. According to Kac's theorem [25] $\int_A \tau_A \, \mathrm{d}\mu = 1$ for any ergodic T-invariant probability measure μ and measurable $A \subset \Omega$ with positive measure. We then can define the induced map $\hat{T}_A : A \circlearrowleft$ given by $\hat{T}_A(x) = T^{\tau_A(x)}(x)$ $\forall x \in A$, and the kth return time τ_A^k by putting $\tau_A^1 = \tau_A$ $(k = 1)$ and recursively for $k > 1$

$$\tau_A^k(x) = \inf\left\{\ell > \tau_A^{k-1}(x) \colon T^\ell(x) \in A\right\} = \tau_A(\hat{T}_A^{k-1}(x)) + \tau_A^{k-1}(x)$$

(for convenience we put $\tau_A^0 = 0$). Following [8] the *period* of $A \subset \Omega$, under the map T, is defined to be

$$r_A = \inf\{n \in \mathbb{N} | A \cap T^{-n}(A) \neq \emptyset\},$$

or, equivalently, $r_A = \inf_{x\in A} \tau_A(x)$. From the mixing property we conclude that $r_A \leqslant \min\{\ell : \phi(\ell) < 1\}$.

For $A \in \sigma(\mathcal{A}^n)$ (union of n-cylinders) let us define

$$\delta_A(j) = \min_{1\leqslant w\leqslant j\wedge n} \{\mu(A_w(A)) + \phi(j - w)\},$$

where $A_w(A) \in \sigma(\mathcal{A}^w)$ is smallest so that $A \subset A_w(A)$, that is $A_w(A) = \bigcup_{B\in\mathcal{A}^w \colon B\cap A\neq\varnothing} B$.

Remark. In a similar way one can define a measure μ to be *left ϕ-mixing*[1] if

$$\left|\frac{\mu(A\cap T^{-n-k}(B))}{\mu(A)}-\mu(B)\right|\leqslant\phi(k)$$

for all $A\in\mathcal{A}^n$, $B\in\sigma(\bigcup_j\mathcal{A}^j)$ and n,k. A right ϕ-mixing measure is not necessarily also left ϕ-mixing. However, the results in this paper on the distribution of return times (theorems 1 and 2 and corollary 1 and also lemma 1) also apply to left ϕ-mixing systems since the techniques involved are symmetric. If the measure is left ϕ-mixing then $\delta_A(j)$ has to be replaced by

$$\hat{\delta}_A(j)=\min_{1\leqslant w\leqslant j\wedge n}\left\{\mu(A^{(w)}(A))+\phi(j-w)\right\},$$

where $A^{(w)}(A)=T^{-(n-w)}T^{n-w}A\in\sigma(T^{-(n-w)}\mathcal{A}^w)$ is the smallest element in $\sigma(T^{-(n-w)}\mathcal{A}^w)$ which contains A ($w\leqslant n$).

Theorem 1. *Let μ be a T-invariant probability measure which is ϕ-mixing with respect to a generating and at most countably infinite partition $\mathcal{A}$. Then there exists a constant C_1 so that*

$$\left|\mathbb{P}\left(\tau_A^k>\frac{t}{\mu(A)}\right)-\sum_{i=0}^{k-1}\mathrm{e}^{-t}\frac{t^i}{i!}\right|\leqslant C_1t(t\vee 1)\inf_{\Delta>0}\left(\Delta\mu(A)+\sum_{j=r_A}^{\Delta}\delta_A(j)+\frac{\phi(\Delta)}{\mu(A)}\right)|\log\mu(A)|.$$

for all $k,n\in\mathbb{N}$ and $A\in\sigma(\mathcal{A}^n)$.

Theorem 2 ([33]). *Let μ be a ϕ-mixing T-invariant probability measure with respect to the generating and at most countable infinite partition $\mathcal{A}$. Let $\eta\geqslant 1$ be so that $n^\eta\phi(n)\to 0$ as $n\to\infty$. Let $K>0$. Then for $A\in\sigma(\mathcal{A}^n)$ a finite or infinite union of n-cylinders such that $|\log\mu(A)|\leqslant Kn^\eta$ and $r_A>\frac{n}{2}$ the following applies:*

(i) Exponential mixing rate. Suppose $\phi(n)=\mathcal{O}(\vartheta^n)$, with $0<\vartheta<1$ and $\mu(A_w(A))=\mathcal{O}(\vartheta^w)$ for $w\leqslant n$. Then there exists $\gamma=\gamma(\vartheta)>0$ and $C_2>0$ such that

$$\left|\mathbb{P}\left(\tau_A^k>\frac{t}{\mu(A)}\right)-\sum_{i=0}^{k-1}\mathrm{e}^{-t}\frac{t^i}{i!}\right|\leqslant C_2t(t\vee 1)\mathrm{e}^{-\gamma n},\qquad\forall t>0\quad\text{and }\forall n\in\mathbb{N}.\tag{1}$$

(ii) Polynomial mixing rate. Suppose $\phi(n)=\mathcal{O}(n^{-\beta})$ with $\beta>1+\eta$ and $\mu(A_w(A))=\mathcal{O}(w^{-\beta})$ for $w\leqslant n$. Then there exists $C_2>0$ such that

$$\left|\mathbb{P}\left(\tau_A^k>\frac{t}{\mu(A)}\right)-\sum_{i=0}^{k-1}\mathrm{e}^{-t}\frac{t^i}{i!}\right|\leqslant C_2t(t\vee 1)\frac{1}{n^{\beta-1-\eta}},\qquad\forall t>0\quad\text{and }\forall n\in\mathbb{N}.\tag{2}$$

Remarks.

(I) The statements of these two theorems also apply to left ϕ-mixing measures. In this case, however, the quantity $\delta_A(j)$ in theorem 1 has to be replaced by $\hat{\delta}_A(j)$ and in theorem 2 the decay rate for $\mu(A_w(A))$ has to apply to $\mu(A^{(w)}(A))$ instead. Here we present the proof in the case when μ is right ϕ-mixing.

(II) The assumption of theorem 2 that the period r_A be greater than $\frac{n}{2}$ can be substituted with any other number of the order of n. This assumption is in place to ensure that the reference cylinder A does not exhibit a periodic behaviour. By its very definition, the set A consists of points that travel together for at least n iterates of the map F. In view of this property if the set A revisited itself very early on by the means of a single point x that would have caused an

[1] This is sometimes also called *reversed ϕ-mixing*.

entire neighbourhood of A to fall into A at that same iterate. Considering the extreme case, if the entire set falls into A at the same iterate of F that renders A periodic. In this case the set A would act like a 'trap'. By asking that more time passes by before any of A's points comes back to A we ensure that the system is nearer to the time where the set will start spreading all over the space, by virtue of the mixing properties that govern the dynamics. In particular, for cylinders around periodic points the limiting distribution of return times is a compound Poissonian distribution [21].

(III) Commenting on the assumption that $|\log\mu(A_n)| \leqslant Kn^\eta$ recall that in the finite entropy case, when $H(\mathcal{A}) < \infty$, the theorem of Shannon–MacMillan–Breiman [31] implies that for a.e. point $x \in \Omega$ there exists $C > 0$ such that

$$|\log\mu(A_n(x))| \leqslant Cn \qquad \forall n \in \mathbb{N}, \tag{3}$$

i.e. $\eta = 1$, where we denote by $A_n(x)$ the n-cylinder centred at x. On the other hand, if $H(\mathcal{A}) < \infty$ and $\eta > 1$ then we can give a rough estimate on the set of cylinders that do not satisfy the condition $|\log\mu(A)| \leqslant Kn^\eta$. Denote by $B(n) \subset \mathcal{A}^n$ the set of all the n-cylinders A that satisfy $|\log\mu(A)| > Kn^\eta$. Then, since $H(\mathcal{A}^n) = \sum_{A\in\mathcal{A}^n}\mu(A)|\log\mu(A)| \leqslant nH(\mathcal{A})$, we obtain

$$nH(\mathcal{A}) \geqslant \sum_{A_n\in B(n)} \mu(A_n)|\log\mu(A_n)| \geqslant \sum_{A_n\in B(n)} Kn^\eta\mu(A_n) = Kn^\eta\mu(B(n))$$

which implies

$$\mu(B(n)) \leqslant \frac{H(\mathcal{A})}{Kn^{\eta-1}} \leqslant \frac{c}{n^{\eta-1}}.$$

This shows that for $\eta > 1$ as n increases the exception set, or 'bad' set, gets smaller. The bigger the η we choose the bigger coverage we achieve, where the estimates hold, but making η larger that has a direct effect on the error estimates. As pointed out above, Abadi's result does not allow us to choose η to be less than 1.

In the rest of this section we will look at the return time distribution for cylinder sets. Let $x \in \Omega$ and denote by $\pi_n = r_{A_n(x)}$ the period of the n-cylinder neighbourhood $A_n(x) \in \mathcal{A}^n$. Since $A_{n+1}(x) \cap T^jA_{n+1}(x) \subset A_n(x) \cap T^jA_n(x)$ $\forall n, j$, one sees that π_n is an increasing sequence which implies that either $\pi_n \to \infty$ or π_n converges to a limit π_∞ (which is a function of x).

In the finite case, $\pi_\infty < \infty$, the point x is a periodic point with period π_∞. This follows from the fact that $x \in A_n(x) \cap T^{\pi_\infty}A_n(x)$ for all n large enough. Since $\mathcal{A}$ is generating, the periodicity of x follows from taking a limit $n \to \infty$ as $\{x\} = \bigcap_n A_n(x)$. For ψ-mixing measures it was shown in [21] that the limiting distribution of $\mathbb{P}(\tau_A^k > \frac{t}{\mu(A)})$ converges to the Pólya–Aeppli compound Poisson distribution. For the limiting first return-time distribution at a periodic point a complete description for ϕ-mixing measures was given in [8] where it was shown that the density has a point mass at $t = 0$ of weight $\lim_{n\to\infty}\mathbb{P}_{A_n(x)}(\tau_{A_n(x)} = \pi_\infty)$ and is exponential otherwise. This generalizes a result of Pitskel [32] for equilibrium states on axiom A systems.

In the infinite case, when $\pi_n \to \infty$ as $n \to \infty$, x is non-periodic and we can estimate δ_A as follows:

$$\delta_{A_n(x)}(j) = \inf_{0\leqslant k\leqslant j\wedge n}\{\mu(A_k(x)) + \phi(j-k)\} \leqslant K\mathrm{e}^{-\Lambda(j\wedge n)/2} + \phi(j/2)$$

($k = j/2$), where we used the property that $\mu(A_k(x)) \leqslant K\mathrm{e}^{-\Lambda k}$ ($\Lambda > 0$). Hence, with some c_1,

$$\mathcal{E}_n(\Delta) = \sum_{j=\pi_n}^{\Delta} \delta_{A_n(x)}(j) \leqslant c_1\mathrm{e}^{-\Lambda(\pi_n\wedge n)/2} + \sum_{j=\pi_\infty}^{\infty}\phi(j/2) \longrightarrow 0$$

as $n \to \infty$ if we assume that $\phi(j)$ is summable. Also note that if ϕ is summable then we get that $\lim_{j\to\infty} j\phi(j) = 0$. Hence there exist a sequence Δ_n, $n = 1, 2, \ldots$, so that $\phi(\Delta_n)/\mu(A_n(x)) \to 0$ and also $\Delta_n \mu(A_n(x)) \to 0$ as $n \to \infty$.

As a consequence of theorem 1 we thus have the following result.

Corollary 1. *Let μ be a ϕ-mixing w.r.t. the generating partition $\mathcal{A}$ that is at most countably infinite. Assume $\phi(j)$ is summable. If $x \in \Omega$ is not periodic, then*

$$\mathbb{P}\left(\tau^k_{A_n(x)} > \frac{t}{\mu(A_n(x))}\right) \longrightarrow \sum_{i=0}^{k-1} \mathrm{e}^{-t}\frac{t^i}{i!}$$

as $n \to \infty$ for all $t > 0$.

This is sometimes expressed using the counting function $\zeta^t_A = \sum_{j=0}^m \chi_A \circ T^j$, where $m = [t/\mu(A)]$ and χ_A is the characteristic function of A. Then $\mathbb{P}(\tau^k_A > t/\mu(A)) = \sum_{i=0}^{k-1} \mathbb{P}(\zeta^t_A = i)$ and the statement of the corollary reads

$$\mathbb{P}\left(\zeta^t_{A_n(x)} = k\right) \longrightarrow \mathrm{e}^{-t}\frac{t^k}{k!}$$

as $n \to \infty$ for all non-periodic $x \in \Omega$ and all $t > 0$. As remarked earlier, this result equally applies to left ϕ-mixing measures.

2.1. Application

As an application of theorem 2 we will indicate how one can obtain the limiting distribution for metric balls for maps on metric spaces. We will still require that there be a generating partition with respect to which the measure is ϕ-mixing. The balls will then be approximated by unions of cylinders. This approach was also used by Pitskel [32] for toral automorphisms on $\mathbb{T}^2$ and in [18] for rational maps.

Let T be a map on a metric space Ω and let $\mathcal{A} = \{A_j : j\}$ a generating finite or countable infinite partition of Ω, that is $\Omega = \bigcup_j A_j$ and $A_j \cap A_i = \varnothing$ for $i \neq j$. As before we denote by $\mathcal{A}^n$ the nth joint of the partition. Assume there is a T-invariant probability measure μ on Ω. Then we put for parameters $t > 0$ and radii $\rho > 0$

$$\zeta^t_{B_\rho(x)} = \sum_{j=0}^{m} \chi_{B_\rho(x)} \circ T^j$$

for the counting function of the returns to the metric ball $B_\rho(x)$ in the space Ω, where $m = [t/\mu(B_\rho(x)]$.

Theorem 3. *Let μ be an invariant measure on the metric space Ω and suppose there is a partition (finite or countably infinite) $\mathcal{A}$. Let $x \in \Omega$ and assume the following conditions are satisfied.*

(i) μ is ϕ-mixing with rate $\phi(k)$ decaying at least polynomially with power larger than 2;
(ii) diam($\mathcal{A}^n$) decays exponentially fast as $n \to \infty$;
(iii) There exists $w > 1$ such that $\frac{\mu(B_{\rho+\rho^w}(x))}{\mu(B_\rho(x))} \longrightarrow 1$ as $\rho \to 0^+$ almost everywhere;
(iv) μ has finite and positive dimension almost everywhere;
(v) $r_{B_\rho(x)} \geqslant$ const.$|\log\rho|$ for small enough ρ.

Then

$$\mathbb{P}\left(\zeta^t_{B_\rho(x)} = k\right) \longrightarrow \mathrm{e}^{-t}\frac{t^k}{k!}$$

as $\rho \to 0^+$ for almost every $x \in \Omega$ and $k \in \mathbb{N}_0$.

Proof. We approximate the balls $B_\rho(x)$ by unions of cylinders. By assumption (ii) there exists a $v \in (0,1)$ such that $\mathrm{diam}(\mathcal{A}) \leqslant v^n$ (for n large enough). Let $n = [w\frac{\log\rho}{\log v}] + 1$, fix x and denote by

$$C^t_{\rho,n} = \bigcup_{A\in\mathcal{A}^n : A\cap B_\rho(x)\neq\varnothing} A$$

the smallest union of n-cylinders that contains $B_\rho(x)$. By assumption (iv) we have $|\log\mu(B_\rho(x))| \leqslant c_1|\log\rho|$ for some constant $c_1 < \infty$ and consequently the sets $C^t_{\rho,n} \in \sigma(\mathcal{A}^n)$ satisfy the assumption of theorem 2 for $\eta = 1$. By assumption (v) we have $r_{B_\rho(x)} \geqslant \mathrm{const}.n$ thus satisfying the short return times condition. Hence we obtain by theorem 2 that $\mathbb{P}(\zeta^t_{C_{\rho,n}} = k) \longrightarrow \mathrm{e}^{-t}\frac{t^k}{k!}$ as $\rho \to 0$ (and $n \to \infty$).

By assumption (iii) on the regularity of the measure μ we have

$$\left|\mathbb{P}\left(\zeta_{B_{\rho+\rho^w}} = k\right) - \mathbb{P}\left(\zeta_{B_\rho} = k\right)\right| \leqslant \left[\frac{t}{\mu(B_\rho(x))}\right]\mu(B_{\rho+\rho^w}\setminus B_\rho) \longrightarrow 0$$

as $\rho \to 0$. Since $B_\rho(x) \subset C^t_{\rho,n} \subset B_{\rho+v^n}(x) \subset B_{\rho+\rho^w}$ (as $v^n < \rho^w$) we obtain $\mathbb{P}(\zeta^t_{B_\rho(x)} = k) \longrightarrow \mathrm{e}^{-t}\frac{t^k}{k!}$ ∎

Remarks.

(I) The requirement (i) that μ is ϕ-mixing appears somewhat artificial, but it can occur in the following simple way: an Anosov map T on a manifold Ω admits the construction of an arbitrarily fine Markov partition $\mathcal{A}$ which then can be used to model the dynamics of T by the shift transform σ a subshift of finite type Σ. The projection $\pi : \Sigma \to \Omega$ semiconjugates the shift transform $\sigma : \Sigma \circlearrowleft$ to the map $T : \Omega \circlearrowleft$; that is $\pi\circ\sigma = T\circ\pi$. A ϕ-mixing measure ν on Σ then maps to a ϕ-mixing measure $\mu = \pi^*\nu$ on Ω. Theorem 3 then implies that the limiting return times distribution for metric balls is Poissonian (provided conditions (iii)–(v) are met).

(II) If Ω is a manifold and μ is an absolutely continuous measure then the regularity condition (iii) $\frac{\mu(B_{\rho+\rho^w}(x))}{\mu(B_\rho(x))} \longrightarrow 1$ as $\rho \to 0^+$ is satisfied everywhere for any $w > 1$.

(III) Condition (v) on the short returns is satisfied for many measures. For instance in [13], lemma 4.1, it was shown that for the SRB measure on codimension one attractors with exponentially decaying tails there exists an $\mathfrak{a} > 0$ so that the measure of the set of very short returns

$$\mathcal{V}_\rho = \{x \in \Omega : r_{B_\rho(x)} > \mathfrak{a}|\log\rho|\}$$

is bounded by $\mu(\mathcal{V}_\rho) = \mathcal{O}(\rho^a)$ for some $a > 0$. Although the proof uses Young towers it does not rely on the decay of correlations or a mixing property. This was in [22, 35] extended to invariant measures for more general maps that allow for a Young tower construction with polynomially decaying tails where one gets the estimate $\mu(\mathcal{V}_\rho) = \mathcal{O}(|\log\rho|^{-a})$ for some $a > 0$. In both cases every point $x \notin \mathcal{V}_\rho$ satisfies condition (v).

(IV) The theorem cannot in general directly be applied to systems that are modelled by a Young tower since the invariant measure is only α-mixing and not necessarily ϕ-mixing (see equation (18)). A more elaborate method will be used to exploit the $\mathscr{L}^1$ convergence of the densities (see theorem 5).

3. Proof of theorem 1

3.1. Short returns

Abadi has shown that for ϕ-mixing systems the measure of cylinder sets decay exponentially, i.e. there are strictly positive constants K and Λ such that $\mu(A) \leqslant K\mathrm{e}^{-\Lambda n}$ for any integer $n \in \mathbb{N}$ and any n-cylinder A. Recall that $\delta_A(k) = \min_{1\leqslant w<k}\{\mu(A_w(A)) + \phi(k-w)\}$ where $A_w(A) \in \sigma(\mathcal{A}^w)$ is smallest so that $A \subset A_w(A)$.

Recall that the period r_A of the set A is defined as the smallest j for which $A \cap T^{-j}(A) \neq \emptyset\}$.

Lemma 1. $\mathbb{P}_A(\tau_A \leqslant t) \leqslant \sum_{j=r_A}^{t} \delta_A(j)$.

Proof. For numbers $w_j \leqslant j$ we have

$$\begin{aligned}\mu\left(A \cap \{\tau_A \leqslant t\}\right) &= \sum_{j=r_A}^{t} \mu\left(A \cap \{\tau_A = j\}\right)\\ &\leqslant \sum_{j=r_A}^{t} \mu\left(A_{w_j}(A) \cap T^{-(n-j)}A\right)\\ &\leqslant \sum_{j=r_A}^{t} \mu(A)\delta_A(j)\end{aligned}$$

using the right ϕ-mixing property and optimizing for w_j. The result then follows. □

In the same way one proves that $\mathbb{P}_A(\tau_A \leqslant t) \leqslant \sum_{j=r_A}^{t} \hat{\delta}_A(j)$ if μ is left ϕ-mixing since then $\mu(A \cap \{\tau_A = j\}) \leqslant \mu(A \cap T^{-(n-j)}A^{(w)}(A)) \leqslant \hat{\delta}_A(j)$ for the optimal choice of $w \in [1, n]$.

3.2. The Stein method

Stein's method of proving limiting theorems was first introduced by Stein [34] for the central limit theorem and then subsequently developed for the Poisson distribution [10, 11]. As mentioned before this method has been used in dynamics several times: Abadi [4] used it by way of a result in [9] to obtain the Poisson distribution for cylinder sets in ϕ-mixing systems. Denker *et al* [14] used the Chen–Stein method to obtain the Poisson distribution for limiting return times to ball-like sets for torus maps. Their approach involved extensive use of harmonic analysis. Here we develop a more practical approach that does not use [9] and does not require the target set to be a single cylinder, but could possibly be an infinite union of cylinders. Also, since entropy does not play any role, this approach works for infinite entropy systems and infinite alphabets. In the following we give a short description of the method as it is relevant for our purpose.

Let μ be a probability measure on $\mathbb{N}_0$ which is equipped with the power σ-algebra $\mathcal{B}_{\mathbb{N}_0}$. Additionally we denote by μ_0 the Poisson-distribution measure with mean t, i.e. $\mathbb{P}_{\mu_0}(\{k\}) = \frac{\mathrm{e}^{-t}t^k}{k!}$ $\forall k \in \mathbb{N}_0$. Also let $\mathcal{F}$ be the set of all real-valued functions on $\mathbb{N}_0$. The Stein operator $\mathcal{S} : \mathcal{F} \to \mathcal{F}$ is defined by

$$\mathcal{S}f(k) = tf(k+1) - kf(k), \qquad \forall k \in \mathbb{N}_0. \tag{4}$$

The Stein equation

$$\mathcal{S}f = h - \int_{\mathbb{N}_0} h \,\mathrm{d}\mu_0 \tag{5}$$

for the Stein operator in (4) has a solution f for each μ_0-integrable $h \in \mathcal{F}$ (see [10]). The solution f is unique except for $f(0)$, which can be chosen arbitrarily. Moreover, f can be computed recursively from the Stein equation, namely [10]

$$f(k) = \frac{(k-1)!}{t^k} \sum_{i=0}^{k-1} (h(i) - \mu_0(h)) \frac{t^i}{i!} \tag{6}$$

$$= -\frac{(k-1)!}{t^k} \sum_{i=k}^{\infty} (h(i) - \mu_0(h)) \frac{t^i}{i!}, \qquad \forall k \in \mathbb{N}. \tag{7}$$

In particular, if $h : \mathbb{N}_0 \to \mathbb{R}$ is bounded then so is the associated Stein solution f.

Proposition 1 ([10]). *A probability measure μ on $(\mathbb{N}_0, \mathcal{B}_{\mathbb{N}_0})$ is Poisson (with parameter t) if and only if*

$$\int_{\mathbb{N}_0} \mathcal{S} f \, \mathrm{d}\mu = 0 \quad \text{for all bounded functions } f : \mathbb{N}_0 \to \mathbb{R}.$$

A probability measure μ on $(\mathbb{N}_0, \mathcal{B}_{\mathbb{N}_0})$ which approximates the Poisson distribution μ_0 can be estimated as follows:

$$|\mu(E) - \mu_0(E)| = \left| \int_{\mathbb{N}_0} \mathcal{S} f \, \mathrm{d}\mu \right| = \left| \int_{\mathbb{N}_0} (tf(k+1) - kf(k)) \, \mathrm{d}\mu \right|, \tag{8}$$

where $E \subset \mathbb{N}_0$ and f is the Stein solution that corresponds to the indicator function χ_E. Sharp bounds for the quantity on the right-hand side of (8) is what one is after when the Stein method is used for Poisson approximation.

Lemma 2. *For the Poisson distribution μ_0, the Stein solution of the Stein equation (5) that corresponds to the indicator function $h = \chi_E$, with $E \subset \mathbb{N}_0$, satisfies*

$$\left| f_{\chi_E}(k) \right| \leqslant \begin{cases} 1 & \text{if } k \leqslant t \\ \frac{2+t}{k} & \text{if } k > t. \end{cases} \tag{9}$$

In particular,

$$\sum_{i=1}^{m} \left| f_{\chi_E}(i) \right| \leqslant \begin{cases} m & \text{if } m \leqslant t \\ t + (2+t) \log \frac{m}{t} & \text{if } m > t. \end{cases} \tag{10}$$

Proof.

We consider the two cases: (i) $k > t$ and (ii) $k \leqslant t$.

(i) $k > t$: for $h = \chi_E$, from the representation (7) for the Stein solution we have

$$f_{\chi_E}(k) = -\frac{(k-1)!}{t^k} \sum_{i=k}^{\infty} (h(i) - \mu_0(h)) \frac{t^i}{i!}.$$

Therefore,

$$\begin{aligned} \left| f_{\chi_E}(k) \right| &\leqslant \frac{(k-1)!}{t^k} \sum_{i=k}^{\infty} |h(i) - \mu_0(h)| \frac{t^i}{i!} \\ &\leqslant \frac{(k-1)!}{t^k} \sum_{i=k}^{\infty} \frac{t^i}{i!} \\ &= \frac{(k-1)!}{t^k} \frac{t^k}{k!} \left(1 + \sum_{i=1}^{\infty} \frac{t}{k+1} \frac{t}{k+2} \cdots \frac{t}{k+i} \right). \end{aligned} \tag{11}$$

If $i > t$ then each term in the infinite sum in (11) is no greater than $(\frac{1}{2})^{i-t}$. If $i \leqslant t$, all terms in the sum in (11) are clearly no greater than 1. Hence

$$\left|f_{\chi_E}(k)\right| \leqslant \frac{(k-1)!}{t^k}\frac{t^k}{k!}\left(1+t+\sum_{i=1}^{\infty}\left(\frac{1}{2}\right)^i\right) = \frac{2+t}{k}.$$

(ii) $k \leqslant t$: using the alternative representation (6) for the Stein solution f_{χ_E}, this time, we obtain

$$|f_{\chi_E}(k)| \leqslant \frac{(k-1)!}{t^k}\sum_{i=0}^{k-1}|h(i)-\mu_0(h)|\frac{t^i}{i!} \leqslant \frac{(k-1)!}{t^k}\sum_{i=0}^{k-1}\frac{t^i}{i!} \leqslant \frac{(k-1)!}{t^k}\frac{t^{k-1}}{(k-1)!}k \leqslant 1$$

as the sequence $\{\frac{t^j}{j!}\}_{j\in\mathbb{N}}$ is increasing for $j \leqslant t$ and decreasing for $j > t$. This completes the proof of inequality (9). The second statement is now obvious for $m \leqslant t$. On the other hand if $m > t$ then it follows from the inequality $\sum_{i=t+1}^{m}\frac{1}{i} \leqslant \log\frac{m}{t}$.

□

3.3. Return time distribution

Now we want to approximate the function $\mathbb{P}(\tau_A^k \leqslant m)$ for all $k \geqslant 1$ and all $m \in \mathbb{R}^+$. Let $A \in \sigma(\mathcal{A}^n)$ and denote by $W_m(x)$ the number of visits of the orbit $\{T(x), T^2(x), \ldots, T^{[m]}(x)\}$ to the set A, i.e.

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

where χ_A is the characteristic function of the set A that is $\chi_A(x) = 1$ if $x \in A$ and $\chi_A(x) = 0$ otherwise (and $[m]$ is the integer part of m). Then

$$\mathbb{P}(\tau_A^k \leqslant [m]) = 1 - \mathbb{P}(\tau_A^k > [m]) = 1 - \mathbb{P}(W_{[m]} < k).$$

Therefore, our problem of approximating the distribution of τ_A^k becomes equivalent to approximating the distribution of W_m for all $m \in \mathbb{N}$. The Poisson parameter t is the expected value of W_m (i.e. $t = \mu(W_m)$). If we put $p_i = \mu(T^{-i}A) = \mu(A) \quad \forall i = 1, 2, \ldots$, then

$$t = \mu(W_m) = \sum_{i=1}^{m}\mu\left(\chi_A T^i\right) = \sum_{i=1}^{m}p_i = m\mu(A),$$

i.e. $m = [t/\mu(A)]$. If $h = \chi_E$ with E an arbitrary subset of the positive integers, $E \subset \mathbb{N}_0$, then we obtain from (8)

$$\left|\int_{\mathbb{N}_0}\mathcal{S}f\,\mathrm{d}\mu\right| = \left|\int_{\mathbb{N}_0}h\,\mathrm{d}\mu - \int_{\mathbb{N}_0}h\,\mathrm{d}\mu_0\right| = |\mathbb{P}(W_m \in E) - \mu_0(E)|$$

and in turn, since the Stein operator $\mathcal{S}$ for the Poisson distribution is given by (4), we obtain

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| = |E\,(tf(W_m+1) - W_m f(W_m))| \quad \forall E \subset \mathbb{N}_0.$$

Note that the difference $|\mathbb{P}(W_m \in E) - \mu_0(E)|$ above gives exactly the error of the Poisson approximation. We hence estimate

$$\begin{aligned}|\mathbb{P}(W_m \in E) - \mu_0(E)| &= \left|t\mathbb{E}f(W_m+1) - \mathbb{E}\left(\sum_{i=1}^{m}I_i f(W_m)\right)\right| \\ &= \left|\sum_{i=1}^{m}p_i\mathbb{E}f(W_m+1) - \sum_{i=1}^{m}p_i\mathbb{E}(f(W_m)|I_i = 1)\right|\end{aligned}$$

$$= \left| \sum_{i=1}^{m} p_i \left(\mathbb{E} f(W_m + 1) - \mathbb{E}(f(W_m)|I_i = 1) \right) \right|$$
$$= \sum_{i=1}^{m} p_i \left(\sum_{a=0}^{m} f(a+1)\mathbb{P}(W_m = a) - \sum_{a=0}^{m} f(a)\mathbb{P}(W_m = a|I_i = 1) \right)$$
$$= \sum_{i=1}^{m} p_i \sum_{a=0}^{m} f(a+1)\epsilon_{a,i}, \tag{12}$$

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

$$\epsilon_{a,i} = |\mathbb{P}(W_m = a) - \mathbb{P}(W_m = a + 1|I_i = 1)|\,. \tag{13}$$

The function f above is the solution of the Stein equation (5) that corresponds to the indicator function $h = \chi_E$ in the Stein method. In fact bounds on f have been obtained in corollary 2.

Now, in view of the new representation for $|\mathbb{P}(W_m \in E) - \mu_0(E)|$ we need to look at the term $\epsilon_{a,i}$ more closely. If we put $W_m^i = W_m - \chi_A \circ T^i$ then the mixing condition yields the following estimates on $\epsilon_{a,i}$:

$$\begin{aligned}
\epsilon_{a,i} &= |\mathbb{P}(W_m = a) - \mathbb{P}(W_m = a + 1|I_i = 1)| \\
&= \left| \mathbb{P}(W_m = a) - \frac{\mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right)}{\mu(A)} \right| \\
&= \left| \mathbb{P}(W_m = a) - \frac{\mathbb{P}(W_m^i = a)\mu(A) + \epsilon'_{a,i}}{\mu(A)} \right| \\
&\leqslant \left| \mathbb{P}(W_m = a) - \mathbb{P}(W_m^i = a) \right| + \frac{\xi_a}{\mu(A)},
\end{aligned}$$

where $\epsilon'_{a,i} = \mathbb{P}(\{W_m^i = a\} \cap T^{-i}A) - \mathbb{P}(W_m^i = a)\mu(A)$ ($\epsilon'_{a,i} = 0$ if all I_j are independent) and $\xi_a = \max_i |\mathbb{P}(\{W_m^i = a\} \cap T^{-i}A) - \mathbb{P}(W_m^i = a)\mu(A)|$. The bound on $\epsilon_{a,i}$ has two terms, the first of which is

$$\left| \mathbb{P}(W_m = a) - \mathbb{P}(W_m^i = a) \right| \leqslant \mathbb{P}(I_i = 1) = \mu(A).$$

The second term, which contains ξ_a, is the error due to dependence for which we get estimates in proposition 2.

Proposition 2. *There exists a positive constant C so that for all $n \in \mathbb{N}$ and for all $A \in \sigma(\mathcal{A}^n)$ the following estimate holds true:*

$$\left| \mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}(W_m^i = a)\mu(A) \right| \leqslant C\mu(A) \inf_{\Delta > 0} \left(\Delta\mu(A) + \sum_{j=r_A}^{\Delta} \delta_A(j) + \frac{\phi(\Delta)}{\mu(A)} \right)$$

where $W_m = \sum_{j=1}^{m} \chi_A \circ T^j$ and $W_m^i = \sum_{\substack{1 \leqslant j \leqslant m \\ j \neq i}} \chi_A \circ T^j$.

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

$$W_m^{i,-} = \sum_{j=1}^{i-(\Delta+1)} \chi_A \circ T^j, \qquad W_m^{i,+} = \sum_{j=i+\Delta+1}^{m} \chi_A \circ T^j,$$
$$U_m^{i,-} = \sum_{j=i-\Delta}^{i-1} \chi_A \circ T^j, \qquad U_m^{i,+} = \sum_{j=i+1}^{i+\Delta} \chi_A \circ T^j,$$
$$U_m^i = U_m^{i,-} + U_m^{i,+}, \qquad \tilde{W}_m^i = W_m^i - U_m^i = W_m^{i,-} + W_m^{i,+}$$

with the obvious modifications if $i < \Delta$ or $i > m-\Delta$. With these partial sums we distinguish between the hits that occur near the ith iteration, namely $U_m^{i,-}$ and $U_m^{i,+}$, and the hits that occur away from the ith iteration, namely $W_m^{i,-}$ and $W_m^{i,+}$.

The 'gap' of length $2\Delta+1$ allows us to use the mixing property in the terms $W_m^{i,\pm}$ and its size will be determined later by optimizing the error term.

We then have, for $0 \leqslant a \leqslant m-1$, $a \in \mathbb{N}_0$, that

$$\mathbb{P}(\{W_m = a+1\} \cap T^{-i}A) = \mathbb{P}(\{W_m^i = a\} \cap T^{-i}A) = \sum_{\substack{\vec{a} = (a^-, a^{0,-}, a^{0,+}, a^+) \\ \text{s.t } |\vec{a}| = a}} \mathbb{P}\big(\{W_m^{i,\pm} = a^\pm\} \cap \{U_m^{i,\pm} = a^{0,\pm}\} \cap T^{-i}A\big)$$

(intersection of five terms). For $0 \leqslant a \leqslant m-1$ we have

$$\left|\mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(W_m^i = a\right)\mu(A)\right| \leqslant R_1 + R_2 + R_3$$

and will estimate the three terms

$$R_1 = \left|\mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right)\right|$$

$$R_2 = \left|\mathbb{P}\left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}_m^i = a\right)\mathbb{P}\left(I_i = 1\right)\right|$$

$$R_3 = \left|\mathbb{P}\left(\tilde{W}_m^i = a\right) - \mathbb{P}\left(W_m^i = a\right)\right|\mu(A)$$

separately as follows.

Estimate of R_1. Here we show that short returns are rare when conditioned on $T^{-i}A$. Observe that

$$\{W_m^i = a\} \cap T^{-i}A \subset \left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right) \cup \left(\{U_m^i > 0\} \cap T^{-i}A\right)$$

$$\{\tilde{W}_m^i = a\} \cap T^{-i}A \subset \left(\{W_m^i = a\} \cap T^{-i}A\right) \cup \left(\{U_m^i > 0\} \cap T^{-i}A\right).$$

Since $U_m^i > 0$ implies that either $U_m^{i,+} > 0$ or $U_m^{i,-} > 0$ we obtain

$$\left|\mathbb{P}\big(\{W_m^i = a\} \cap T^{-i}A\big) - \mathbb{P}\big(\{\tilde{W}_m^i = a\} \cap T^{-i}A\big)\right| \leqslant \mathbb{P}\big(\{U_m^i > 0\} \cap T^{-i}A\big) \leqslant b_i^- + b_i^+$$

where

$$b_i^- = \mathbb{P}\big(\{U_m^{i,-} > 0\} \cap T^{-i}A\big) \quad \text{and} \quad b_i^+ = \mathbb{P}\big(\{U_m^{i,+} > 0\} \cap T^{-i}A\big).$$

We now estimate the two terms, b_i^- and b_i^+, separately as follows:

(i) Estimate of b_i^+: By lemma 1

$$\begin{aligned} b_i^+ &= \mathbb{P}\big(\{U_m^{i,+} > 0\} \cap T^{-i}A\big) \\ &= \mathbb{P}(U_m^{i,+} > 0 | I_i = 1)\mu(A) \\ &= \mathbb{P}_A(\tau_A \leqslant \Delta)\mu(A) \\ &\leqslant C\mu(A)\sum_{j=r_A}^{\Delta}\delta_A(j). \end{aligned}$$

(ii) Estimate of b_i^-: if $U_m^{i,-} > 0$ then $\{U_m^{i,-} > 0\} \subset \bigcup_{k=1}^{\Delta} T^{-(i-k)}A$ and therefore

$$\mathbb{P}\left(\{U_m^{i,-} > 0\} \cap T^{-i}A\right) \leqslant \mu\left(T^{-i}A \cap \bigcup_{k=1}^{\Delta} T^{-(i-k)}A\right).$$

We show the following symmetry

$$\mu\left(T^{-i}A\cap\bigcup_{k=1}^{\Delta}T^{-(i-k)}A\right)=\mu\left(T^{-i}A\cap\bigcup_{k=1}^{\Delta}T^{-(i+k)}A\right).$$

For that purpose let $S_i=\bigcup_{k=1}^{\Delta}J_{i,k}$ where $J_{i,k}=T^{-i}A\cap T^{-(i-k)}A$ and similarly $\tilde{S}_i=\bigcup_{k=1}^{\Delta}\tilde{J}_{i,k}$, $\tilde{J}_{i,k}=T^{-i}A\cap T^{-(i+k)}A$. We now want to show that $\mu(S_i)=\mu(\tilde{S}_i)$. We decompose S_i into a disjoint union as follows:

$$S_i=\bigcup_{k=1}^{\Delta}V_{i,k},$$

where

$$V_{i,k}=J_{i,k}\setminus\bigcup_{j=1}^{k-1}J_{i,k}\cap J_{i,j}.$$

Then

$$\mu(S_i)=\mathbb{P}\left(\dot{\bigcup}_{k=1}^{\Delta}V_{i,k}\right)=\sum_{k=1}^{\Delta}\mu(V_{i,k}).$$

Similarly, $\tilde{S}_i$ is the disjoint union of $\tilde{V}_{i,k}=\tilde{J}_{i,k}\setminus\bigcup_{j=1}^{k-1}\tilde{J}_{i,k}\cap\tilde{J}_{i,j}$, $k=1,\ldots,\Delta$. Then

$$F^{-k}V_{i,k}=F^{-k}J_{i,k}\setminus\bigcup_{j=1}^{k-1}F^{-k}\left(J_{i,k}\cap J_{i,j}\right)=\tilde{J}_{i,k}\setminus\bigcup_{j=1}^{k-1}\tilde{J}_{i,k}\cap\tilde{J}_{i,k-j}=\tilde{V}_{i,k},$$

where we have used that $F^{-k}J_{i,k}=\tilde{J}_{i,k}$ and $F^{-k}(J_{i,k}\cap J_{i,j})=\tilde{J}_{i,k}\cap\tilde{J}_{i,k-j}$, $0\leqslant j\leqslant k-1$. Therefore, by the invariance of the measure $\mu(\tilde{V}_{i,k})=\mu(V_{i,k})$ and consequently

$$\mu(S_i)=\sum_{k=1}^{\Delta}\mu(V_{i,k})=\sum_{k=1}^{\Delta}\mu(\tilde{V}_{i,k})=\mu(\tilde{S}_i).$$

We therefore obtain

$$b_i^-=\mu\left(\bigcup_{k=1}^{\Delta}T^{-(i-k)}A\cap T^{-i}A\right)=\mu\left(\bigcup_{k=1}^{\Delta}T^{-(i+k)}A\cap T^{-i}A\right)$$
$$=\mathbb{P}\left(\{U_m^{i,+}>0\}\cap T^{-i}A\right)=b_i^+$$

Combining (i) and (ii) yields

$$R_1\leqslant C\sum_{j=r_A}^{\Delta}\delta_A(j).$$

Estimate of R_3. Now we show that short returns are rare. We proceed similarly to the estimate of R_1. The set inclusions

$$\{W_m^i=a\}\subset\{\tilde{W}_m^i=a\}\cup\{U_m^i>0\}$$
$$\{\tilde{W}_m^i=a\}\subset\{W_m^i=a\}\cup\{U_m^i>0\}$$

let us estimate

$$\left|\mathbb{P}\left(\tilde{W}_m^i=a\right)-\mathbb{P}\left(W_m^i=a\right)\right|\leqslant\mathbb{P}\left(U_m^i>0\right)\leqslant 2\mathbb{P}\left(\bigcup_{k=1}^{\Delta}\{I_{i+k}=1\}\right)\leqslant 2\Delta\mu(A).$$

Hence

$$R_3\leqslant 2\Delta\mu(A)^2.$$

Estimate of R_2. This is the principal term and the speed of mixing now becomes relevant. Recall that $\tilde{W}_m^i(x) = W_m^{i,-}(x) + W_m^{i,+}(x)$ and

$$R_2 = \left|\mathbb{P}\left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}_m^i = a\right)\mu(A)\right|$$
$$= \left|\sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i}A\right) - \sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \mathbb{P}\left(W_m^{i,\pm} = a^\pm\right)\mu(A)\right|.$$

For each $\vec{a} = (a^-, a^+)$ for which $|\vec{a}| = a$ we have

$$\left|\mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i}A\right) - \mathbb{P}\left(W_m^{i,\pm} = a^\pm\right)\mu(A)\right| \leqslant R_{2,1} + R_{2,2} + R_{2,3}$$

where

$$R_{2,1} = \left|\mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i}A\right) - \mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right)\mathbb{P}\left(W_m^{i,-} = a^-\right)\right|$$
$$R_{2,2} = \left|\mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right) - \mathbb{P}\left(W_m^{i,+} = a^+\right)\mu(A)\right|\mathbb{P}\left(W_m^{i,-} = a^-\right)$$
$$R_{2,3} = \left|\mathbb{P}\left(W_m^{i,+} = a^+\right)\mathbb{P}\left(W_m^{i,-} = a^-\right) - \mathbb{P}\left(W_m^{i,\pm} = a^\pm\right)\right|\mu(A).$$

We now bound the three terms separately:

Bounds for $R_{2,1}$. Due to the mixing property

$$\left|\mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i}A\right) - \mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right)\mathbb{P}\left(W_m^{i,-} = a^-\right)\right|$$
$$\leqslant \phi(\Delta)\mathbb{P}\left(W_m^{i,-} = a^-\right)$$

we obtain

$$\left|\sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t}|\vec{a}|=a}} \mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i}A\right) - \sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right)\mathbb{P}\left(W_m^{i,-} = a^-\right)\right|$$
$$\leqslant \sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \phi(\Delta)\mathbb{P}\left(W_m^{i,-} = a^-\right)$$
$$\leqslant \phi(\Delta).$$

Bounds for $R_{2,2}$. We have

$$R_{2,2} = \mathbb{P}\left(W_m^{i,-} = a^-\right)\left|\mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right) - \mathbb{P}\left(W_m^{i,+} = a^+\right)\mu(A)\right|$$
$$\leqslant \phi(\Delta)\mathbb{P}\left(W_m^{i,-} = a^-\right)\mu(A)$$

and therefore

$$\sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \left|\mathbb{P}\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right)\mathbb{P}\left(W_m^{i,-} = a^-\right) - \mathbb{P}\left(W_m^{i,+} = a^+\right)\mathbb{P}\left(W_m^{i,-} = a^-\right)\mu(A)\right|$$

$$\leqslant \sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \phi(\Delta)\mathbb{P}\left(W_m^{i,-}=a^-\right)\mu(A)$$

$$\leqslant \phi(\Delta)\mu(A).$$

Bounds for $R_{2,3}$. Here we obtain

$$\sum_{\substack{\vec{a}=(a^-,a^+)\\ \text{s.t } |\vec{a}|=a}} \left|\mathbb{P}\left(W_m^{i,+}=a^+\right)\mathbb{P}\left(W_m^{i,-}=a^-\right)-\mathbb{P}\left(W_m^{i,\pm}=a^\pm\right)\right|\mu(A)\leqslant \phi(2\Delta)\mu(A).$$

Combining the estimates for $R_{2,1}$, $R_{2,2}$ and $R_{2,3}$ we obtain that

$$R_2 \leqslant R_{2,1}+R_{2,2}+R_{2,3}\leqslant C\phi(\Delta)$$

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

$$\left|\mathbb{P}(\{W_m^i=a\}\cap T^{-i}A)-\mathbb{P}(W_m^i=a)\mu(A)\right|\leqslant C\inf_{\Delta>0}\left(\mu(A)^2\Delta+\mu(A)\sum_{j=r_A}^{\Delta}\delta_A(j)+\phi(\Delta)\right),$$

for some $C\in\mathbb{R}^+$ independent of A. □

Proof of theorem 1. By proposition 2

$$\xi_a\leqslant C\inf_{\Delta>0}\left(\mu(A)^2\Delta+\mu(A)\sum_{j=r_A}^{\Delta}\delta_A(j)+\phi(\Delta)\right)$$

and therefore

$$\epsilon_{a,i}\leqslant \mu(A)+\frac{\xi_a}{\nu(A)}\leqslant C\inf_{\Delta>0}\left(\mu(A)\Delta+\sum_{j=r_A}^{\Delta}\delta_A(j)+\frac{\phi(\Delta)}{\mu(A)}\right).$$

Let us note that replacing the value t by $t^*=[\frac{t}{\mu(A)}]\mu(A)$ results in an error of order $\mathcal{O}(\mu(A))$. With the new estimates for the error term $\epsilon_{a,i}$ in hand we can now use lemma 2 to obtain (as $\log m=\mathcal{O}(|\log\mu(A)|)$ and $m=[t/\mu(A)]$) with $E=\{0,1,\ldots,k-1\}$:

$$\left|\mathbb{P}\left(\tau_A^k>\frac{t}{\mu(A)}\right)-\sum_{i=0}^{k-1}e^{-t}\frac{t^i}{i!}\right|\leqslant Ct(t\vee 1)\inf_{\Delta>0}\left(\mu(A)\Delta+\sum_{j=r_A}^{\Delta}\delta_A(j)+\frac{\phi(\Delta)}{\mu(A)}\right)|\log\mu(A)|.$$

∎

Proof of theorem 2.

(i) *Polynomial mixing.* In the polynomial case where $\phi(k)=\mathcal{O}(k^{-\beta})$ with some $\beta>2$ we have by assumption $\mu(A_w)=\mathcal{O}(w^{-\beta})$ which implies that $\delta_A(j)\leqslant\mathcal{O}((\frac{j}{2})^{-\beta})+\phi(\frac{j}{2})=\mathcal{O}(j^{-\beta})$ if $r_A\leqslant j\leqslant 2n$ where we used $w=\frac{j}{2}$. Similarly we obtain $\delta_A(j)\leqslant\mu(A)+\phi(j-n)=\mu(A)+\mathcal{O}(j^{-\beta})$ for $j\geqslant 2n$. This gives the estimate $\sum_{j=r_A}^{\Delta}\delta_A(j)=\mathcal{O}(r_A^{-(\beta-1)})+\Delta\mu(A)=\mathcal{O}(n^{-(\beta-1)})+\Delta\mu(A)$ and consequently

$$\left|\mathbb{P}\left(\tau_A^k>\frac{t}{\mu(A)}\right)-\sum_{i=0}^{k-1}\mathrm{e}^{-t}\frac{t^i}{i!}\right|\leqslant Ct(t\vee 1)\inf_{\Delta>0}\left(\Delta\mu(A)+\frac{1}{n^{\beta-1}}+\frac{\Delta^{-\beta}}{\mu(A)}\right)|\log\mu(A)|.$$

In order to optimize Δ put $\Delta = \frac{1}{\mu(A_n)^\omega}$ for some $\omega \in (0,1)$. Then we obtain

$$\inf_{\Delta>0}\left(\Delta\mu(A) + \frac{1}{n^{\beta-1}} + \frac{\Delta^{-\beta}}{\mu(A)}\right) \leqslant \mu(A)^{1-\omega} + \frac{1}{n^{\beta-1}} + \mu(A)^{\beta\omega-1}.$$

The best value for $w \in (0,1)$ is $\omega = \frac{2}{\beta+1}$ and therefore

$$\inf_{0<\omega<1}\left(\mu(A)^{1-\omega} + \frac{1}{n^{\beta-1}} + \mu(A)^{\beta\omega-1}\right) \leqslant 2\mu(A)^{\frac{\beta-1}{\beta+1}} + \frac{1}{n^{\beta-1}} \leqslant \frac{C}{n^{\beta-1}} \quad \forall n \in \mathbb{N},$$

for some constant C. Since by assumption $|\log(\mu(A))| \leqslant Kn^\eta$ we obtain

$$\inf_{0<\omega<1}\left(\mu(A)^{1-\omega} + \frac{1}{n^{\beta-1}} + \mu(A)^{\beta\omega-1}\right)|\log(\mu(A))| \leqslant C\frac{1}{n^{\beta-1-\eta}}$$

for some $C > 0$. Finally we obtain

$$\left|\mathbb{P}\left(\tau_A^k > \frac{t}{\mu(A)}\right) - \sum_{i=0}^{k-1} e^{-t}\frac{t^i}{i!}\right| \leqslant Ct(t \vee 1)\frac{1}{n^{\beta-1-\eta}}. \tag{14}$$

(ii) *Exponential mixing.* In this case $\phi(k) = \mathcal{O}(\vartheta^k)$ with $\vartheta < 1$ which combined with the assumption $\mu(A_w(A)) = \mathcal{O}(\vartheta^w)$ implies that $\delta_A(j) = \tilde{\theta}^j$ for some $\tilde{\theta} < 1$ (take e.g. $w = \min\{n, \frac{j}{2}\}$. Hence $\sum_{j=r_A}^{\Delta}\delta_A(j) = \mathcal{O}(\tilde{\theta}^{r_A}) = \mathcal{O}(\theta^n)$ for some $\theta < 1$. Hence

$$\left|\mathbb{P}\left(\tau_A^k > \frac{t}{\mu(A)}\right) - \sum_{i=0}^{k-1} e^{-t}\frac{t^i}{i!}\right| \leqslant Ct(t \vee 1)\inf_{\Delta>0}\left(\Delta\mu(A) + \theta^n + \frac{\theta^\Delta}{\mu(A)}\right)|\log\mu(A)|\,. \tag{15}$$

In order to estimate the RHS let us put $\Delta = (1+\epsilon)\frac{|\log\mu(A)|}{|\log\theta|}$ for some $\epsilon > 0$. Then

$$\inf_{\Delta>0}\left(\Delta\mu(A) + \theta^n + \frac{\theta^\Delta}{\mu(A)}\right) = (1+\epsilon)\frac{|\log\mu(A)|}{|\log\theta|}\mu(A) + \theta^n + \mu(A)^\epsilon$$

and therefore

$$\inf_{\Delta>0}\left(\Delta\mu(A) + \theta^n + \frac{\theta^\Delta}{\mu(A)}\right)|\log\mu(A)| \\ \leqslant \left((1+\epsilon)\frac{|\log\mu(A)|}{|\log\theta|}\mu(A) + \theta^n + \mu(A)^\epsilon\right)|\log\mu(A)|.$$

Since for any $\delta \in (0,1)$ $|\log x| = \mathcal{O}(\frac{1}{x^\delta})$ as $x \to 0^+$ we obtain $|\log\mu(A)| \leqslant C\frac{1}{\mu(A)^\delta}$ for some constant C independent of A. Hence, as the measure of cylinder sets decay exponentially fast we obtain

$$\inf_{\Delta>0}\left(\Delta\mu(A) + \theta^n + \frac{\theta^\Delta}{\mu(A)}\right)|\log\mu(A)| \leqslant Ce^{-\gamma n}$$

for some $\gamma > 0$. Therefore,

$$\left|\mathbb{P}\left(\tau_A^k > \frac{t}{\mu(A)}\right) - \sum_{i=0}^{k-1} e^{-t}\frac{t^i}{i!}\right| \leqslant Ct(t \vee 1)e^{-\gamma n}.$$

∎

4. Return times on Markov towers

4.1. Mixing properties derived on the Markov tower

Let F be a differentiable map on a manifold M and Ω_0 a subset of M. As in [36, 37] we assume that Ω_0 is partitioned into sets $\Omega_{0,i}$, $i = 1, 2, \dots$ so that there is a return time function $R : \Omega_0 \to \mathbb{N}$ which is constant on the partition elements $\Omega_{0,i}$ and which satisfies that F^R maps $\Omega_{0,i}$ bijectively to the entire set Ω_0. Let us put $\Omega_{j,i} = \{(x, j) : x \in \Omega_{0,i}\}$ for $j = 0, 1, \dots, R(\Omega_{0,i}) - 1$. The space $\Omega = \bigcup_{i=1}^{\infty} \bigcup_{j=0}^{R(\Omega_{0,i})-1} \Omega_{j,i}$ is called a *Markov tower* for the map T. It has the associated partition $\mathcal{A} = \{\Omega_{j,i} : 0 \leqslant j < R(\Omega_{0,i}), i = 1, 2, \dots\}$ which typically is countably infinite. On the tower Ω we have the map T which for $x \in \Omega_{0,i}$ is given by $T(x, j) = (x, j + 1)$ if $j < R(\Omega_{0,i}) - 1$ and $T(x, R(\Omega_{0,i}) - 1) = (F^{R(\Omega_{0,i})}, 0)$.

For points $x, y \in \Omega_0$ one defines the function $s(x, y)$ as the largest positive n so that $(T^R)^j x$ and $(T^R)^j y$ for $0 \leqslant j < n$ lie in the same sub-partition elements, that is $(T^R)^j x, (T^R)^j y \in \Omega_{0,i_j}$ for some $i_0, i_1, \dots, n - 1$.

The space of Hölder continuous functions $\mathcal{C}_\gamma$ consists of all functions φ on Ω for which $|\varphi(x) - \varphi(y)| \leqslant C_\varphi \gamma^{s(x,y)}$. The norm on $\mathcal{C}_\gamma$ is $\|\varphi\|_\gamma = |\varphi|_\infty + C_\varphi$, where C_φ is smallest possible.

Let ν be a finite given 'reference' measure on Ω and assume that the Jacobian JT^R with respect to the measure ν is Hölder continuous in the following sense: there exists a $\gamma \in (0, 1)$ so that

$$\left|\frac{JT^R x}{JT^R y} - 1\right| \leqslant \text{const}\gamma^{s(T^R x, T^R y)}$$

for all $x, y \in \Omega_{0,i}$, $i = 1, 2, \dots$.

If the return time R is integrable with respect to m then by [37] theorem 1 there exists a T-invariant probability measure μ (SRB measure) on Ω which is absolutely continuous with respect to ν. Moreover, the density function $h = \frac{d\mu}{d\nu} = \lim_{n\to\infty} \mathcal{L}^n \lambda$ is Hölder continuous, where λ can be any initial density distribution in $\mathcal{C}_\gamma$. The transfer operator $\mathcal{L} : \mathcal{C}_\gamma \to \mathcal{C}_\gamma$ is defined by $\mathcal{L}\varphi(x) = \sum_{x' \in T^{-1}x} \frac{\varphi(x')}{JT(x')}$, $\varphi \in \mathcal{C}_\gamma$, and has the property that ν is a fixed point of its adjoint, i.e. $\mathcal{L}^*\nu = \nu$. In [37] theorem 2(II) the $\mathscr{L}^1$-convergence was proven:

$$\|\mathcal{L}^k \lambda - h\|_{\mathscr{L}^1} \leqslant p(k)\|\lambda\|_\gamma \tag{16}$$

where the 'decay function' $p(k) = \mathcal{O}(k^{-\beta})$ if the tail decays polynomially with power β, that is if $\nu(R > j) \leqslant \text{const}.j^{-\beta}$. If the return times decay exponentially, i.e. if $\nu(R > j) \leqslant \text{const}.\vartheta^j$ for some $\vartheta \in (0, 1)$, then there is a $\tilde{\vartheta} \in (0, 1)$ so that $p(k) \leqslant \text{const}.\tilde{\vartheta}^k$.

Recall that for each $n \in \mathbb{N}$ the elements of the nth join $\mathcal{A}^n = \bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}$ of the partition $\mathcal{A} = \{\Omega_{i,j}\}$ are called n-cylinders. For each $n \in \mathbb{N}$ the n-cylinders $\mathcal{A}^n$ form a new partition of the space, a refinement of the original partition. The σ-algebra $\mathcal{F}$ generated by all n-cylinders $\mathcal{A}^\ell$, for all $\ell \geqslant 1$, is the σ-algebra of the system $(\Omega, \mathcal{F}, \mu)$.

We will need the following standard arithmetic lemma to carry estimates for cylinders over to union of cylinders.

Lemma 3. *Let $a_1, a_2, \dots$ and $b_1, b_2, \dots$ be positive reals. Then*

$$\left|1 - \frac{a_1 + a_2 + \cdots}{b_1 + b_2 + \cdots}\right| \leqslant \sup_i \left|1 - \frac{a_i}{b_i}\right|.$$

Proof. If we put $\epsilon = \sup_i |1 - \frac{a_i}{b_i}|$ then we have by assumption $(1 - \epsilon)b_i \leqslant a_i \leqslant (1 + \epsilon)b_i$. Summation over i yields

$$(1 - \epsilon)\sum_i b_i \leqslant \sum_i a_i \leqslant (1 + \epsilon)\sum_i b_i$$

and therefore

$$(1-\epsilon) \leqslant \frac{\sum_i a_i}{\sum_i b_i} \leqslant (1+\epsilon)$$

which implies the statement. □

Lemma 4. *There exists a constant* C_6 *so that* $\|\mathcal{L}^n \chi_A\|_\gamma \leqslant C_6$ *for all* $A \in \sigma(\mathcal{A}^n)$ *and* n.

Proof. We first show that

$$\left|\log \frac{JT^n(x)}{JT^n(y)}\right| \leqslant c_1 \gamma^{s(T^n x, T^n y)},$$

for all pairs $x, y \in A$, $A \in \mathcal{A}^n$ and $\forall n \in \mathbb{N}$. For $A \in \mathcal{A}^n$ and $x, y \in A$ we have $x, y \in \Omega_{i,j}$ for some $i < R_j = R(\Omega_{0,j})$. Put $n_0 = R_j - i$ and then successively $n_\ell = R_j - i + \sum_{k=1}^{\ell-1} R_{j_k}$, where the j_ℓ are such that $T^{n_\ell} x \in \Omega_{0,j_\ell}$. Clearly $T^{n_\ell} x, T^{n_\ell} y \in \Omega_{k,j_\ell}$ for $k < R_{j_\ell}$ for all ℓ for which $n_\ell \leqslant n$. Put $L = \max_{n_\ell \leqslant n} \ell$ and we get from the distortion property

$$\begin{aligned}
\left|\log \frac{JT^n(x)}{JT^n(y)}\right| &\leqslant \sum_{k=0}^{L-1} \left|\log \frac{JT^{R_{j_k}}(T^{n_k}(x))}{JT^{R_{j_k}}(T^{n_k}(y))}\right| \\
&\leqslant c_2 \sum_{k=0}^{L-1} \gamma^{s(T^{n_k}(x), T^{n_k}(y))} \\
&\leqslant c_3 \gamma^{s(T^{n_L}(x), T^{n_L}(y))} \\
&\leqslant c_1 \gamma^{s(T^n(x), T^n(y))}
\end{aligned}$$

for some c_1.

Now, if $x, y \in \Omega_{i,j}$ for some i, j, then let $A \in \mathcal{A}^n$ and $x', y' \in A$ be so that $T^n x' = x$ and $T^n y' = y$ (for x', y' to exist one needs $A \subset \Omega_{i,j}$). Then we obtain

$$\frac{\mathcal{L}^n h \chi_A(y)}{\mathcal{L}^n h \chi_A(x)} = \frac{h(y')}{h(x')} \frac{JT^n(x')}{JT^n(y')}$$

which implies by the above estimate and the regularity of the density h that

$$\left|\log \frac{\mathcal{L}^n h \chi_A(y)}{\mathcal{L}^n h \chi_A(x)}\right| \leqslant \left|\log \frac{JT^n(x')}{JT^n(y')}\right| + \left|\log \frac{h(y')}{h(x')}\right| \leqslant c_1 \gamma^{s(x,y)} + c_4 \gamma^{s(x',y')} \leqslant c_5 \gamma^{s(x,y)},$$

for which we can also write

$$\left|1 - \frac{\mathcal{L}^n h \chi_A(y)}{\mathcal{L}^n h \chi_A(x)}\right| \leqslant c_6 \gamma^{s(x,y)} \qquad \forall A \in \mathcal{A}^n.$$

Now any $A \in \sigma(\mathcal{A}^n)$ is the disjoint union of some $A_j \in \mathcal{A}^n$. We now apply lemma 3 with the identification $a_j = \mathcal{L}^n h \chi_{A_j}(x)$, $b_j = \mathcal{L}^n h \chi_{A_j}(y)$. Since $\mathcal{L}^n h \chi_A = \sum_j \mathcal{L}^n h \chi_{A_j}$ we obtain

$$\left|1 - \frac{\mathcal{L}^n h \chi_A(y)}{\mathcal{L}^n h \chi_A(x)}\right| \leqslant c_6 \gamma^{s(x,y)} \qquad \forall A \in \sigma(\mathcal{A}^n), \forall x, y \text{ in some } \Omega_{i,\ell}, \forall n.$$

Let us note that in particular (see [37] theorem 1(ii) and sublemma 1) that (as $\sum_{A \in \mathcal{A}^n} \chi_A = 1$)

$$\left|1 - \frac{\mathcal{L}^n 1(y)}{\mathcal{L}^n 1(x)}\right| \leqslant c_1 \gamma^{s(x,y)}.$$

Since $|\mathcal{L}^n 1|_\infty \leqslant 1$, we now obtain

$$\left|\mathcal{L}^n h \chi_A(x) - \mathcal{L}^n h \chi_A(y)\right| \leqslant |\mathcal{L}^n h \chi_A(y)| \cdot \left|1 - \frac{\mathcal{L}^n h \chi_A(x)}{\mathcal{L}^n h \chi_A(y)}\right| \leqslant C_6 \gamma^{s(x,y)}$$

for some constant C_6. Hence $\mathcal{L}^n h\chi_A \in \mathcal{C}_\gamma$ and, moreover, is bounded in the $\mathcal{C}_\gamma$-norm uniformly in $A \in \sigma(\mathcal{A}^n)$ and $n \in \mathbb{N}$. □

We proceed as in the proof of [37] theorem 3 and put $\lambda = \mathcal{L}^n h\chi_A$ which is a strictly positive function. Then $\eta = \frac{\lambda}{\mu(A)}$ is a density function as $\nu(\lambda) = \nu(\mathcal{L}^n h\chi_A) = \nu(h\chi_A) = \mu(A)$. Moreover, $\|\lambda\|_\gamma$ is by lemma 4 bounded by C_6 uniformly in n and $A \in \sigma(\mathcal{A}^n)$. Denote by $p(k)$, $k = 1, 2, \ldots$, the rate of the decay of correlations which is $p(k) = \mathcal{O}(k^{-\beta})$ if the return times tail decays like $k^{-\beta}$ and $p(k) = \mathcal{O}(\tilde{\vartheta}^k)$ for some $\tilde{\vartheta} \in (0, 1)$ if the return times tail decays exponentially. We obtain

$$\begin{aligned}\mu(A \cap T^{-k-n}B) - \mu(A)\mu(B) &= \nu(h\chi_A(\chi_B \circ T^{k+n})) - \nu(h\chi_A)\nu(h\chi_B)\\ &= \mu(A)\nu(\chi_B\mathcal{L}^k\eta) - \nu(h\chi_B)\\ &= \mu(A)\int \chi_B(\mathcal{L}^k\eta - h)\,\mathrm{d}\nu\\ &= \int_B (\mathcal{L}^k\lambda - \mu(A)h)\,\mathrm{d}\nu. \end{aligned} \tag{17}$$

In particular, the $\mathscr{L}^1$-convergence of $\mathcal{L}^k\eta - h$ from (16) yields

$$\begin{aligned}\left|\mu(A \cap T^{-k-n}B) - \mu(A)\mu(B)\right| &\leqslant \mu(A)\int \chi_B|\mathcal{L}^k\lambda - h|\,\mathrm{d}\nu\\ &\leqslant \begin{cases}\mu(B)\\ \mu(A)c_1\|\eta\|_\gamma p(k)\end{cases}\\ &\leqslant \begin{cases}\mu(B)\\ c_2 p(k)\end{cases}\end{aligned} \tag{18}$$

as $\|\eta\|_\gamma = \frac{1}{\mu(A)}\|\lambda\|_\gamma \leqslant \frac{C_6}{\mu(A)}$. The upper estimate which only uses boundedness of h and the pullbacks of the density η is useful for small k and $\mu(B)$. In particular this shows that the invariant measure on a Young tower is α-mixing but not ϕ-mixing.

Denote by $\hat{T} = T^R$ the induced map on Ω_0 given by $\hat{T}(x) = T^{R(x)}$ for $x \in \Omega_0$ and extended to the entire tower by putting $\hat{T}(x) = T^{R(\Omega_{0,i})-j}x$ for $x \in \Omega_{j,i}$. Similarly we extend R to the entire space Ω by putting $R(x) = R(\Omega_{j,i}) - j$ for $x \in \Omega_{j,i}$. To deal with short returns let $A \subset \Omega$ be a set with period r_A and put $\mathscr{S}(A) = \bigcup_j A_j$ for the smallest disjoint union so that $A \subset \mathscr{S}(A)$, where $A_j \in \sigma(\mathcal{A}^{\ell_j})$, and $\ell_j = \sum_{k=0}^{K_j-1} R(\hat{T}^k\tilde{A}_j)$ (for $K_j \geqslant 1$) is such that $\ell_j \leqslant \min(n, r_A)$.

Theorem 4. *As described above let T be a map on the Markov Tower structure Ω with a reference measure ν and return time function R. Let μ be the absolutely continuous invariant measure. Then for a sequence $A_n \in \sigma(\mathcal{A}^n)$ the following result holds true ($\tau^k_{A_n}$ is the kth entry time to A_n):*

(I) *If $\mu(A_n) \geqslant \mathrm{e}^{-Kn}$, $\mu(\mathscr{S}(A_n)) \leqslant \mathrm{e}^{-Ln}$ for some $0 < L \leqslant K$ then*

$$\left|\mathbb{P}\left(\tau^k_{A_n} > \frac{t}{\mu(A_n)}\right) - \mathrm{e}^{-t}\sum_{i=0}^{k-1}\frac{t^i}{i!}\right| \leqslant C_7(t \vee 1)\mathrm{e}^{-Gn} \quad \forall t > 0 \qquad \text{and} \qquad \forall n \in \mathbb{N},$$

for all $G < L$ if $p(k)$ is exponential and $G = \frac{1}{\beta+1}(\beta L - K)$ if $p(k) \sim k^{-\beta}$ is polynomial with $\beta > K/L$.

(II) *If $\mu(A_n) \geqslant n^{-\kappa}$, $\mu(\mathscr{S}(A_n)) \leqslant n^{-\lambda}$ for some $1 < \lambda \leqslant \kappa$ then*

$$\left|\mathbb{P}\left(\tau^k_{A_n} > \frac{t}{\mu(A_n)}\right) - \mathrm{e}^{-t}\sum_{i=0}^{k-1}\frac{t^i}{i!}\right| \leqslant C_7(t \vee 1)n^{-\gamma} \quad \forall t > 0 \qquad \text{and} \qquad \forall n \in \mathbb{N},$$

where $\gamma = \lambda - 1$ *if* $p(k)$ *is exponential and* $\gamma = \frac{\beta\lambda-\kappa}{\beta+1}$ *if* $p(k) \sim k^{-\beta}$ *is polynomial of order* $\beta > \kappa/\lambda$.

Note that in both cases, exponentially and polynomially decreasing sets A_n and $\mathscr{S}(A_n)$, the lowest possible bound for the value β is 1 for polynomially decaying return times tail $\nu(R > n) \sim n^{-\beta}$. In these cases one must have $K = L$ (exponential case) or $\kappa = \lambda$ (polynomial case).

4.2. Return time distribution

Here again we denote by $p(k), k = 1, 2, \ldots$, the rate of the decay of correlations as in (18), that is $p(k) = \mathcal{O}(k^{-\beta})$ if the return time tail decays like $k^{-\beta}$ and $p(k) = \mathcal{O}(\tilde{\vartheta}^k)$ for some $\tilde{\vartheta} \in (0, 1)$ if the return time tail decays exponentially. Let us now prove the main result for Markov towers.

Theorem 5. *Let* $T\colon \Omega \to \Omega$ *be a Markov tower as above with a 'reference measure'* m *and a return time function* R*. Let* μ *be the absolutely continuous invariant measure for* T *and* $p(k), k = 1, 2, \ldots$*, the rate of the decay of correlations.*

Let $A \in \sigma(\mathcal{A}^n)$*. Then for all* Δ $(n < \Delta \ll m)$ *and* $m \geqslant t$*:*

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant \text{const.} \left(\Delta\mu(\mathscr{S}(A)) + (2+t)\frac{p(\Delta - n)}{\mu(A)}\log m\right)$$

Proof. As before we put $W_m = \sum_{j=1}^m \chi_A \circ T^j$ and $W_m^i = \sum_{\substack{1 \leqslant j \leqslant m \\ j \neq i}} \chi_A \circ T^j$. We have to estimate the following quantity:

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| = \sum_{i=1}^m p_i \sum_{a=0}^m f(a+1)\epsilon_{a,i},$$

where

$$\epsilon_{a,i} = |\mathbb{P}(W_m = a) - \mathbb{P}(W_m = a+1 | I_i = 1)| \leqslant \left|\mathbb{P}(W_m = a) - \mathbb{P}(W_m^i = a)\right| + \frac{\xi_a}{\mu(A)},$$

and

$$\xi_a = \max_i \left|\mathbb{P}(\{W_m^i = a\} \cap T^{-i}A) - \mathbb{P}(W_m^i = a)\mu(A)\right|.$$

Clearly

$$\left|\mathbb{P}(W_m = a) - \mathbb{P}(W_m^i = a)\right| \leqslant \mathbb{P}(I_i = 1) = \mu(A)$$

which leaves us to estimate ξ_a and to execute the sum over a where we will use the bounds from lemma 4 for f.

Let $\Delta \ll m$ be the halfwith of the 'gap' and for $i \in (0, m]$ define as before

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

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

$$U_m^i = U_m^{i,-} + U_m^{i,+}, \qquad \tilde{W}_m^i = W_m^i - U_m^i = W_m^{i,-} + W_m^{i,+}$$

(with the obvious modifications if $i < \Delta$ or $i > m - \Delta$). For $a \in [0, m]$ we have

$$\mathbb{P}(\{W_m = a+1\} \cap T^{-i}A) = \mathbb{P}(\{W_m^i = a\} \cap T^{-i}A)$$
$$= \sum_{\substack{\vec{a}=(a^-, a^{0,-}, a^{0,+}, a^+) \\ \text{s.t } |\vec{a}|=a}} \mathbb{P}\big(\{W_m^{i,\pm} = a^\pm\} \cap \{U_m^{i,\pm} = a^{0,\pm}\} \cap T^{-i}A\big)$$

where the terms inside the sum are measures of intersections of five sets. Then

$$\mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(W_m^i = a\right)\mu(A) = R_1(a) + R_2(a) + R_3(a),$$

where

$$R_1(a) = \mathbb{P}\left(\{W_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right)$$
$$R_2(a) = \mathbb{P}\left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right) - \mathbb{P}\left(\tilde{W}_m^i = a\right)\mathbb{P}(I_i = 1)$$
$$R_3(a) = \left(\mathbb{P}\left(\tilde{W}_m^i = a\right) - \mathbb{P}\left(W_m^i = a\right)\right)\mu(A)$$

estimated separately as follows in increasing order of difficulty.

Estimate of R_3. We first show that short returns are rare. The set inclusions

$$\{W_m^i = a\} \subset \{\tilde{W}_m^i = a\} \cup \{U_m^i > 0\}$$
$$\{\tilde{W}_m^i = a\} \subset \{W_m^i = a\} \cup \{U_m^i > 0\}$$

let us estimate

$$\left|\mathbb{P}\left(\tilde{W}_m^i = a\right) - \mathbb{P}\left(W_m^i = a\right)\right| \leqslant \mathbb{P}\left(U_m^i > 0\right) \leqslant 2\mathbb{P}\left(\bigcup_{k=1}^{\Delta}\{I_{i+k} = 1\}\right) \leqslant 2\Delta\mu(A).$$

Hence

$$|R_3(a)| \leqslant 2\Delta\mu(A)^2$$

for every $a = 0, \ldots, m$.

Estimate of R_1. Here we show that short returns are rare when conditioned on $T^{-i}A$. Observe that

$$\{W_m^i = a\} \cap T^{-i}A \subset \left(\{\tilde{W}_m^i = a\} \cap T^{-i}A\right) \cup \left(\{U_m^i > 0\} \cap T^{-i}A\right)$$
$$\{\tilde{W}_m^i = a\} \cap T^{-i}A \subset \left(\{W_m^i = a\} \cap T^{-i}A\right) \cup \left(\{U_m^i > 0\} \cap T^{-i}A\right).$$

Since $U_m^i > 0$ implies that either $U_m^{i,+} > 0$ or $U_m^{i,-} > 0$ we obtain

$$\left|\mathbb{P}\big(\{W_m^i = a\} \cap T^{-i}A\big) - \mathbb{P}\big(\{\tilde{W}_m^i = a\} \cap T^{-i}A\big)\right| \leqslant \mathbb{P}\big(\{U_m^i > 0\} \cap T^{-i}A\big) \leqslant b_i^- + b_i^+$$

where

$$b_i^- = \mathbb{P}\big(\{U_m^{i,-} > 0\} \cap T^{-i}A\big) \qquad \text{and} \qquad b_i^+ = \mathbb{P}\big(\{U_m^{i,+} > 0\} \cap T^{-i}A\big).$$

It was shown in proposition 2 that $b_i^+ = b_i^-$.

Now let $\mathscr{S}(A)$ be a disjoint union of cylinders $A_j \in \sigma(\mathcal{A}^{\ell_j})$, where $\ell_j = \sum_{k=0}^{K_j-1} R(\hat{T}^k A_j)$ for some $K_j \geqslant 1$ is so that $\ell_j \leqslant \min(n, r_A)$. The set $\mathscr{S}(A)$ is chosen so that it contains A and is a disjoint union of A_j. This can be achieved since if there is a non-empty intersection of some A_j with some other cylinder A_k, then, say, $\ell_j < \ell_k$ which implies that $A_k \subset A_j$. It is then sufficient to retain A_j and to omit A_k. In order to estimate $\mu(A_j)$ put $\lambda_{A_j} = \mathcal{L}^{\ell_j} h \chi_{A_j}$.

Then $\lambda_{A_j}(x) = \frac{h(y)}{JT^{\ell_j}(y)}$, where $y \in A_j$ is such that $T^{\ell_j} y = x$, and x is any point in Ω_0. Since by [37] sublemma 2

$$\left|\log \frac{JT^{\ell_j}(y)}{JT^{\ell_j}(y')}\right| \leqslant c_1 \qquad \forall\, y, y' \in A_j,$$

for some c_1, and as the density $h \in \mathcal{C}_\gamma$ is positive, we obtain

$$\left|\log \frac{\lambda_{A_j}(x)}{\lambda_{A_j}(x')}\right| \leqslant c_2 \qquad \forall\, x, x' \in \Omega_0,$$

and thus $|\lambda_{A_j}|_\infty \in [\frac{1}{c_3}, c_3]\frac{1}{JT^{\ell_j}(y)} \quad \forall\, y \in A_j$. As a consequence $\nu(A_j)$ is similarly comparable to $\frac{1}{JT^{\ell_j}(y)} \quad \forall\, y \in A_j$ as $T^{\ell_j} : A_j \to \Omega_0$ is one-to-one ($c_3 > 0$) as $\ell_j = R(A_j)$. One also has $|\lambda_{A_j}|_\infty \leqslant c_4 \mu(A_j)$. Clearly $\{\tau_A \leqslant \Delta\} \subset \bigcup_{\ell=r_A}^{\Delta} T^{-\ell} A$ and thus

$$\mu(A \cap \{\tau_A \leqslant \Delta\}) \leqslant \sum_{\ell=r_A}^{\Delta} \mu(A \cap T^{-\ell} A),$$

where we can estimate as follows for $\ell \geqslant \ell_j$

$$\begin{aligned}
\mu(A \cap T^{-\ell} A) &\leqslant \sum_j \mu(A_j \cap T^{-\ell} A) \\
&= \sum_j \int_{T^{-(\ell-\ell_j)} A} \lambda_{A_j}\, \mathrm{d}\nu \\
&\leqslant \sum_j |\lambda_{A_j}|_\infty \nu(T^{-(\ell-\ell_j)} A) \\
&\leqslant c_5 \sum_j \mu(A_j)\mu(A).
\end{aligned}$$

Since $\mu(\mathscr{S}(A)) = \sum_j \mu(A_j)$ we obtain

$$b_i^+ = \mu_A(\{\tau_A \leqslant \Delta\}) \leqslant \sum_{\ell=r_A}^{\Delta} \frac{\mu(A \cap T^{-\ell} A)}{\mu(A)} \leqslant c_5 \Delta \mu(\mathscr{S}(A))$$

and thus

$$R_1(a) \leqslant b_i^+ + b_i^- \leqslant 2c_5 \Delta \mu(\mathscr{S}(A))$$

for all $a \in [0, m]$

Estimate of R_2. Here the decay of correlations play a central role. For $\tilde{W}_m^i(x) = W_m^{i,-}(x) + W_m^{i,+}(x)$ we obtain as in proposition 2

$$R_2(a) = \sum_{\substack{\vec{a} = (a^-, a^+) \\ \text{s.t } |\vec{a}| = a}} \left(\mathbb{P}\left(\{W_m^{i,\pm} = a^\pm\} \cap T^{-i} A\right) - \mathbb{P}\left(W_m^{i,\pm} = a^\pm\right) \mu(A)\right)$$

where $a^- + a^+ = a$. As before we split the summands into three separate parts $R_{2,1}$, $R_{2,2}$, $R_{2,3}$ which we sum over a and bound separately as follows.

Bounds for $R_{2,1}$. The mixing of sets formula (17) gives us

$$R_{2,1}(a^-,a^+) = \mu\left(\{W_m^{i,\pm}=a^\pm\}\cap T^{-i}A\right) - \mu\left(\{W_m^{i,+}=a^+\}\cap T^{-i}A\right)\mu\left(W_m^{i,-}=a^-\right)$$
$$= \int_{Y_{a^+}} \left(\mathcal{L}^{\Delta-n}\lambda_{a^-} - h\mu(X_{a^-})\right) \mathrm{d}\nu,$$

where $\lambda_{a^-} = \mathcal{L}^{i+n}h\chi_{X_{a^-}}$, $X_{a^-} = \{W_m^{i,-}=a^-\}$ and $Y_{a^+} = T^{\Delta-n}(\{W_m^{i,+}=a^+\}\cap T^{-i}A)$. According to lemma 4 $\|\lambda_{a^-}\|_\gamma \leqslant C_6$ for any value of a^-, i, m and n. Thus, summing over $a=0,\ldots,m$, we obtain

$$\left|\sum_{a=0}^m f(a+1)R_{2,1}(a^-,a^+)\right| \leqslant \sum_{a^-,a^+}\left|f(a^-+a^++1)\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-})\right)\mathrm{d}\nu\right|$$
$$\leqslant \sum_{a^+=0}^m\sum_{a^-=0}^m |f(a^-+a^++1)|\varepsilon_{a^-,a^+}\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-})\right)\mathrm{d}\nu$$

where ε_{a^-,a^+} is the sign of the integral $\int_{Y_{a^+}}(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-}))\,\mathrm{d}m$. We now split the sum over a^-, a^+ in geometric progression and use the bounds on $|f|$ from lemma 2 to obtain

$$\left|\sum_{a=0}^m f(a+1)R_{2,1}(a^-,a^+)\right| \leqslant \sum_{k=0}^{[\log_2 2m]}\sum_{a^-,a^+=0}^{[2m2^{-k}]}\frac{2+t}{a^-+a^++1}\varepsilon_{a^-,a^+}\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-})\right)\mathrm{d}\nu$$
$$+\sum_{a^-,a^+=0}^{[t]}\varepsilon_{a^-,a^+}\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-})\right)\mathrm{d}\nu.$$

The first (triple) sum is estimated by $I+II$, where I is for the terms with $\varepsilon=+1$ and II contains the terms for which $\varepsilon=-1$. For every k we use the fact that $\frac{2+t}{a^-+a^++1}\leqslant\frac{2+t}{m2^{-k}}$ for $a^-+a^+\in[m2^{-k},m2^{-(k-1)})$. Hence

$$I = \sum_{k=0}^{[\log_2 2m]}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{[2m2^{-k}]}\sum_{\substack{a^-\in[0,2m2^{-k}]\\ \text{s.t } \varepsilon_{a^-,a^+}=1}}\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}\lambda_{a^-}-h\mu(X_{a^-})\right)\mathrm{d}\nu$$
$$= \sum_{k=0}^{[\log_2 2m]}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{[2m2^{-k}]}\int_{Y_{a^+}}\left(\mathcal{L}^{\Delta-n}L_{k,a^+,1}-h\mu(\tilde{X}_{a^+,1})\right)\mathrm{d}\nu$$

(note that all terms are positive), where

$$L_{k,a^+,1} = \sum_{\substack{a^-\in[0,2m2^{-k}]\\ \text{s.t } \varepsilon_{a^-,a^+}=1}}\lambda_{a^-} = \mathcal{L}^{i+n}\chi_{\tilde{X}_{a^+,1}}$$

and $\tilde{X}_{a^+,1} = \bigcup_{\substack{a^-\in[0,2m2^{-k}]\\ \text{s.t } \varepsilon_{a^-,a^+}=1}} X_{a-}$ is a disjoint union in $\sigma(\mathcal{A}^{i+n})$. Hence by lemma 4 we have $\|L_{k,a^+,1}\|_\gamma\leqslant C_6$ for all values of a^+, i, n. We thus obtain

$$I \leqslant \sum_{k=0}^{[\log_2 2m]}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{[2m2^{-k}]}\|L_{k,a^+,1}\|_\gamma\, p(\Delta-n)$$
$$\leqslant C_6\sum_{k=0}^{[\log_2 2m]}\frac{2+t}{m2^{-k}}2m2^{-k}p(\Delta-n)$$
$$\leqslant c_6(2+t)p(\Delta-n)\log m.$$

Similarly one estimates the second contribution II by putting $L_{k,a^+,2} = \sum_{\substack{a^- \in [0,2m2^{-k}] \\ \text{s.t } \varepsilon_{a^-,a^+} = -1}} \lambda_{a^-} = \mathcal{L}^{i+n}\chi_{\tilde{X}_{a^+,2}}$ where $\tilde{X}_{a^+,2}$ is the disjoint union $\bigcup_{\substack{a^- \in [0,2m2^{-k}] \\ \text{s.t } \varepsilon_{a^-,a^+} = -1}} X_{a-}$. We then get as above in estimating the part I (again for every k we estimate $|f(a+1)| \leqslant \frac{2+t}{m2^{-k}}$ for $a^- + a^+ \in [m2^{-k}, m2^{-(k-1)})$):

$$
\begin{aligned}
II &= \sum_{k=0}^{[\log_2 2m]} \frac{2+t}{m2^{-k}} \sum_{a^+=0}^{[2m2^{-k}]} \sum_{\substack{a^- \in [0,m2^{-k}] \\ \text{s.t } \varepsilon_{a^-,a^+} = -1}} -\int_{Y_{a^+}} \left(\mathcal{L}^{\Delta-n}\lambda_{a^-} - h\mu(X_{a^-})\right) \mathrm{d}\nu \\
&= \sum_{k=0}^{[\log_2 2m]} \frac{2+t}{m2^{-k}} \sum_{a^+=0}^{[2m2^{-k}]} -\int_{Y_{a^+}} \left(\mathcal{L}^{\Delta-n}L_{k,a^+,2} - h\mu(\tilde{X}_{a^+,2})\right) \mathrm{d}\nu \\
&\leqslant \sum_{k=0}^{[\log_2 2m]} \frac{2+t}{m2^{-k}} \sum_{a^+=0}^{[2m2^{-k}]} \|L_{k,a^+,2}\|_\gamma p(\Delta-n) \\
&\leqslant c_6(2+t)p(\Delta-n)\log m
\end{aligned}
$$

as $\|L_{k,a^+,2}\|_\gamma \leqslant C_6$ by lemma 4.

In the same way one estimates the second sum above which does not involve a sum over k:

$$
\sum_{a^-,a^+=0}^{[t]} \varepsilon_{a^-,a^+} \int_{Y_{a^+}} \left(\mathcal{L}^{\Delta-n}\lambda_{a^-} - h\mu(X_{a^-})\right) \mathrm{d}\nu \leqslant C_6 t p(\Delta-n).
$$

These estimates combined yield ($c_7 \leqslant 2c_6 + C_6$)

$$
\left|\sum_{a=0}^{m} f(a+1)R_{2,1}(a^-,a^+)\right| \leqslant c_7(2+t)p(\Delta-n)\log m.
$$

Bounds for $R_{2,2}$. Here we obtain

$$
\begin{aligned}
R_{2,2}(a^-,a^+) &= \left(\mu\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right) - \mu\left(W_m^{i,+} = a^+\right)\mu(A)\right)\mu(W_m^{i,-} = a^-) \\
&= \mu\left(W_m^{i,-} = a^-\right) \int_{T^{\Delta-n}\{W_m^{i,+}=a^+\}} \left(\mathcal{L}^{\Delta-n}\lambda_* - h\mu(A)\right) \mathrm{d}\nu
\end{aligned}
$$

where $\lambda_* = \mathcal{L}^{i+n}h\chi_{T^{-i}A}$ and therefore we obtain the following estimate which is independent of the value of a:

$$
\begin{aligned}
&\left|\sum_{a^-+a^+=a} R_{2,2}(a^-,a^+)\right| \\
&\quad\leqslant \sum_{\substack{\vec{a} = (a^-,a^+) \\ \text{s.t } |\vec{a}| = a}} \mu(W_m^{i,-} = a^-) \left|\mu\left(\{W_m^{i,+} = a^+\} \cap T^{-i}A\right) - \mu(W_m^{i,+} = a^+)\mu(A)\right| \\
&\quad\leqslant \sum_{a^+} \int_{T^{\Delta-n}\{W_m^{i,+}=a^+\}} \left|\mathcal{L}^{\Delta-n}\lambda_* - h\mu(A)\right| \mathrm{d}\nu \\
&\quad\leqslant \int_{T^{\Delta-n}\bigcup_{a^+}\{W_m^{i,+}=a^+\}} \left|\mathcal{L}^{\Delta-n}\lambda_* - h\mu(A)\right| \mathrm{d}\nu \\
&\quad\leqslant C_6 p(\Delta-n)
\end{aligned}
$$

again using the fact that for different a^+ the sets $T^{\Delta-n}\{W_m^{i,+} = a^+\}$ are disjoint in $\sigma(\bigcup_{\ell=i}^\infty \mathcal{A}^\ell)$.

Bounds for $R_{2,3}$. We proceed as in the estimates about $R_{2,1}$. Put

$$R_{2,3}(a^-, a^+) = \mu(A)\left(\mu\left(W_m^{i,+} = a^+\right)\mu\left(W_m^{i,-} = a^-\right) - \mu\left(W_m^{i,\pm} = a^\pm\right)\right)$$

and we obtain in the same way that

$$\left|\sum_{0\leqslant a^-+a^+\leqslant m} f(a^- + a^+ + 1)R_{2,3}(a^-, a^+)\right| \leqslant c_7\mu(A)(2+t)p(2\Delta - n)\log m.$$

Combining the estimates for $R_{2,1}$, $R_{2,2}$ and $R_{2,3}$ we obtain that ($c_8 \leqslant 2c_7 + C_6$)

$$\left|\sum_{a=0}^{m} f(a+1)R_2(a)\right| \leqslant c_8(2+t)p(\Delta - n)\log m.$$

On the other hand, using the estimates on R_1 and R_3 together with the lemma 2 we obtain

$$\left|\sum_{a=0}^{m} f(a+1)(R_1(a) + R_3(a))\right| \leqslant \Delta(2\mu(A) + c_5\mu(\mathscr{S}(A)))\sum_{a=0}^{m}|f(a+1)|$$

$$\leqslant \Delta(2\mu(A) + c_5\mu(\mathscr{S}(A)))\left(t + (2+t)\log\frac{m}{t}\right)$$

Hence

$$\left|\sum_a f(a+1)\xi_a\right| \leqslant \left|\sum_{a=0}^{m} f(a+1)(R_1(a) + R_2(a) + R_3(a))\right|$$

$$\leqslant 2\Delta\mu(A)(\mu(A) + c_5\mu(\mathscr{S}(A)))\left(t + (2+t)\log\frac{m}{t}\right)$$

$$+ c_8(2+t)p(\Delta - n)\log m$$

if $m > t$, and therefore

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant c_9\Delta\mu(\mathscr{S}(A))\,(t + (2+t)\,|\log\mu(A)|) + c_8(2+t)\frac{p(\Delta - n)}{\mu(A)}\log m$$

as $m = [t/\mu(A)]$ for some $c_8, c_9 \in \mathbb{R}^+$ independent of A. □

Proof of theorem 4. Optimizing the error terms requires the gaps $\Delta = (\mu(\mathscr{S}(A_n))\mu(A_n))^{\frac{1}{1+\beta}}$. We now look at different decay rates, namely the two cases when (i) $\mu(A_n)$ decays polynomially and (ii) $\mu(A_n)$ decays exponentially.

(i) If the target set A_n has polynomially decaying measure, $\mu(A_n) \sim n^{-\kappa}$ and $\mu(\mathscr{S}(A_n)) \sim n^{-\lambda}$, then if $p(k) = \mathcal{O}(k^{-\beta})$ and the gaps Δ are of the order $n^{\frac{\kappa+\lambda}{\beta+1}}$ (where $\kappa/\lambda < \beta$ implies that $\Delta \ll m = [t/\mu(A_n)]$). If $p(k) = \mathcal{O}(\tilde{\vartheta}^k)$ is exponentially decaying then the best choice for the gaps is $\Delta \sim n + \log n$. Hence

$$\begin{cases} p(k) = \mathcal{O}(k^{-\beta}) \Rightarrow |\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant c_1 n^{-\frac{\beta\lambda-\kappa}{\beta+1}} \\ p(k) = \mathcal{O}(\tilde{\vartheta}^k) \Rightarrow |\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant c_1 n^{-(\lambda-1)} \end{cases}$$

for some c_1.

(ii) In the case when the return set A_n has exponentially decaying measure, $\mu(A_n) \leqslant \mathrm{e}^{-Kn}$ (e.g. single n-cylinders) and $\mu(\mathscr{S}(A_n)) \leqslant e^{-Ln}$ then theorem 5 implies in the polynomial case $p(k) \sim k^{-\beta}$:

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant c_1(t \vee 1)\Delta\mu(\mathscr{S}(A_n)) \leqslant c_2\mathrm{e}^{-\frac{\beta}{1+\beta}L + \frac{1}{\beta+1}K} \leqslant c_2\mathrm{e}^{-Gn},$$

where $G = \frac{1}{\beta+1}(\beta L - K)$ and in the exponential case $p(k) \sim \tilde{\vartheta}^k$:

$$|\mathbb{P}(W_m \in E) - \mu_0(E)| \leqslant c_3(t \vee 1)\mu(\mathscr{S}(A_n))\log n \leqslant c_3\mathrm{e}^{-Gn},$$

for any $G < L$. ∎

Acknowledgments

The work of NTAH was partially supported by a grant from the NSF (DMS-0301910) and the work of YP was partially supported by a grant from the NSF (DMS-0301910).

References

[1] Aaronson J, Denker M and Urbanski M 1993 Ergodic theory for Markov fibred systems and parabolic rational maps *Trans. Am. Math. Soc.* **337** 495–548
[2] Abadi M 2001 Exponential approximation for hitting times in mixing processes *Math. Phys. Electron. J.* **7** 343–63
[3] Abadi M 2004 Sharp error terms and necessary conditions for exponential hitting times in mixing processes *Ann. Probab.* **32** 243–64
[4] Abadi M 2001 Instantes de ocorrencia de eventos raros em processos misturadores *PhD Thesis* Instituto de Mathematica e Estatistica, Universidade de Sao Paulo
[5] Abadi M and Saussol B 2010 Hitting and returning into rare events for all alpha-mixing processes (arXiv:1003.4856)
[6] Abadi M and Vergne N Sharp error terms for return time statistics under mixing conditions *J. Theor. Probab.* **22** 11–37
[7] Abadi M and Vergne N 2008 Sharp errors for point-wise Poisson approximations in mixing processes *Nonlinearity* **21** 2871–85
[8] Abadi M and Vergne N 2009 Sharp error terms for return time statistics under mixing conditions *J. Theor. Probab.* **22** 18–37
[9] Arratia R, Goldstein L and Gordon L 1990 Poisson approximation and the Chen–Stein method *Stat. Sci.* **5** 403–25
[10] Barbour A D and Chen L H Y 2005 *An Introduction to Stein's Method* (*Lecture Notes Series* vol 4) (Singapore: World Scientific)
[11] Barbour A D and L H Y Chen 2005 *Stein's Method and Applications* (*Lecture Notes Series* vol 5) (Singapore: World Scientific)
[12] Brin M and Stuck G 2002 *Introduction to Dynamical Systems* 1st edn (New York: Cambridge University Press)
[13] Chazottes J-R and Collet P 2013 Poisson approximation for the number of visits to balls in nonuniformly hyperbolic dynamical systems *Ergod. Theory Dyn. Syst.* **33** 49–80
[14] Denker M, Gordin M and Sharova A 2004 A Poisson limit theorem for toral automorphisms *Illinois J. Math.* **48** 1–20
[15] Doeblin W 1940 Remarques sur la théorie métrique des fraction continue *Compos. Math.* **7** 353–71
[16] Doukhan P 1994 *Mixing: Properties and Examples* (*Lecture Notes in Statistices* vol 85) (New York: Springer)
[17] Galves A and Schmitt B 1997 Inequalities for hitting times in mixing dynamical systems *Random Comput. Dyn.* **5** 337–47
[18] Haydn N 2000 Statistical properties of equilibrium states for rational maps *Ergod. Theory Dyn. Syst.* **20** 1371–90
[19] Haydn N T A 2013 Entry and return times distribution *Dyn. Syst.: Int. J.* **28** 333–53
[20] Haydn N T A, Lacroix Y and Vaienti S 2005 Hitting and return times in ergodic dynamical systems *Ann. Probab.* **33** 2043–50
[21] Haydn N T A and Vaienti S 2009 The compound Poisson distribution and return times in dynamical systems *Probab. Theory Relat. Fields* **144** 517–42
[22] Haydn N T A and Wasilewska K Limiting distribution and error terms for the number of visits to balls in non-uniformly hyperbolic dynamical systems arXiv:1402.2990
[23] Hirata M 1993 Poisson law for axiom A diffeomorphisms *Ergod. Theory Dyn. Sys.* **13** 533–56
[24] Hirata M 1995 Poisson law for the dynamical systems with the 'self-mixing' conditions *Dyn. Syst. Chaos* vol 1 (River Edge, NY: Worlds Scientific) pp 87–96
[25] Kac M 1947 On the notion of recurrence in discrete stochastic processes *Bull. Am. Math. Soc.* **53** 1002–10
[26] Kifer Y 2011 Nonconventional Poisson limit theorems (arXiv:1110.2155v1)
[27] Kifer Y and Rapaport A 2012 Poisson and compound Poisson approximations in a nonconventional setup (arXiv:1211.5238)
[28] Kupsa M and Lacroix Y 2005 Asymptotics for hitting times *Ann. Probab.* **33** 610–4

[29] Lacroix Y 2002 Possible limit laws for entrance times of an ergodic aperiodic dynamical system *Israel J. Math.* **132** 253–64

[30] Liverani C, Saussol B and Vaienti S 1999 A probabilistic approach to intermittency *Ergod. Theory Dyn. Syst.* **19** 671–85

[31] Mañé R 1985 *Ergodic Theory and Differentiable Dynamics* (New York: Springer)

[32] Pitskel B 1991 Poisson law for Markov chains *Ergod. Theory Dyn. Syst.* **11** 501–13

[33] Psiloyenis Y 2008 Mixing conditions and return times on Markov towers *PhD Thesis* USC

[34] Stein C 1986 *Approximate Computation of Expectations IMS Lecture Notes #7*

[35] Wasilewska K 2013 Limiting distribution and error terms for the number of visits to balls in mixing dynamical systems *PhD Thesis* USC

[36] Young L-S 1998 Statistical properties of dynamical systems with some hyperbolicity *Ann. Math.* **7** 585–650

[37] Young L-S 1999 Recurrence time and rate of mixing *Israel J. Math.* **110** 153–88