ENTRY AND RETURN TIMES DISTRIBUTIONS

NICOLAI HAYDN

ABSTRACT. We begin with the Poincaré recurrence theorem and Kac's theorem which for ergodic measures provides the value of the average return time. An example that shows that limiting entry and return times can be arbitrary if one only choses a suitable sequence of subsets along which to take the limit underlines the importance to let the return sets shrink to a point in a dynamically meaningful way. We then show that return and hitting times are related by a simple formula. Then for the induced map on a subset (of positive measure) we show that limiting entry times are the same for the induced map and measure as they are for the original map and measure. For α mixing measures we then provide Abadi's proof that the limiting distribution along cylinder sets is exponential almost surely. In the last section we present the method of Stein and Chen to show that the limiting distribution for α-mixing measures is Poissonian almost surely.

CONTENTS

1. Introduction 1
2. Poincaré recurrence theorem 5
3. Return times and the induced map 5
3.1. Kac's theorem and the induced map 5
3.2. Example 7
3.3. Entry and return times distributions 9
3.4. Relation between entry and return times distributions 11
3.5. Entry and return times for the induced map 12
3.6. Limiting entry times distributions 16
4. Higher order returns 21
4.1. Stein-Chen method 21
4.2. Higher order returns for α-mixing measures 22
4.3. ϕ-mixing measures 27
References 28

1. INTRODUCTION

In dynamics the deterministic development under the application of a map yields nevertheless to random behaviour on a long time scale, This is on an elementary level for ergodic measures this is expressed through the ergodic theorem which postulates that time averages converge in the limit to the space average or the given function. Systems

Date: March 28, 2019.

that have more structured independence over long times can satisfy a number of limiting theorems which otherwise are familiar from i.i.d. random variables such as decay of correlation, the central limit theorem for sufficiently regular function or the limiting distribution of entry and return times. In these notes we address this latter topic. The first result in this direction is due to Poincaré [35] from 1899 which states that for every finite invariant measure μ almost every point of any positive measure set U returns in finite time without however providing any estimate on how long this time is expected to be. That means that the return time function

$$\tau_U(x) = \min\{j \geq 1 : T^j x \in U\}$$

where T is the tranformation on the finite measure space Ω (see below). It was left to Kac [30] to show in 1947 that the average time is the reciprocal of the measure of the return set under the assumption that the measure be ergodic, more precisely $\int_U \tau_U \, d\mu = 1$.

Interest in return times distribution took off in the 90s starting with the paper by Pitskel [34] who showed that for Axiom A systems the limiting return times to cylinders are almost surely Poisson distributed. He used that Axiom A systems allow Markov partitions (see e.g. [10]) and thus a symbolic representation by subshifts of finite type. To be more precise, if $\mathcal{A}$ is the partiton and $\mathcal{A}^n$ denotes the nth refinement where $A_n(x)$ denotes the unique element in $\mathcal{A}^n$ that contains the point x, then he showed that

$$\mathbb{P}(\xi^t_{A_n(x)} = r) \to e^{-t} t^r / r!, \quad \forall t > 0,$$

as $n \to \infty$ for almost every point x, where

$$\xi^t_{A_n(x)} = \sum_{j=0}^{t/\mu(A_n(x))} \chi_{A_n(x)} \circ T^j$$

counts the number of hits to $A_n(x)$ along the orbit segment of length $t/\mu(A_n(x))$. (Notice that $\xi^t_{A_n(x)}(y) = 0$ exactly if $\tau_{A_n(x)}(y) > t/\mu(A_n(x))$.) He used the method of moments where the factorial moments can be directly linked to the higher order mixing properties which for such systems are implied by the ψ-mixing property (for the definition of various kinds of mixing see section 3.6). For two dimensional tori he moreover showed using an approximation argument that for hyperbolic maps the limiting return times to metric discs are Poisson almost surely. Interestingly enough, Pitskel also showed that at periodic points the return times distribution converges to a combination of a point mass at zero which encapsulates the periodicity followed by an exponential distribution. In fact he showed that for a periodic point x with minimal period m one has

$$\mathbb{P}_{A_n(x)}(\tau_{A_n(x)} > t/\mu(A_n(x))) \to \vartheta e^{-t\vartheta}, \quad \forall t > 0,$$

where he determined that the value of the extremal index ϑ is given by $1 - \exp \sum_{j=0}^{m-1} f(T^j x)$ with f being the potential of the invariant measure (we assume the pressure of f is zero).

Nearly concurrently, Hirata [27, 28] showed a similar result using the Laplace transform. This method considers the transfer operator restricted to the complement of the target cylinder and requires delicate estimates on the convergence of the principal eigenvalue.

The limiting distribution of hitting times was shown by Galves and Schmitt [16] to be exponentially distributed almost surely for ψ-mixing systems where ψ is only required to be summable. Their argument pivots on the estimate that up to a small error $\mathbb{P}(\tau_U > t+s)$

equals $\mathbb{P}(\tau_U > t)\mathbb{P}(\tau_U > s)$. By choosing $t = rs$ it then follows that for a good choice of r and s the quantity $\mathbb{P}(\tau_U > rs)$ is by a recursion argument approximated by $\mathbb{P}(\tau_U > s)^r$. This then leads to an exponential law since for small values of s one has $\mathbb{P}(\tau_U < s) \approx s\mu(U)$ which then yields $\mathbb{P}(\tau_U > t/\mu(U)) \sim (1 - s\mu(U))^{\frac{t}{s\mu(U)}} \approx e^{-t}$.

The direct approach which this paper pioneered also yields rates of convergence but cannot be practically used for higher order returns. The method however has been shown to be quite powerful and allowed for generalisations to ϕ-mixing and then also α-mixing measures [1, 2, 3, 4, 6]. We will present this result in section 3.6 for ϕ-mixing measures where the conclusion is particularly strong as it establishes the exponential limit law at every non-periodic point and for periodic points one obtains, as Pitskel did for equilibrium states, that the limiting return times distribution has a point mass at $t = 0$ and then falls off exponentially beginning at a lower level which is referred to as the extremal value.

A number of results on the limiting distribution of entry times are mirrored by extremal values laws. EVLs looks at the probability that a sequence of random variables X_n exceeds a given threshold which depends on the observation length. Typically the random variable X_n is taken to be $-\log d(T^j\cdot, x)$ for a fixed point x and this can be translated into an entry times problem. For details see [15].

If $A \subset \Omega$ and $t > 0$ a parameter, then as before $\xi_A^t = \sum_{j=1}^{[t/\mu(A)]} \chi_A \circ T^j$ is the counting function that counts the number of times the orbit of a point x visits the set A on the orbit segment of length $\frac{t}{\mu(A)}$. If the return times $\tau_A \circ \hat{T}^j$, $j = 0, 2, \dots$, are independent, where $\hat{T} = T^{\tau_A}$ is the induced map on A, and are exponentially distributed then ξ_A^t will be Poisson distributed. Since in the deterministic systems we consider, independence of entries takes plade in the limit as separation goes to infinity, we are lead to conclude that entry times should be in the limit Poisson distributed in much wider settings as considered by Pitskel in his trendsetting paper.

The first result on higher returns however was due to Doeblin in 1940 for the Gauss map at the origin. Then there was a long gap and nothing much seems to have happened until 1991 when Pitskel followed by several people began to work in this area with different methods. To recall Doeblin's result on the Gauss map let $\Omega = (0, 1]$ be the unit interval. The Gauss map is then given by $Tx = \frac{1}{x}$ mod 1 and is related to the continued fraction expansion of real numbers. If $[a_0, a_1, a_2, \dots]$ is the continued fraction expansion of a point $x \in \Omega$ then

$$x = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}}$$

where the integers $a_j \in \mathbb{N}$ are uniquely determined by x and are given by $a_j = \frac{1}{x} - Tx$. The Gauss measure μ on $(0, 1]$ is the unique absolutely continuous T-invariant probability measure. Its density is $\frac{1}{\log 2}\frac{1}{1+x}$ and Doeblin [**?**] showed that for every $\theta > 0$:

$$\mathbb{P}\left(|\{j : 1 \le j \le n, a_j(x) \ge \theta n\}| = p\right) \to e^{-1/(\theta \log 2)} \frac{1}{(\theta \log 2)^p p!}$$

as $n \to \infty$. Since $x = \frac{1}{a_j + Tx} = \frac{1}{a_j + \mathcal{O}(1)}$ (as $0 < Tx \le 1$) we see that a point $x \in (0,1]$ for which $|\{j : 1 \le j \le n, a_j(x) \ge \theta n\}| = p$ visits the interval $\left(0, \frac{1}{n\theta}\right)$ typically exactly p times on the orbit segment of length n. Since $\mu\left(\left(0, \frac{1}{n\theta}\right)\right) = \frac{1}{\log 2}\log(1 + \frac{1}{\theta n}) \approx \frac{1}{\log 2}\frac{1}{\theta n}$ we can put $A_m = (0, \frac{1}{m})$ and see that Doeblin's statement translates into

$$\mathbb{P}\left(\xi^t_{A_m} = p\right) \to e^{-t}\frac{t^p}{p!}, \quad \forall t > 0,$$

as $m \to \infty$. In other words, the limiting distribution of return times at the origin is Poissonian.

For higher order returns, Pitskel's result was generalised by various different methods:
(i) For interval maps an ad hoc method was provided in [29] to show that the limiting return times are Poisson. The same paper incidentally also shows that first return and entry times can only agree if they are both exponential. A more general result that ties entry times distributions to return times distributions is given in Theorem 5.
(ii) Using generating functions and moment estimates it was shown in [22] that for ψ-mixing measures, which include equilibrium states for Axiom A systems, the limiting distribution is almost surely Poisson. A similar approach determined [23] that at periodic points the limiting distribution is Pólya-Aeppli distributed, that is compound Poisson with a geometric distribution of the cluster size.
(iii) For torus automorphisms it was shown in [14] by the Chen-Stein method and using harmonic analysis that the return times distribution converges to Poisson for non-periodic points.
(iv) Using the Chen-Stein method the limiting distribution was shown to be Poisson a.s. in [5] for ψ and β-mixing measures by employing an estimate due to Arratia, Goldstein and Gordon [8] and also in [19] using a more direct approach which also yields the limiting Poisson distribution for all non-periodic points.
(v) By an elementary approach, Abadi and Vergne [7] determined that for ϕ-mixing measures returns to cylinder set approximations are in the limit Poisson distributed and also provide a rate of convergence.
(vi) In [19] we showed for ϕ-mixing measure that for suitable sets the limiting distribution at non-periodic points converges to Poisson. This uses the Chen-Stein method and will be described in section 4 for α-mixing and ϕ-mixing measures.

One of the underlying facts is that the limiting distributions a priori must depend on the way the target sets shrink to a single point. This was addressed in [33, 32] where it was shown that for ergodic measures one can arbitrarily prescribe any (reasonable) distribution and then find a sequence of sets U_n so that $\bigcap_n U_n$ is a single point x and so that the limiting distribution for either entry times or return times is equal to that arbitrarily given one. We will state this result and provide its proof in section 3.3. It is consequently clear that the approximating return sets U_n must be dynamically or geometrically representative. Naturally, cylinder sets encode the dynamics of the map and consequently yield relevant limiting distributions. Two other choices are for metric spaces to consider Bowen balls and geometric balls.

For non-uniformly hyperbolic maps Chazaottes and Collet [13] proved the limiting distribution for approximating balls to be Poisson. This applies to all maps on manifolds that allow for a Young tower construction (see [36, 37]) with one-dimensional unstable leaves and exponentially decaying correlations. This was extended in [20] to arbitrary dimensional unstable leaves and polynomial decay of correlations. A more general version is contained in [24]. In both cases the approach is to estimate the total variation distance between the counting function $\xi_{U_n}^t$ to a similar sum of i.i.d. zero-one random variables.

A third possible choice of approximating return set U_n is to take Bowen balls which are given by $B_{\varepsilon,n}(x) = \{y : d(T^j x, T^j y) < \varepsilon, j = 0, \dots, n-1\}$ where $\varepsilon > 0$ will ultimately go to zero. For such sets the limiting return times distribution was shown to be Poisson for ϕ-mixing measures and also for some α-mixing measures in [25, 26].

Finally let us mention that non-conventional return times limit results were obtained by Kifer and Papaport [31] for ψ-mixing measures. Also, recently there have been quenched limiting results for some random systems.

2. Poincaré recurrence theorem

Let (Ω, T, μ) be a dynamical system that consists of a space Ω, a map $T : \Omega \to \Omega$ and a T-invariant probability measure μ on Ω (μ is *T-invariant* if $\mu(U) = \mu(T^{-1}U)$ for all measurable $U \subset \Omega$).

Theorem 1. *(Poincaré recurrence theorem) Let $T : \Omega \to \Omega$ and μ be a T-invariant probability measures. For $U \subset \Omega$ put $\tau_U(x) = \min\{k \geq 1 : T^k x \in U\}$ for the return time of $x \in \Omega$ (we have $\tau_U(x) = \infty$ if the forward orbit of x never intersects U). If $\mu(U) > 0$, then $\tau_U(x) < \infty$ for almost every $x \in U$.*

Proof. Let $U \subset \Omega$ have positive measure and put $U_n = \bigcup_{j=n}^{\infty} T^{-j}U$ for the set of points $x \in \Omega$ that enter U at least once after time n. Obviously $U_0 \supset U_1 \supset U_2 \supset \cdots$. We also have $U_n = T^{-1}U_{n+1}$ which implies by invariance of the measure that $\mu(U_n) = \mu(T^{-1}U_{n+1}) = \mu(U_{n+1})$ and consequently $\mu(U_0) = \mu(U_n)\ \forall\ n$. Now $W = \bigcap_{n=1}^{\infty} U_n = \{x \in \Omega \text{ enters } U \text{ infinitely often}\}$ and $V = W \cap U = \{x \in U \text{ enters } U \text{ infinitely often}\}$. Since $\mu(U_0) = \mu(U_n)$ we obtain that $\mu(W) = \mu(U_0)$ and since $U \subset U_0$ we conclude that $\mu(V) = \mu(U)$. ∎

The recurrence statement is not true if the measure is infinite. As an example one can take the Lebesgue measure on $\mathbb{R}$ and the map $T : \mathbb{R} \to \mathbb{R}$ given by $Tx = x + 1$. No set of positive measure is recurrent.

3. Return times and the induced map

3.1. **Kac's theorem and the induced map.** For $U \subset \Omega$ we define the return time $\tau_U(x) = \min\{j \geq 1 : T^j x \in U\}$. By the Poincaré recurrence theorem $\tau_U(x) < \infty$ for almost every $x \in U$ for any finite T-invariant measure μ on Ω. Poincaré's theorem doesn't tell us anything about how big τ_U is. The next result gives us the expected value of τ_U on U which in particular implies that τ_U is integrable on U (assuming $\mu(U) > 0$).

Theorem 2. *(Kac 1947) If μ is an ergodic T-invariant probability measure on Ω then for any $U \subset \Omega$ of positive measure one has*

$$\int_U \tau_U(x)\, d\mu(x) = 1.$$

Proof. Let us put τ_U^k for the kth return time, that is we put $\tau_U^1 = \tau_U$ and then define recursively

$$\tau_U^k(x) = \tau_U(\hat{T}^{k-1}x) + \tau_U^{k-1}(x)$$

where we put $\hat{T}(x) = T^{\tau_U(x)}(x)$ for the *induced transformation* on U (which exists almost surely by Poincaré). Inductively we also get

$$\tau_U^k = \tau_U + \tau_U \circ \hat{T} + \tau_U \circ \hat{T}^2 + \cdots + \tau_U \circ \hat{T}^{k-1},$$

i.e. the kth return time is the kth ergodic sum of τ_U on $(U, \hat{T})$ By the pointwise ergodic theorem we get

$$\int_U \tau_U \, d\mu = \int_\Omega \chi_U \tau_U \, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} (\chi_U \tau_U)(T^j x)$$

for μ-almost every $x \in \Omega$ as μ is ergodic. If we take the limit along a subsequence $n_\ell = \tau_U^\ell(x)$ and use the fact that

$$(\chi_U \tau_U)(T^j x) = \begin{cases} 0 & \text{if} \quad T^j x \notin U \\ \tau_U(T^j x) & \text{if} \quad T^j x \in U \end{cases}$$

then we get (with $n = \tau_U^\ell$)

$$\int_U \tau_U \, d\mu = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\tau_U^\ell - 1} (\chi_U \tau_U)(T^j x) = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\ell-1} \tau_U(\hat{T}^j x) = \lim_{\ell\to\infty} \frac{1}{\tau_U^\ell(x)} \tau_U^\ell(x) = 1.$$

It remains to show that $\chi_U \tau_U \in \mathscr{L}^1$. We use the same argument again but this time cut off the values of τ_U. For R large we put

$$\varphi_R(x) = \begin{cases} 1 & \text{if} \quad \tau_U(x) \le R \\ 0 & \text{if} \quad \tau_U(x) > R \end{cases}.$$

Now, since $\varphi_R \chi_U \tau_U \in \mathscr{L}^1$, we get by the pointwise ergodic theorem

$$\int_U \varphi_R \tau_U \, d\mu = \int_\Omega \varphi_R \chi_U \tau_U \, d\mu = \lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n-1} (\varphi_R \chi_U \tau_U)(T^j x) = \lim_{n\to\infty} \frac{1}{\tau_U^\ell(x)} \sum_{j=0}^{\tau_U^\ell - 1} (\varphi_R \tau_U)(\hat{T}^\ell x) \le 1$$

for all values of R. Let $R \to \infty$ which implies that $\chi_U \tau_U \in \mathscr{L}^1$. ∎

Second proof of Kac's theorem if T is invertible. This one uses a representation of Ω which is called a *Rokhlin tower*. Given $U \subset \Omega$ ($\mu(U) > 0$), then we put $U_k = \{x \in U : \tau_U(x) = k\}$, $k = 1, 2, \ldots$, for the level sets of τ_U. Then $U = \dot{\bigcup}_{k=1}^\infty U_k$ is a disjoint union and the sets $T^j U_k$ for $j = 0, 1, \ldots, k-1, k = 1, 2, \ldots$, are pairwise disjoint. Since μ is ergodic, $\Omega = \bigcup_k \bigcup_{j=1}^{k-1} T^j U_k$ and as T is invertible $\mu(T^j U_k) = \mu(U_k)$. Hence

$$1 = \mu(\Omega) = \sum_{k=1}^\infty \sum_{j=0}^{k-1} \mu(T^{-1} U_k) = \sum_k k \mu(U_k) = \int_U \tau_U \, d\mu$$

This uses the representation of Ω by the following tower construction which is due to Rokhlin. Let F be a map on a space Δ_0 and assume Δ_0 is decomposed into a disjoint union $\Delta_0=\bigcup_k\Delta_{k,0}$. Then, given a (roof) function $r:\mathbb{N}\to\mathbb{N}$, we put

$$\Delta=\bigcup_{k=1}^{\infty}\bigcup_{j=0}^{r(k)-1}\Delta_{k,j}$$

(disjoint union), where $\Delta_{k,j}=\{(x,j):x\in U_k\}$. Then there is a map S on Δ which is defined by

$$\begin{aligned} S &: \Delta_{k,j}\to\Delta_{k,j+1} \text{ if } j\le r(k)-1\\ S &: \Delta_{k,r(k)-1}\to\Delta_0=\bigcup_k\Delta_{k,0}\end{aligned}$$

where $S(x,j)=(x,j+1)$ for $(x,j)\in\Delta_{k,j}$ and if $j<f(k)-1$. If $(x,j)\in\Delta_{k,r(k)-1}$ then the map is $S(x,j)=(F(x),0)$. We call the pair (S,Δ) a *Rokhlin tower*. In the case of Kac's theorem $\Delta_0=U$ and the roof function is $f(k)=k$.

For a subset $U\subset\Omega$, $\mu(U)>0$, let us denote by $\hat{T}=T^{\tau_U}:U\to U$ the *induced map*. $\hat{T}$ exists by Poincaré's (or Kac's) theorem almost everywhere. We also have the *induced measure* $\hat{\mu}$ which is defined on U by $\hat{\mu}(A)=\frac{\mu(A)}{\mu(U)}$ for all measurable $A\subset U$. Moreover $\hat{\mu}$ is $\hat{T}$-invariant and also ergodic w.r.t. $\hat{T}$ if μ is ergodic w.r.t. T. Ergodicity of μ can only be inferred from the ergodicity of $\hat{\mu}$ if the additional assumption $\Omega=\bigcup_{j=0}^{\infty}T^{-j}U$ is satisfied.

3.2. **Example.** Kac's theorem states that the return time function τ_U is integrable over U and also gives the value of the integral. Here we give an example of a system for which τ_U is not integrable over the entire space Ω and yet the measure is ergodic under the shift map.

Let $\Omega=\mathbb{N}^{\mathbb{Z}}$ where on the state space $\mathbb{N}$ we give the transition probabilities: Let $p_i\in(0,1), i=1,2,\dots,$ be a sequence, then we allow for the transition $i\to i+1$ with probability p_i and for the transition $i\to1$ with probability $q_i=1-p_i$. In other words, we can define a stochastic matrix M by

$$\left\{\begin{array}{rcl} M_{j,1} &=& q_j\\ M_{j,j+1} &=& p_j\\ M_{j,k} &=& 0 \text{ otherwise, i.e. if } k\ne1 \text{ or } k\ne j+1\end{array}\right. ,$$

where the transition probability of the transition $j\to k$ is given by the entry $M_{j,k}$. Then $M\mathbb{1}=\mathbb{1}$ as $\sum_{k=1}^{\infty}M_{j,k}=M_{j,1}+M_{j,j+1}=q_j+p_j=1\forall j$ and M has the left eigenvector $\vec{x}=(x_1,x_2,\dots)$ (for the dominant eigenvalue 1) which satisfies

$$\begin{aligned} q_1x_1+q_2x_2+q_3x_3+\cdots &= x_1\\ x_jp_j &= x_{j+1}\quad\text{for } j=1,2,\dots.\end{aligned}$$

One sees that the components of the left eigenvector are $x_j=x_1P_j, j=2,3,\dots,$ where $P_j=\prod_{i=1}^{j-1}p_i$ ($P_1=1$) and x_1 is chosen to make $\vec{x}$ a probability vector ($x_1^{-1}=\sum_jP_j$). The first equation above is satisfied as $\sum_jq_jx_j=\sum_j(1-p_j)x_1P_j=x_1\sum_j(P_j-P_{j+1})=x_1P_1=x_1$ if $P_j\to0$ as $j\to\infty$. In this way we obtain a shift invariant probability measure μ on Ω which is ergodic as one can go from any state i to any other state j with positive probability.

Put $A_j = \{\vec{\omega} \in \Omega : \omega_0 = j\}, j = 1, 2, \dots,$ let $U = A_1$ be the return set and τ_U its return/entry time function. If we put $A_{j,k} = A_j \cap \{\tau_U = k\}$ then $\vec{\omega} \in A_{j,k}$ is of the form $\omega_0\omega_1 \cdots \omega_k = j(j+1)(j+2)\cdots(j+k-2)(j+k-1)1$ (symbol sequence of length $k+1$). One has

$$\mu(A_{j,k}) = \mu(A_j)p_jp_{j+1}\cdots p_{j+k-2}q_{j+k-1} = x_1P_j\frac{P_{j+k-1}}{P_j}q_{j+k-1} = x_1P_{j+k-1}q_{j+k-1}$$

as $\mu(A_j) = x_j = x_1P_j$. The integral of τ_U over the entire space is

$$\int_\Omega \tau_U\, d\mu = \sum_{j,k} k\mu(A_{j,k}) = \sum_{j,k} kx_1P_{j+k-1}q_{j+k-1}.$$

If we choose $p_i = \left(\frac{i}{i+1}\right)^\alpha$ for some $\alpha \in (1,2)$ then $P_j = \prod_{i=1}^{j-1}\left(\frac{i}{i+1}\right)^\alpha = \frac{1}{j^\alpha}$ and since the P_j are summable, $x_1 = \left(\sum_j P_j\right)^{-1}$ is well defined and positive. Then

$$\begin{aligned}
\int_\Omega \tau_U\, d\mu &= x_1\sum_k k\sum_j \frac{1}{(j+k-1)^\alpha}q_{j+k-1}\\
&\geq c_1x_1\sum_k k\sum_j \frac{1}{(j+k-1)^{\alpha+1}}\\
&\geq c_2\sum_k \frac{k}{k^\alpha} = \infty,
\end{aligned}$$

as $\alpha < 2$, where we used that $q_{j+k-1} = 1 - \left(1 - \frac{1}{j+k-1}\right)^\alpha \geq c_1\frac{1}{j+k-1}$ for some $c_1 > 0$. We thus see that the integral of τ_U over the entire space Ω diverges.

This can be converted to an example on a two-state shiftspace $\Sigma \subset \{0,1\}^{\mathbb{Z}}$ by the single element mapping $\pi : \Omega \to \Sigma$ which maps $\pi(1) = 1$ and collapses all other symbols to 0, i.e. $\pi(j) = 0, j = 2, 3, \dots$. The measure μ is sent to the probability measure $\nu = \pi^*\mu$ which is invariant under the shift map.

In fact we have $\int_\Omega \tau_U\, d\mu$ is finite if and only if $\int_U \tau_U^2\, d\mu$ is finite. So the above example is an example where the return time to U is not square integrable over U. For this consider the following lemma.

Lemma 3. *Let μ be T-invariant and $U \subset \Omega$, $\mu(U) > 0$. Then*

$$\mu(\{x \in U : \tau_U(x) \geq n\}) = \mu(\{x \in \Omega : \tau_U(x) = n\}).$$

Proof. Put $V_j^k = \bigcap_{i=j}^k T^{-i}U^c$ for the points x for which $T^ix \notin U$ for $i = j, \dots, k$. Clearly $V_{j'}^{k'} \subset V_j^k$ if $j' \leq j$ and $k' \geq k$. In particular $\{x \in \Omega : \tau_U(x) > n\} = V_1^n$ and $\{x \in U : \tau_U(x) \geq n\} = U \cap V_1^{n-1}$. Also

$$\{x \in \Omega : \tau_U(x) = n\} = V_1^{n-1} \setminus V_1^n$$

as $V_1^n \subset V_1^{n-1}$ and

$$T^{-1}V_0^{n-2} = V_1^{n-1} = (U \cap T^{-1}V_0^{n-2}) \cup (U^c \cap T^{-1}V_0^{n-2}) = (U \cap V_1^{n-1}) \cup V_0^{n-1}$$

(disjoint union) as $T^{-1}V_0^{n-1} = V_1^n$. Hence

$$\begin{aligned}\mu(\{x\in U:\tau_U(x)\geq n\}) &= U\cap V_1^{n-1}\\ &= \mu(T^{-1}V_0^{n-2})-\mu(V_0^{n-1})\\ &= \mu(V_1^{n-1}\setminus V_1^n)\\ &= \mu(\{x\in\Omega:\tau_U(x)=n\}).\end{aligned}$$

∎

Now Put $A_n = \{x\in\Omega:\tau_U(x)=n\}$. Then $U_n = U\cap A_n$ and by the previous lemma $\mu(A_n)=\sum_{j=n}^\infty \mu(U_j)$. Hence

$$\int_\Omega \tau_U\,d\mu = \sum_n n\mu(A_n) = \sum_{n=1}^\infty\sum_{j=n}^\infty n\mu(U_j) = \sum_{j=1}^\infty\sum_{n=1}^j n\mu(U_j) = \sum_j \frac{j(j+1)}{2}\mu(U_j)$$

and thus $\int_\Omega \tau_U\,d\mu = \frac12\int_U \tau_U(\tau_U+1)\,d\mu = \frac12\left(\int_U\tau_U^2\,d\mu+1\right)$ (the last identity if μ is ergodic). {entry.ret

3.3. **Entry and return times distributions.** Let $B\subset\Omega$ ($\mu(B)>0$) and put for (parameter values) $t>0$

$$F_B(t) = \mathbb{P}\left(\tau_B > \frac{t}{\mu(B)}\right) = \mu\left(\left\{x\in\Omega:\tau_B(x)>\frac{t}{\mu(B)}\right\}\right)$$

for the entry time distribution to B. The entry times distribution $F_B(t)$ is locally constant on intervals of length $\mu(B)$ and has jump discontinuities at values t which are integer multiples of $\mu(B)$. For any $s\in\mathbb{N}_0$ one has

$$\{\tau_B > s+1\} = T^{-1}\{\tau_B>s\}\setminus T^{-1}B$$

and consequently

$$\mathbb{P}(\tau_B = s+1) = \mathbb{P}(\tau_B>s)-\mathbb{P}(\tau_B>s+1)\leq\mu(B)$$

which shows that the jumps at the discontinuities are at most $\mu(B)$.

If $B_n\subset\Omega$ ($\mu(B_n)>0$) is a sequence of subsets so that $\mu(B_n)\to0^+$ as $n\to\infty$, then we would like to know what happens to $F_{B_n}(t)$ as $n\to\infty$. If the F_{B_n} converge weakly to a limiting distribution $F(T)$ (that is pointwise at all points of continuity), then we see that$F(t)$ is Lipschitz continuous with Lipschitz constant 1 and in particular continuous.

In a similarly way we can define the return times distribution by putting

$$\hat{F}_B(t) = \mathbb{P}_B\left(\tau_B>\frac{t}{\mu(B)}\right) = \hat{\mu}\left(\left\{x\in\Omega:\tau_B(x)>\frac{t}{\mu(B)}\right\}\right),$$

where $\hat{\mu}$ is the induced measure on B. This function too is constant except for jump discontinuities at multiples of $\mu(B)$. By Kac's theorem we have in this case

$$\int_0^\infty \hat{F}_B(t)\,dt = \sum_{j=1}^\infty \mu(B)\hat{\mu}(\{x:\tau_B(x)\geq j\}) = \sum_{j=1}^\infty \mu(\{x\in B:\tau_B(x)\geq j\}) = \int_B\tau_B\,d\mu = 1.$$

However, if for a sequence of sets B_n, $\mu(B_n)\to0$, the functions $\hat{F}_{B_n}$ converge to a limiting distribution $\hat{F}$ then $\int_0^\infty\hat{F}(t)\,dt = 1$ if the the sequence $\hat{F}_{B_n}$ is tight. In general we can only say that $\int_0^\infty\hat{F}(t)\,dt\leq1$

Lacroix has shown in 2002 that if $\hat{F}(t)$ is an eligible limiting distribution, that is it satisfies $F(0) = 1$, is right continuous, convex, monotonically decreasing on $(0, \infty)$ and $F(t) \to 0^+$ as $t \to \infty$, then for any ergodic T-invariant probability measure μ there exists a sequence of positive measure sets $B_n \subset \Omega$ so that $\mu(B_n) \to 0$ and such that $F(t) = \lim_{n\to\infty} F_{B_n}(t)$ for every point of continuity $t \in (0, \infty)$. Of course, the sets B_n are typically pretty wild looking and in particular they are not topological balls or cylinder sets (if there is a given partition).

Theorem 4. [33, 32] *Let μ be ergodic. Then for every eligible limiting distribution $\hat{F}$ ($\int_0^\infty \hat{F}(t)\,dt = 1$) there exists a sequence of sets $U_m \subset \Omega$, $\mu(U_m) > 0$, so that $\bigcap_m U_m = \{x\}$ and $\hat{F}_{U_m}$ converges to $\hat{F}$.*

Proof. We proceed in two steps. In the first step one constructs the 'stamp' for rational distribution functions and in the second step $\hat{F}$ will be approximated by rational distributions.

(I) A distribution function $G(t)$, $t \geq 0$, is rational if it is piecewise constant with finitely many discontinuities $0 < t_1 < t_2 < \cdots < t_k$, $t_j \in \mathbb{N}$, $\beta_j = \frac{p_j}{q}$, where $\beta_j = G(t_j^-) - G(t_j^+)$, and where the return set B has measure $\alpha = \frac{p}{q}$. One clearly has $\sum_j \beta_j = 1$ which implies $\sum_j p_j = q$. One now constructs a periodic system with invariant probability measure μ in which the subset B has measure α and return times distribution $\mu(\{x \in B : \tau_B(x) = \ell\})$ is equal to $\alpha\beta_j$ is $\ell = t_j$ and equal to 0 otherwise. Notice that by Kac's theorem $\sum_j t_j \mu(\{x \in B : \tau_B(x) = t_j\}) = 1$

Let m be some positive integer and put $H = Mq^2$ and $n_j = Mpp_j$, $j = 1, 2, \ldots, k$. Notice that $\sum_j n_j = Mp\sum_j p_j = Mpq$. The unit interval B will be divided into Mpq equal intervals which are grouped into blocks of $n_1, n_2, \ldots, n_k$ subintervals. On the jth block of n_j subintervals we construct over each subinterval a tower with t_j floors. The map T is then given by mapping one floor to the next floor above and mapping the top most floor to the bottom most floor (which is an interval of B) of the next tower. The invariant measure μ is then equal distribution.

By Kac's theorem $1 = \sum_j t_j \alpha\beta_j = \sum_j t_j \frac{p}{q}\frac{p_j}{q}$ implies $\sum_j t_j p_j = \frac{q^2}{p}$. The total height of the tower construction is then $\sum_j t_j n_j = Mp\sum_j t_j p_j = Mq^2 = H$ and the measure of the return set B is thus $\mu(B) = \frac{Mpq}{H} = \frac{p}{q} = \alpha$.

For the return times we obtain $\mu(\{x \in B : \tau_B(x) = t_j\}) = \frac{1}{H} n_j = \frac{Mpp_j}{Mq^2} = \frac{pp_j}{q^2} = \alpha\beta_j$, or $\hat{\mu}(\tau_B = t_j) = \beta_j$ where $\hat{\mu}$ is the induced measure on B. All other return times have zero measure. Hence, the distribution function of the system is G.

(II) For almost every$x \in \Omega$ we can find neighbourhoods $B_n \in \Omega$, $\mu(B_n) \to 0$, $\{x\} = \bigcap_n B_n$ and so that $\tau(B_n) = \inf_{y \in B_n} \tau_{B_n}(y) \to \infty$ as $n \to \infty$. Let $\hat{F}$ be a given distribution. Let $\alpha_m \to 0$ be rational G_m rational distribution functions as in part (I) with discontinuities $t_1 < t_2 < \cdots < t_k$ (dropping the index m) and jumps β_j, $j = 1, \ldots, k$, $\sum_j \beta_j$, such that $G_m(\alpha_m t) \to \hat{F}(t)$ at points of continuity. By erodicity Ω is described by the Rokhlin tower $\Delta = \bigcup_{\ell=1}^\infty \bigcup_{j=0}^{\ell-1} T^j B_{n,\ell}$, where $B_{n,\ell}$ are the level sets $\{y \in B_n : \tau_{B_n}(y) = \ell\}$. Choose an integer M and construct a stamp S_m for G_m as described in part (I). If $n = n(m)$ is large enough, then $\tau(B_n)$ is much larger than $L_m = Mq^2$. We label the floors in the Rokhlin tower according to repeated application of the stamp over the level sets $B_{n,\ell}$ and say that $T^j B_{n,\ell}$ belongs to U_m whenever we hit one of the Mpq base subintervals of the stamp. That is $T^j B_{n,\ell} \subset U_m$ if

$$j \in \{0, t_1, 2t_1, \cdots, n_1 t_1, n_1 t_1 + t_2, \ldots, L_m - t_k\} \mod L_m.$$

Let $\varepsilon > 0$ then if $\tau(B_n)$ is large enough we achieve that $|\mu(U_m) - \alpha_m| < \varepsilon\mu(U_m)$ and $\mu(y \in U_n : \tau_{U_m}(y) \notin \{t_1, t_2, \ldots, t_k\}) < \varepsilon\mu(U_m)$. This implies that $\hat{F}_{U_m}(t) - \hat{F}(t) + \mathcal{O}(\varepsilon)$. ∎

3.4. **Relation between entry and return times distributions.** The restriction of the function τ_B to the set $B \subset \Omega$ is called the *return time function* and we correspondingly call

$$\tilde{F}_B(t) = \mathbb{P}_B\left(\tau_B > \frac{t}{\mu(B)}\right)$$

the *return times distribution.* For instance, if Ω is the shiftspace Σ and $B = U(x_0 x_1 \cdots x_{n-1})$ is an n-cylinder then $\tau_B(\vec{x})$ for $\vec{x} \in B$ measures the 'time' it take to see the word $x_0 x_1 \cdots x_{n-1}$ again, that is

$$\tau_B(\vec{x}) = \min\{j \geq 1 : x_j x_{j+1} \cdots x_{j+n-1} = x_0 x_1 \cdots x_{n-1}\}.$$

The function $\tilde{F}_B(t)$ then measures the probability to see the first n-word again after rescaled time $t/\mu(B)$.

The following theorem relates the limiting entry times distribution to the limiting return times distribution. It turns out that a simple formula allows us to compute one from the other one.

{HLV}

Theorem 5. [18] *Let $B_n \subset \Omega$ ($\mu(B_n) > 0$) be a sequence of sets so that $\mu(B_n) \to 0^+$. If the limits $F(t) = \lim_{n\to\infty} F_{B_n}(t)$, $\tilde{F}(t) = \lim_{n\to\infty} \tilde{F}_{B_n}(t)$ exist (pointwise) then*

$$F(t) = \int_t^\infty \tilde{F}(s)\, ds.$$

Observe that the limiting entry times distribution and return times distribution are the same only if they are exponential, that is $\tilde{F} = F$ if only if $F(t) = \tilde{F}(t) = e^{-t}$. Also note, that in conjunction with the previous theorem, we conclude that the limiting return times distribution of the restricted system $(U, \hat{T}, \hat{\mu})$ (for some positive measure $U \subset \Omega$) is the same as the limiting return times distribution of the entire system (Ω, T, μ).

We will first prove the following lemma from which the theorem follows by taking limits.

Lemma 6. *For any $B \subset \Omega$ ($\mu(B) > 0$) let F_B' be the right sided derivative of the largest continuous piecewise linear function that lies below F_B and is linear on the intervals $[k\mu(B), (k+1)\mu(B)], k \in \mathbb{N}$. Then*

$$-F_B'(t) = \tilde{F}(t).$$

Proof. We have that $F_B(t) = \mathbb{P}(\tau_B > t/\mu(B))$, $\tilde{F}_B(t) = \mathbb{P}_B(\tau_B > t/\mu(B))$. By a previous lemma we have that

$$\mu\left(\{x \in B : \tau_B(x) \geq s\}\right) = \mathbb{P}(\tau_B = s).$$

If we put $s=\frac{t}{\mu(B)}$, then we can also write

$$\begin{aligned}\mathbb{P}(\tau_B=s) &= \mathbb{P}(\tau_B>s-1)-\mathbb{P}(\tau_B>s)\\ &= \mathbb{P}\left(\tau_B>\frac{t}{\mu(B)}-1\right)-\mathbb{P}\left(\tau_B>\frac{t}{\mu(B)}\right)\\ &= F_B(t-\mu(B))-F_B(t)\\ &= -\mu(B)F_B'(t-\mu(B)).\end{aligned}$$

Combining this with the previous identity for $\mathbb{P}(\tau_B=s)$ yields

$$\begin{aligned}-F_B'(t-\mu(B)) &= \frac{\mu(\{x\in B:\tau_B>(t-\mu(B))/\mu(B)\})}{\mu(B)}\\ &= \mathbb{P}_B\left(\tau_B>\frac{t-\mu(B)}{\mu(B)}\right)\\ &= \tilde{F}_B(t-\mu(B)).\end{aligned}$$

∎

Proof of the theorem. We apply the lemma to the sets B_n. Integration yields

$$F_{B_n}(t)=\int_t^\infty -F_{B_n}'(s)\,ds=\int_t^\infty \tilde{F}_{B_n}(s-\mu(B_n))\,ds$$

as $F_B(t),\tilde{F}_B(t)\to 0$ as $t\to\infty$ (Poincaré and ergodicity). Now let n go to infinity ($\mu(B_n)\to 0^+$). ∎

As noted above the limiting entry distribution F is Lipschitz continuous. Since the limiting return distribution $\tilde{F}(t)$ is monotonically decreasing to zero, we now easily see that $F(t)$ is in fact always convex.

3.5. **Entry and return times for the induced map.** In this section we show that the limiting distributions for the map are the same as for the induced map. We first prove the following useful identity:

Lemma 7. *Assume μ is ergodic. Let $U\subset\Omega$, $\mu(U)>0$ and put $V_k=\{x\in U:\tau_U(x)\geq k\}$. Then for every $A\subset\Omega$ one has*

$$\mu(A)=\sum_{k=1}^\infty \mu(T^{-k}A\cap V_k).$$

Proof. Put

$$A_j^k=T^{-k}A\setminus\bigcup_{i=j}^{k-1}T^{-i}U=\{x\in T^{-k}A:T^ix\notin U, i=j,\dots,k-1\}.$$

Then $A_0^0=A$, $A_0^1=T^{-1}A\setminus U$, $A_1^k=T^{-1}A_0^{k-1}$ and $T^{-k}A\cap V_k=U\cap A_1^k=A_1^k\setminus A_0^k$. Hence

$$\sum_{k=1}^N\mu(T^{-k}A\cap V_k)=\sum_{k=1}^N\left(\mu(A_0^{k-1})-\mu(A_0^k)\right)=\mu(A_0^0)-\mu(A_0^N)=\mu(A)-\mu(A_0^N).$$

To show that the last term goes to zero as $N\to\infty$ we use that $\mu(A_0^N)=\mu(A_1^{N+1})$ and estimate as follows:

$$\begin{aligned}
\mu(A_0^N) &= \sum_{\ell=N+1}^\infty \mu(x\in\Omega:\tau_U(x)=\ell)\\
&= \sum_{\ell=N+1}^\infty \mu(x\in U:\tau_U(x)\ge\ell)\\
&\le \sum_{\ell=N+1}^\infty \ell\mu(x\in U:\tau_U(x)=\ell)\\
&= \int_U \tau_U\chi_{\tau_U>N}\,d\mu
\end{aligned}$$

which decays to 0 as $N\to\infty$ since τ_U is integrable over U by Kac's theorem. ∎

In the same way one can also show that $\mu(A\setminus U)=\sum_{k=1}^\infty\mu(T^{-k}A\cap V_{k+1})$. The assumption of ergodicity of μ can be replaced with the assumption that $\Omega=\bigcup_n T^{-n}U$ (up to nullsets).

The following theorem shows that a restricted system $(U,\hat T,\hat\mu)$ has the same limiting entry times distribution as the original system (Ω,T,μ).

Theorem 8. [17, 21] *Let μ be ergodic, $U\subset\Omega$, $\mu(U)>0$. Assume there exists a sequence of sets $B_n\subset U$, $\mu(B_n)\to 0^+$, so that*

$$\begin{aligned}
F(t) &= \lim_{n\to\infty}F_{B_n}(t), &\qquad F_{B_n}(t)=\mathbb{P}\left(\tau_{B_n}>\frac{t}{\mu(B_n)}\right)\\
\hat F(t) &= \lim_{n\to\infty}\hat F_{B_n}(t), &\qquad \hat F_{B_n}(t)=\mathbb{P}\left(\hat\tau_{B_n}>\frac{t}{\hat\mu(B_n)}\right)
\end{aligned}$$

where

$$\tau_B(x)>\min\{j\ge1:T^jx\in B\},\quad \hat\tau_B(x)>\min\{j\ge1:\hat T^jx\in B\}$$

and $\hat T=T^{\tau_U}$ is the induced transformation on U.

Then $F(t)=\hat F(t)$ for all $t\in\mathbb{R}^+$.

For ergodic Radon measures μ this was proven in [11] in 2003 where the Lebesgue Density theorem[1] was used. The limit there was along metric balls B_n that shrink to a point $x\in\Omega$.

Proof. We first relate τ_B to $\hat\tau_B$ ($B\subset U,\mu(B)>0$). If we put $m=\hat\tau_B(x)$, $x\in U$, then

$$\tau_B(x)=\tau_U(x)+\tau_U(\hat Tx)+\tau_U(\hat T^2x)+\cdots+\tau_U(\hat T^{m-1}x)=n^m(x),$$

where we wrote the ergodic sum of the function $n=\tau_U|_U$ for the return time on $(U,\hat T)$. By the Birkhoff ergodic theorem on $(U,\hat T,\hat\mu)$ we get as $\hat\mu$ is ergodic:

$$\frac1m\tau_B(x)=\frac1m n^m(x)\to\int_U n(x)\,d\mu(x)=\int_U\tau_U(x)\,\frac{d\mu(x)}{\mu(U)}=\frac{1}{\mu(U)}$$

by Kac's theorem for almost every $x\in U$.

[1]It says that if G is a full measure set then $\frac{\mu(G\cap B_r(x))}{\mu(B_r(x))}\to1$ almost surely as the radius $r\to0$.

Let $\varepsilon > 0$, then there exists $G_\varepsilon \subset U$, and $M_\varepsilon \in \mathbb{N}$ so that

$$\left|\frac{1}{m}n^m(x) - \frac{1}{\mu(U)}\right| < \varepsilon \qquad \forall\, x \in G_\varepsilon,\ m \geq M_\varepsilon$$

and $\mu(G_\varepsilon^c) < \varepsilon$. Thus

$$\tau_B(x) = \sum_{j=0}^{\hat\tau_B(x)-1} \tau_U \circ \hat T^j = \frac{\hat\tau_B(x)}{\mu(U)} + \mathcal{O}(\hat\tau_B(x)\varepsilon)$$

for all $x \in G_\varepsilon$ such that $\hat\tau_B(x) \geq M_\varepsilon$. Since τ_U is integrable on U there exists a $\delta > 0$ (depending on ε) so that $\int_S \tau_U\, d\mu < \varepsilon$ for any set $S \subset U$ for which $\mu(S) < \delta$. We can assume that $\mu(G_\varepsilon^c) < \min(\delta, \varepsilon)$.

With $V_n = \bigcup_{j\geq n} U_j$, and $U_j = \{x \in U : \tau_U = j\}$ we obtain by the previous lemma

$$F_B(t) = \int_\Omega \chi_{\tau_B > s}\, d\mu = \sum_n \int_{V_n} \chi_{\tau_B > s} \circ T^n\, d\mu = \sum_{n \leq M_\varepsilon} \int_{V_n} \chi_{\tau_B > s} \circ T^n\, d\mu + E_1,$$

where $s = \frac{t}{\mu(B)}$. The error E_1 is by assumption bounded by

$$|E_1| \leq \sum_{n > M\varepsilon} \mu(V_n) \leq \int_U \tau_U \chi_{\tau_U > M_\varepsilon}\, d\mu < \varepsilon.$$

Hence

$$F_B(t) = \sum_{n \leq M_\varepsilon} \int_{V_n} \chi_{\tau_B > s}\, d\mu + E_1 + E_2,$$

where

$$\begin{aligned}
|E_2| &\leq \sum_{n \leq M_\varepsilon} \int_{V_n} |\chi_{\tau_B > s} \circ T^n - \chi_{\tau_B > s}|\, d\mu \\
&\leq \sum_{n \leq M_\varepsilon} \int_{V_n} \sum_{j=0}^{n} \left(\chi_B \circ T^j + \chi_B \circ T^{s+j}\right)\, d\mu \\
&\leq \sum_{n \leq M_\varepsilon} n\mu(B) \\
&\leq M_\varepsilon^2 \mu(B)
\end{aligned}$$

and therefore $|E_2| < \varepsilon$ if $\mu(B)$ is small enough. Consequently

$$F_B(t) = \int_U \tau_U \chi_{\tau_B > s}\, d\mu + E_1 + E_2 + E_3,$$

as $\sum_n \chi_{V_n} = \tau_U$, where

$$|E_3| \leq \sum_{n > M_\varepsilon} \int_{V_n} \chi_{\tau_B > s}\, d\mu \leq \sum_{n > M_\varepsilon} \mu(V_n) = \int_U \tau_U \chi_{\tau_U > M_\varepsilon}\, d\mu < \varepsilon$$

by choice of M_ε.

Next we replace τ_B by $\hat{\tau}_B$ using the relation $\frac{\tau_B}{\hat{\tau}_B} = \frac{1}{\mu(U)} + \mathcal{O}(\varepsilon)$ which implies $\tau_B = \frac{\hat{\tau}_B}{\mu(U)} + \eta$ where $\eta : U \to \mathbb{R}$ has the bound $|\eta| \le \varepsilon\mu(U)\hat{\tau}_B \le \varepsilon\hat{\tau}_B$. Hence

$$F_B(t) = \int_U \tau_U \chi_{\hat{\tau}_B > \frac{t}{\hat{\mu}(B)} + \eta} \, d\mu + E_1 + E_2 + E_3.$$

Now we want to introduce a power k of the induced map $\hat{T}$ so that we can average over k and use the ergodic theorem on $(U, \hat{T}, \hat{\mu})$. By $\hat{T}$-invariance of $\hat{\mu}$

$$\begin{aligned} F_B(t) &= \int_U \left(\tau_U \chi_{\hat{\tau}_B > \frac{t}{\hat{\mu}(B)} + \eta}\right) \circ \hat{T}^k \, d\mu + E_1 + E_2 + E_3 \\ &= \int_{G_\varepsilon} \left(\tau_U \chi_{\hat{\tau}_B > \frac{t}{\hat{\mu}(B)} + \eta}\right) \circ \hat{T}^k \, d\mu + E_1 + E_2 + E_3 + H_k, \end{aligned}$$

where we get for the error

$$H_k = \int_{G_\varepsilon^c} \left(\tau_U \chi_{\tau_B > \frac{t}{\mu(B)}}\right) \circ \hat{T}^k \, d\mu \le \int_{G_\varepsilon^c} \tau_U \circ \hat{T}^k \, d\mu = \int_{\hat{T}^{-k} G_\varepsilon^c} \tau_U \, d\mu < \varepsilon$$

since by assumption $\mu(\hat{T}^{-k} G_\varepsilon^c) = \mu(G_\varepsilon^c) < \delta$ as μ restricted to U is $\hat{T}$-invariant.

In the principal term we want to exploit the identity $\sum_j \chi_{V_j} = \tau_U$ on U. For that purpose note that

$$\left\{x \in U : \hat{\tau}_B(\hat{T}^k x) \ge s\right\} \setminus \bigcup_{\ell=1}^{k-1} \hat{T}^{-\ell} B = \{x \in U : \hat{\tau}_B(x) \ge s + k\}$$

which yields (here we use $s = \frac{t}{\hat{\mu}(B)} + \eta$)

$$F_B(t) = \int_{G_\varepsilon} \left(\tau_U \chi_{\hat{\tau}_B > \frac{t}{\hat{\mu}(B)} + \eta + k}\right) \circ \hat{T}^k \, d\mu + E_1 + E_2 + E_3 + H_k + K_k,$$

where the individual errors are bounded by:

$$K_k \le \int_{G_\varepsilon} \tau_U \circ \hat{T}^k \sum_{\ell=1}^{k-1} \chi_B \circ \hat{T}^\ell \, d\mu.$$

We now estimate the average error over $k \in \{0, 1, \dots, n-1\}$:

$$\begin{aligned} \hat{K}_n &= \frac{1}{n} \sum_{k=0}^{n-1} K_k \\ &= \int_{G_\varepsilon} \frac{1}{n} \sum_{k=0}^{n-1} \sum_{\ell=1}^{k-1} (\tau_U \circ \hat{T}^k)(\chi_B \circ \hat{T}^\ell) \, d\mu \\ &= \int_{G_\varepsilon} \sum_{\ell=1}^{n-2} (\chi_B \circ \hat{T}^\ell) \left(\frac{1}{n} \sum_{k=\ell+1}^{n-1} \tau_U \circ \hat{T}^k\right) d\mu \\ &\le c_1 \frac{1}{\mu(U)} \int_{G_\varepsilon} \sum_{\ell=1}^{n-2} (\chi_B \circ \hat{T}^\ell) \, d\mu \\ &\le c_1 n \hat{\mu}(B) \end{aligned}$$

where we used the estimate

$$\frac{1}{n}\sum_{k=\ell+1}^{n-1}\tau_U\circ\hat{T}^k \le \frac{1}{n}\sum_{k=0}^{n-1}\tau_U\circ\hat{T}^k \le \frac{1}{\mu(U)}+\varepsilon \le c_1\frac{1}{\mu(U)}$$

for some constant c_1 and for all $x\in G_\varepsilon$ provided $n\ge M_\varepsilon$. Thus

$$F_B(t)=\frac{1}{n}\sum_{k=1}^{n}\int_{G_\varepsilon}\left(\tau_U\circ\hat{T}^k\right)\chi_{\hat{\tau}_B>\frac{1}{1-\eta'}\frac{t+k\hat{\mu}(B)}{\hat{\mu}(B)}}\,d\mu+E_0+\hat{H}_n+\hat{K}_n,$$

where $|E_0|<3\varepsilon$, $\hat{H}_n=\frac{1}{n}\sum_{k=0}^{n-1}H_k<\varepsilon$ and $\eta'':U\to\mathbb{R}$ satisfies the bound $|\eta''|<\varepsilon$. Consequently (as $|E_0+\hat{K}_n|<3\varepsilon+c_1n\hat{\mu}(B)$)

$$\begin{aligned}F_B(t) &= \int_{G_\varepsilon}\frac{1}{n}\sum_{k=1}^{n}\left(\tau_U\circ\hat{T}^k\right)\chi_{\hat{\tau}_B>(1+\eta'')\frac{t}{\hat{\mu}(B)}}\,d\mu+\mathcal{O}(\varepsilon+n\hat{\mu}(B))\\ &= \int_{G_\varepsilon}\chi_{\hat{\tau}_B>(1+\eta'')\frac{t}{\hat{\mu}(B)}}\,d\hat{\mu}+\mathcal{O}(\varepsilon+n\hat{\mu}(B))\end{aligned}$$

as $\frac{1}{n}\sum_{k=1}^{n}\tau_U\circ\hat{T}^k=\frac{1}{\mu(U)}+\mathcal{O}(\varepsilon)$ on G_ε, where $\eta'':U\to\mathbb{R}$ satisfies $|\eta''|<c_2|\eta'|+\frac{n}{t}\hat{\mu}(B)$ ($c_2>0$). To adjust for the 'time shift' in the lower bound of the entry function, we use the fact that $|\hat{F}_B(t)-\hat{F}(s)|\le|t-s|+\hat{\mu}(B)$ and thus obtain (for a c_3)

$$|F_B(t)-\hat{F}_B(t)|<c_3\varepsilon+\left(\frac{n}{t}+1\right)\hat{\mu}(B)+c_1n\hat{\mu}(B)$$

for all $n\ge M_\varepsilon$. If $\mu(B)$ is small enough so that $\hat{\mu}(B)<\min(\frac{\varepsilon t}{M_\varepsilon},\frac{\varepsilon}{(c_1+1)n})$, then

$$|F_B(t)-\hat{F}_B(t)|<(c_3+1)\varepsilon$$

and as $\mu(B)\to0$ we obtain $|F(t)-\hat{F}(t)|<(c_3+1)\varepsilon$ for any positive ε. Thus $F_B(t)=\hat{F}_B(t)$ for all $t>0$. ∎

g.entry.times}

3.6. **Limiting entry times distributions.**

Definition 9. *(I) We say μ is* mixing *if $\forall U,V\subset\Omega$:*

$$\mu(U\cap T^{-j}V)\to\mu(U)\mu(V)$$

as $n\to\infty$.
(II) We say μ is weakly mixing *if $\forall U,V\subset\Omega$:*

$$\frac{1}{n}\sum_{j=0}^{n-1}\left|\mu(U\cap T^{-j}V)-\mu(U)\mu(V)\right|\to0$$

as $n\to\infty$.

Mixing clearly implies the weak mixing property and both imply ergodicity.

Example: The (irrational) rotation $R_\alpha:[0,1)\circlearrowleft$ given by $R_\alpha x=x+\alpha \mod 1$ is not mixing (nor is it weakly mixing) because there exists a sequence $n_j\to\infty$ so that $R_\alpha 0\to0$ as $j\to\infty$. For instance one can take $n_j=q_j$ where $\frac{p_j}{q_j}$ are the approximats of α. They satisfy $\left|\alpha - \frac{p_j}{q_j}\right| \le \frac{c_1}{q_j}$ for some constant c_1. Then $n_j\alpha = q_j\alpha = p_j + \mathcal{O}(q_j^{-1}) = \mathcal{O}(q_j^{-1})$ mod 1. Then with $U = V = [0, \frac{1}{2}]$, say, one gets $\mu(U \cap R_\alpha^{n-j}V) \to \frac{1}{2} \ne \frac{1}{4} = \mu(U)\mu(V)$. ∎

To obtain meaningful results for the limiting distribution of entry and return times we have to make some assumptions, in particular on the rates of mixing. Let $\mathcal{A} = \{A_1, A_2, \dots, A_M\}$ be a finite measurable partition of Ω, that is $\Omega = \bigcup_j A_j$ and $A_i \cap A_j = \varnothing$ for $i \ne j$. Denote by

$$\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j}\mathcal{A}$$

its *nth joint*. Here we use the notation that for two partitions $\mathcal{A}$ and $\mathcal{B}$ one gets the finer partition $\mathcal{A} \vee \mathcal{B} = \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$. Hence $A \in \mathcal{A}^n$ if $\exists A_{i_j} \in \mathcal{A}$ for $j = 0, 1, \dots, n-1$ so that $A = \bigcap_{j=0}^{n-1} T^{-j}A_{i_j}$, or, in other words, $T^jA \subset A_{i_j}$ for $j = 0, 1, \dots, n-1$.

Example: (Shift space) For the full M shift space $\Sigma = \{1, \dots, M\}^{\mathbb{N}_0}$ with the left shift map $\sigma : \Sigma \circlearrowleft$ one sets $\mathcal{A} = \{U(i) : i = 1, \dots, M\}$, where $U(i) = \{\vec{x} \in \Sigma : x_o = i\}$ are 1-cylinder sets. Then $\mathcal{A}^n = \bigcup_{x_0x_1\cdots x_{n-1}}\{U(x_0x_1\cdots x_{n-1})\}$ where $U(x_0\cdots x_{n-1}) = \{\vec{y} : y_0\cdots y_{n-1} = x_0\cdots x_{n-1}\}$ are n-cylinder sets.

We say $\mathcal{A}$ is *generating* if $\mathcal{A}^\infty$ consists of singletons only containing single points.

Let μ be a T-invariant measure and $\mathcal{A}$ a finite generating partition. Put $\mathcal{A}^* = \bigcup_{n\ge1}\mathcal{A}^n$. Then we say:

(I) μ is *ψ-mixing* if $\forall U \in \sigma(\mathcal{A}^n), V \in \sigma(\mathcal{A}^n)$:

$$\left|\frac{\mu(U \cap T^{-n-k}V)}{\mu(U)\mu(V)} - 1\right| \le \psi(k) \searrow 0.$$

(II) μ is *right ϕ-mixing* if $\forall U \in \sigma(\mathcal{A}^n), V \in \sigma(\mathcal{A}^n)$:

$$\left|\frac{\mu(U \cap T^{-n-k}V)}{\mu(V)} - \mu(U)\right| \le \phi(k) \searrow 0.$$

One similarly calls an invariant measure μ *left ϕ-mixing* measure if $\forall U \in \sigma(\mathcal{A}^n), V \in \sigma(\mathcal{A}^n)$:

$$\left|\frac{\mu(U \cap T^{-n-k}V)}{\mu(U)} - \mu(V)\right| \le \phi(k) \searrow 0.$$

(III) μ is *α-mixing* if $\forall U \in \sigma(\mathcal{A}^n), V \in \sigma(\mathcal{A}^n)$:

$$\left|\mu(U \cap T^{-n-k}V) - \mu(U)\mu(V)\right| \le \alpha(k) \searrow 0.$$

Clearly ψ-mixing implies ϕ-mixing implies α-mixing.

Example: (Bernoulli measures) Let $\Sigma = \{1, \dots, M\}^{\mathbb{N}_0}$ and $\mathcal{A} = \{u(i) : i\}$ be as before. Then a probability vector $\vec{p} = (p_1, \dots, p_M)$ induces a shift invariant probability measure μ on Σ which is ψ-mixing with $\psi(k) = 0$ for all k. ∎

Example: (Markov measures) Let $\Sigma = \{1, \dots, M\}^{\mathbb{N}_0}$ and P an $M \times M$ stochastic matrix with left probability eigenvector $\vec{p}$. This induces a shift invariant probability measure μ on Σ which is ψ-mixing decaying exponentially fast. ∎

For $U\in\sigma(\mathcal{A}^n)$ and $s<<1/\mu(U)$ put $\lambda_s(U)=\frac{-\log\mathbb{P}(\tau_U>s)}{s\mu(U)}$.

{thm.Abadi}

Theorem 10. [1] *Let $T:\Omega\circlearrowleft$ and μ be a T-invariant probability measure. Assume there is a finite generating partition $\mathcal{A}$ of Ω so that μ is α-mixing with $\alpha(k)=\mathcal{O}(k^{-p})$ for some $p>1$. Then there exists a constant C so that*

$$\left|\mathbb{P}(\tau_{A_n(x)}>t)-e^{-\lambda_s(A_n(x))\mu(A_n(x))t}\right|\le C\mu(A_n(x))^\epsilon$$

for all $s<<t/\mu(A_n(x))$, where $A_n(x)$ denotes the unique partition element of $\mathcal{A}^n$ which contains the point x.

Proof. For $U\in\sigma(A^n)$ we notice that

$$\{\tau_U>s\}=\bigcap_{i=0}^{s}T^{-i}U^c\in\sigma(\mathcal{A}^{n+s}).$$

For any $\Delta>0$ (and $t>\Delta$)

$$\{\tau_U>s+t\}\subset\{\tau_U>s\}\cup T^{-t}\{\tau_U>t-\Delta\}\subset\{\tau_U>s+t\}\cup\bigcup_{i=s+1}^{s+\Delta}T^{-i}U$$

and

$$\{\tau_U>t\}\subset\{\tau_U>t-\Delta\}\subset\{\tau_U>t\}\cup\bigcup_{i=t-\Delta+1}^{t}T^{-i}U.$$

By the α-mixing property we thus obtain

$$\begin{aligned}
&|\mathbb{P}(\tau_U>s+t)-\mathbb{P}(\tau_U>s)\mathbb{P}(\tau_U>t)|\\
&\qquad\le \left|\mathbb{P}(\tau_U>s+t)-\mathbb{P}(\{\tau_U>s\}\cap T^{-t}\{\tau_U>t-\Delta\}\right|\\
&\qquad\quad+|\mathbb{P}(\tau_U>s+t)-\mathbb{P}(\tau_U>s)\mathbb{P}(\tau_U>t-\Delta)|\\
&\qquad\quad+\mathbb{P}(\tau_U>s)\,|\mathbb{P}(\tau_U<t)-\mathbb{P}(\tau_U>t-\Delta)|\\
&\qquad\le \Delta\mu(U)(1+\mathbb{P}(\tau_U>s))+\alpha(\Delta-n)\\
&\qquad\le 2\Delta\mu(U)+\alpha(\Delta-n).
\end{aligned}$$

Iterating this estimate yields

$$\begin{aligned}
&\left|\mathbb{P}(\tau_U>ks)-\mathbb{P}(\tau_U>s)^k\right|\\
&\qquad\le \sum_{j=1}^{k}|\mathbb{P}(\tau_U>js)-\mathbb{P}(\tau_U>(j-1)s)\mathbb{P}(\tau_U>s)|\,\mathbb{P}(\tau_U>s)^{k-j}\\
&\qquad\le (2\Delta\mu(U)+\alpha(\Delta-n))\sum_{j=1}^{k}\mathbb{P}(\tau_U>s)^{k-j}\\
&\qquad\le \frac{2\Delta\mu(U)+\alpha(\Delta-n)}{1-\mathbb{P}(\tau_U>s)}\\
&\qquad= \frac{2\Delta\mu(U)+\alpha(\Delta-n)}{\mathbb{P}(\tau_U\le s)}.
\end{aligned}$$

Put $\lambda_s(U) = \frac{-\log \mathbb{P}(\tau_U > s)}{s\mu(U)}$. For $t = ks + r$, $0 \le r < t$ the remainder, we get

$$\begin{aligned}\left|\mathbb{P}(\tau_U > t) - e^{-\lambda_s(U)\mu(U)t}\right| &\le |\mathbb{P}(\tau_U > t) - \mathbb{P}(\tau_U > ks)| + \left|\mathbb{P}(\tau_U > ks) - \mathbb{P}(\tau_U > s)^k\right| \\ &\quad + e^{-\lambda_s(U)\mu(U)ks}\left|1 - e^{-\lambda_s(U)\mu(U)r}\right| \\ &\le \Delta\mu(U) + \frac{2\Delta\mu(U) + \alpha(\Delta - n)}{\mathbb{P}(\tau_U \le s)} + c_1\Delta\mu(U)\end{aligned}$$

for some c_1. Now we choose s so that $\mathbb{P}(\tau_U \le s) \ge \mu(U)^\delta$ for some $\delta > 0$. Then we put $\Delta \sim \mu(U)^{-\beta}$ for some $\beta > 0$ which implies that $\alpha(\Delta - n) \sim \Delta^{-p} \sim \mu(U)^{\beta p}$. Hence

$$\left|\mathbb{P}(\tau_U > t) - e^{-\lambda_s(U)\mu(U)t}\right| \le c_2\mu(U)^{1-\beta} + \frac{\mu(U)^{1-\beta} + \mu(U)^{\beta p}}{\mu(U)^\delta} = \mathcal{O}(\mu(U)^\epsilon),$$

where $\epsilon = \min\{1-\beta, \beta p\} - \delta$ is positive if δ is small enough and $\beta < 1$. Now put $U = A_n(x)$. ∎

To better estimate λ_s we now assume that the measure μ is ϕ-mixing. But first we prove the following topological result.

For a set $U \subset \Omega$ put $\tau(U) = \inf_{y\in U} \tau_U(y)$ for the *period* of U. This is equivalent to $U \cap T^{-j}U = \varnothing$ for $j = 1, \ldots, \tau(U) - 1$ and $U \cap T^{-\tau(U)}U \ne \varnothing$.

Lemma 11. *Let $\mathcal{A}$ be a (finite) generating partition of Ω. Then the sequence $\tau(A_n(x))$, $n = 1, 2, \ldots$ is bounded if and only if x is a periodic point.*

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

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

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

Lemma 12. *Let μ be ϕ-mixing. Then there exists $\vartheta \in (0,1)$ and a constant C so that*

$$\mu(A) \le C\vartheta^n, \quad \forall\, A \in \mathcal{A}^n, \quad \forall\, n.$$

Proof. Since $\max_{A\in\mathcal{A}}\mu(A) < 1$ we can find k so that $\vartheta' = \max_{A\in\mathcal{A}}\mu(A) + \phi(k-1) < 1$. For $A \in \mathcal{A}^n$ there exists $A' \in \bigvee_{j=0}^{r-1} T^{-jk}\mathcal{A}$, where $r = [n/k]$, and so that $A \subset A'$. Clearly, there exist $B_j \in \mathcal{A}$, $j = 0, \ldots, r-1$, so that $A' = \bigcap_{j=0}^{r-1} T^{-jk}B_j$. By the ϕ-mixing property

$$\mu\left(\bigcap_{j=0}^{\ell} T^{-jk}B_j\right) \le \mu\left(\bigcap_{j=0}^{\ell-1} T^{-jk}B_j\right)(\mu(B_\ell) + \phi(k-1))$$

for $\ell = 1, \ldots, r-1$ and therefore

$$\mu(A) \le \mu(A') \le \prod_{j=0}^{r-1}(\mu(B_\ell) + \phi(k-1)) \le \vartheta'^r.$$

Now put $\vartheta=\sqrt[k]{\vartheta'}$. ∎

Lemma 13. *Let μ be ϕ-mixing with $\phi(k)=\mathcal{O}(k^{-p})$ for some $p>1$. Then for $U\in\sigma(\mathcal{A}^n)$:*

$$\frac{\mathbb{P}(\tau_U\le s)}{s\mu(U)}\to 1^-$$

as $s\mu(U)\to 0$ and $\tau(U)\to\infty$.

Proof. One has the simple upper bound $\mathbb{P}(\tau_U\le s)=\mu\Big(\bigcup_{j=1}^s T^{-j}U\Big)\le s\mu(U)$. In order to find a lower bound put $N_s=\sum_{j=1}^s\chi_U\circ T^j$ for the counting function for hitting U up to time s. Clearly $\{\tau_U\le s\}=\{N_s\ge 1\}$ and also $\mathbb{E}(N_s)=s\mu(U)$. By Cauchy-Schwarz

$$(\mu(N_s))^2=(\mu(N_s\chi_{N_s\ge 1}))^2\le\mu(N_s^2)\mu(\chi^2_{N_s\ge 1})=\mu(N_s^2)\mathbb{P}(\tau_U\le s).$$

Thus

$$\mathbb{P}(\tau_U\le s)\ge\frac{s^2\mu(U)^2}{\mu(N_s^2)},$$

where we can write

$$\mu(N_s^2)=\mu\left(\sum_{j=1}^s\chi_U\circ T^j\right)^2=s\mu(U)+2\sum_{j=1}^s(s-j)\mu(\chi_U(\chi_U\circ T^j)).$$

For $j=1,\dots,\tau(U)-1$ the terms in the sum are zero. For $j=\tau(U),\dots,2n$ we put $U^j\in\sigma(\mathcal{A}^{[j/2]})$ so that $U\subset U^j$. Then by the ϕ-mixing property

$$\mu(\chi_U(\chi_U\circ T^j))\le\mu(U^j\cap T^{-j}U)\le\mu(U)(\mu(U^j)+\phi(j/2)).$$

For $j=2n+1,\dots,s$ the ϕ-mixing property yields

$$\mu(\chi_U(\chi_U\circ T^j))=\mu(U\cap T^{-j}U)\le\mu(U)(\mu(U)+\phi(j-n))\le\mu(U)(\mu(U)+\phi(j/2)).$$

Hence

$$\begin{aligned}\mu(N_s^2)&\le s\mu(U)\left(1+2\sum_{j=\tau(U)}^{2n}\mu(U^j)+\sum_{j=2n+1}^{s}\mu(U)+\sum_{j=\tau(U)}^{\infty}\phi(j/2)\right)\\&\le s\mu(U)\left(1+c_1\sum_{j=\tau(U)}^{\infty}\vartheta^{j/2}+s\mu(U)+c_2\sum_{j=\tau(U)}^{\infty}(j/2)^{-p}\right)\\&\le s\mu(U)\big(1+c_3\vartheta^{\tau(U)/2}+s\mu(U)+c_4\tau(U)^{-p+1}\big)\end{aligned}$$

where we used the estimate $\mu(U^j)\le C\vartheta^{[j/2]}$ and that $\phi(k)\sim k^{-p}$. Consequently

$$\mathbb{P}(\tau_U\le s)\ \ge\ \frac{1}{1+c_3\vartheta^{\tau(U)/2}+s\mu(U)+c_4\tau(U)^{-p+1}}$$

and therefore $\mathbb{P}(\tau_U\le s)\to 1^-$ if we let $s\mu(U)\to 0$ and $\tau(U)\to\infty$. ∎

Theorem 14. *Let μ be ϕ-mixing and $\phi(k)=\mathcal{O}(k^{-p})$ for $p>1$. Then*

$$F_n(t)=\mathbb{P}\left(\tau_{A_n(x)}>\frac{t}{\mu(A_n(x))}\right)\to e^{-t}$$

as $n\to\infty$, for every non-periodic point $x\in\Omega$.

For ψ-mixing measures summability of $\psi(k)$ is enough.

In light of Theorem 5 we get the same result for the limiting return times distribution.

Corollary 15. *Let μ be a T-invariant probability measure. Assume there is a finite generating partition $\mathcal{A}$ of Ω so that μ is ϕ-mixing. Then*

$$\tilde{F}_n(t) = \mathbb{P}_{A_n(x)}\left(\tau_{A_n(x)} > \frac{t}{\mu(A_n(x))}\right) \to e^{-t}$$

for every non-periodic x.

4. Higher order returns

{higher.or

If individual returns are independent of one another then we expect that higher order returns are Poisson distributed. Here we will describe a method that allows us to prove precisely such results for α-mixing measures.

4.1. **Stein-Chen method.** Denote by ν_t the Poisson distribution on $\mathbb{N}_0$ with parameter $t > 0$, that is $\nu_t(\{k\}) = e^{-t}\frac{t^k}{k!}$. We will now describe the Stein-Chen method to estimate how close a given probability measure ν is to a Poisson distribution ν_t. Let $\mathcal{F} = \{f : \mathbb{N}_0 \to \mathbb{R}\}$ for the function space of functions on the non-negative integers. We then define the *Stein operator* $\mathscr{S} : \mathcal{F} \to \mathcal{F}$ by

$$\mathscr{S}f(k) = tf(k+1) - kf(k).$$

For given $h \in \mathcal{F}$ we want to find f so that

$$\mathscr{S}f = h - \nu_t(h).$$

This is the Stein equation. For a given h one can in fact compute explicitly f by solving

$$\mathscr{S}f(k) = tf(k+1) - kf(k) = h(k) - \nu_t(h)$$

for $f(k+1)$. Then

$$\begin{aligned} f(k+1) &= \frac{k}{t}f(k) + \frac{1}{t}(h(k) - \nu_t(h) \\ &= \frac{k(k-1)}{t^2}f(k-1) + \frac{k}{t^2}(h(k-1) - \nu_t(h)) + \frac{1}{t}(h(k) - \nu_t(h)) \end{aligned}$$

and recursively

$$f(k+1) = \frac{k!}{t^k}f(0) + \frac{k!}{t^k}\sum_{i=0}^{k}(h(i) - \nu_t(h))\frac{t^i}{i!},$$

where $f(0)$ is arbitrary. Similarly one can also show that

$$f(k+1) = -\frac{k!}{t^k}\sum_{i=k+1}^{\infty}(h(i) - \nu_t(h))\frac{i^t}{i!}$$

Lemma 16. *Let $E \subset \mathbb{N}_0$ and $h = \chi_E$. Then if f solves the Stein equation $\mathscr{S}f = \chi_E - \nu_t(E)$ we get*

$$f_{\chi_E}(k) \le \begin{cases} 1 & \text{if } k \le t \\ \frac{2+t}{k} & \text{if } k > t \end{cases}$$

and

$$\sum_{i=1}^{m} f_{\chi_E}(i) \le \begin{cases} m & \text{if } m \le t \\ t + (2+t)\log\frac{m}{t} & \text{if } m > t \end{cases}.$$

This is proven by using the explicit representation given above.

One has the following result.

Proposition 17. [9] *A probability measure ν on $\mathbb{N}_0$ is Poisson for t if and only if $\int_{\mathbb{N}_0} \mathscr{S} f\,d\nu = 0$ for all bounded $f \in \mathcal{F}$.*

Proof. One has

$$\begin{aligned} \int \mathscr{S} f\,d\nu &= \sum_k \mathscr{S} f(k)\nu_t(\{k\}) \\ &= \sum_k (tf(k+1) - kf(k))\,\nu_t(\{k\}) \\ &= \sum_\ell f(\ell)\,(t\nu_t(\{\ell-1\}) - \ell\nu_t(\{\ell\}))\,. \end{aligned}$$

Hence $\int \mathscr{S} f\,d\nu = 0$ for all $f \in \mathcal{F}$ if and only if $t\nu(\{\ell-1\}) = \ell\nu(\{\ell\})\forall\ell$. Hence $\nu(\{\ell\}) = \frac{t}{\ell}\nu(\{\ell-1\}) = \frac{t^2}{\ell(\ell-1)}\nu(\{\ell-2\}) = \cdots = \frac{t^\ell}{\ell!}\nu(\{0\})$. That means ν is Poisson for t and $\nu(\{0\})$ is the normalising term. ∎

4.2. **Higher order returns for α-mixing measures.** We want to apply this to higher order returns and assume we have a map $T : \Omega \circlearrowleft$ and a T-invariant probability measure μ on Ω. For a subset $A \subset \Omega$ we put

$$W_m = \sum_{i=1}^{m} \chi_A \circ T^i$$

for the counting function on orbit segments of lengths m. Then $t = \mathbb{E}(W_m) = \mu(W_m) = m\mu(A)$ by invariance of μ. That is $m = \frac{t}{\mu(A)}$ (Kac scaling). We now state the main result of this section.

eorem.poisson}

Theorem 18. *Let μ be an α-mixing T-invariant measure on Ω. Then there exists a constant C so that for $A \in \sigma(\mathcal{A}^n)$, ($n \in \mathbb{N}$) and $E \subset \mathbb{N}_0$ one has*

$$|\mathbb{P}(W_m \in E) - \nu_t(E)| \le C \inf_{n<\Delta<m} \left(\Delta\mu(A) + \frac{\alpha(\Delta)}{\mu(A)} + \mathbb{P}_A(\tau_A < \Delta)\right) |\log\mu(A)|$$

For $E \subset N_0$ let again be $h = \chi_E$ and f the solution of the Stein equation $\mathscr{S} f = \chi_E - \nu_t(E)$. Then for a probability measure ν on $\mathbb{N}_0$ we have

$$\begin{aligned}
\nu(E) - \nu_t(E) &= \int \chi_E \, d\nu - \int \chi_E \, d\nu_t \\
&= \int (\chi_E - \nu_t(E)) \, d\nu \\
&= \int \mathscr{S}(f) \, d\nu \\
&= \int (tf(k+1) - kf(k)) \, d\nu(k).
\end{aligned}$$

Let $E \subset \mathbb{N}_0$ and f the solution to Stein's equation $\mathscr{S} f = \chi_E - \nu_t(E)$. Let ν be the probability measure on $\mathbb{N}_0$ given by $\nu(E) = \mathbb{P}(W_m \in E)$. Then $\int \mathscr{S} f \, \nu = \nu(E) - \nu_t(E)$ and consequently

$$\mu(W_m \in E) - \nu_t(E) = \int \mathscr{S} f \, d\nu = \mathbb{E}(tf(W_n + 1)) - \mathbb{E}(W_m f(W_m)).$$

With $I_i = \chi_A \circ T^i$ and $W_m = \sum_{i=1}^m I_i$ we get $W_m f(W_m) = \sum_{i=1}^m I_i f(W_m)$ and

$$\begin{aligned}
|\mathbb{P}(W_m \in E) - \nu_t(E)| &= \left| t\mathbb{E}(f(W_m + 1)) - \sum_{i=1}^m \mathbb{E}(I_i f(W_m)) \right| \\
&= \left| \sum_i \mu(I_i) \left(\mathbb{E}(f(W_m + 1)) - \mathbb{E}(f(W_m)|I_i) \right) \right| \\
&= \mu(A) \sum_i \left| \sum_{a=0}^m (f(a+1)\mathbb{P}(W_m = a) - f(a)\mathbb{P}(W_m = a|I_i)) \right| \\
&\le \mu(A) \sum_a f(a+1) \varepsilon_{a,i},
\end{aligned}$$

where

$$\begin{aligned}
\varepsilon_{a,i} &= |\mathbb{P}(W_m = a) - \mathbb{P}(W_m = a + 1|I_i)| \\
&= \left| \mathbb{P}(W_m = a) - \frac{\mathbb{P}(W_m = a \cap T^{-i}A)}{\mu(A)} \right| \\
&\le \left| \mathbb{P}(W_m = a) - \frac{\mathbb{P}(W_m^i = a)\mu(A) + \xi_{a,i}}{\mu(A)} \right| \\
&\le \left| \mathbb{P}(W_m = a) - \mathbb{P}(W_m^i = a) \right| + \frac{\xi_{a,i}}{\mu(A)},
\end{aligned}$$

where $W_m^i = W_m - I_i$ and

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

Proposition 19. *Let μ be α-mixing. Then there exists a constant C so that for all $A\in\sigma(\mathcal{A}^n)$,*

$$\sum_{a=0}^m f(a+1)\xi_{a,i}\le C\mu(A)\inf_{r_A\le\Delta<m}\left(\Delta\mu(A)+\frac{\alpha(\Delta)}{\mu(A)}+\mathbb{P}_A(\tau_A\le\Delta)\right)\lg m,$$

$i=1,\dots,m$, where

$$r_A=\min\{j\ge1: A\cap T^{-i}A\ne\varnothing\}$$

is the period *of A.*

Remark: If μ is ϕ-mixing, then

$$\mathbb{P}(\tau_A\le\Delta)\le\sum_{i=r_A}^{\Delta}\delta_A(j)$$

where

$$\delta_A(j)=\inf_{0\le w\le j}\left(\mu(A_w(A))+\phi(j-w)\right)$$

and $A_w(A)\in\mathcal{A}^w$ is the w-cylinder that contains A.

Proof. Let us establish the following notation:

$$W_m^i=W_m-I_i=W_m^{i,-}+W_m^{i,+}+U_m^{i,-}+U_m^{i,+},$$

where

$$W_m^{i,-}=\sum_{j<i-\Delta}I_j,\quad W_m^{i,+}=\sum_{j>i+\Delta}I_j,\quad U_m^{i,-}=\sum_{i-\Delta\le j<i}I_j,\quad U_m^{i,+}=\sum_{i+\Delta\ge j>i}I_j$$

and also $\tilde W_m^i=W_m^{i,-}+W_m^{i,+}$ for the total sum with gap and $U_m^i=U_m^{i,-}+U_m^{i,+}$ for the gap terms without I_i. Then

$$\mathbb{P}(W_m^i=a\cap T^{-i}A)=\sum_{\substack{a^\pm,a^{0,\pm}\\ a^-+a^++a^{0,-}+a^{0,+}=a}}\mathbb{P}(\{W_m^{i,\pm}=a^\pm\}\cap\{U_m^{i,\pm}=a^{0,\pm}\}\cap T^{-i}A)$$

and

$$\xi_{a,i}\le R_1(a)+R_2(a)+R_3(a),$$

where

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

for opening the gap,

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

to get independence,

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

for filling in the gap again.

We estimate the terms independently:

(I) For opening the gap we note that

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

and similarly

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

This implies that $R_1 \le b_i^- + b_i^+$, where

$$b_i^- = \mathbb{P}(\{U^{i,-}_m > 0\} \cap T^{-i}A), \quad b_i^+ = \mathbb{P}(\{U^{i,+}_m > 0\} \cap T^{-i}A).$$

Since

$$b_i^+ = \mathbb{P}(\{U^{i,+}_m > 0\}|I_i)\mu(A) = \mathbb{P}_A(\tau_A < \Delta)\mu(A)$$

and similarly for $b_i^- = b_i^+$ one has

$$R_1(a) \le \mu(A) 2\mathbb{P}_A(\tau_A < \Delta).$$

(II) Estimate of R_3. We observe that

$$\{W^i_m = a\} \subset \{\tilde{W}^i_m = a\} \cup \{U^i_m > 0\}, \qquad \{\tilde{W}^i_m = a\} \subset \{W^i_m = a\} \cup \{U^i_m > 0\}$$

which implies that

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

and consequently

$$R_3(a) \le 2\Delta\mu(A)^2.$$

(III) We use the mixing property to estimate the second error term

$$\begin{aligned} R_2(a) &= \left|\mathbb{P}(\{\tilde{W}^i_m = a\} \cap T^{-i}A) - \mathbb{P}(\tilde{W}^i_m = a)\mu(A)\right| \\ &= \left|\sum_{a^- + a^+ = a} \left(\mathbb{P}(\{W^{i,\pm}_m = a^\pm\} \cap T^{-i}A) - \mathbb{P}(W^{i,\pm}_m = a^\pm)\mu(A)\right)\right| \\ &\le \sum_{a^- + a^+ = a} (R_{2,1}(a^-, a^+) + R_{2,2}(a^-, a^+) + R_{2,3}(a^-, a^+)). \end{aligned}$$

The three steps are first splitting off $W^{i,-}_m$ then splitting off $T^{-i}A$ and finally putting $W^{i,-}_m$ and $W^{i,+}_m$ back together again. For the first term we obtain accordingly

$$R_{2,1}(a^-, a^+) = \mathbb{P}(\{W^{i,\pm}_m = a^\pm\} \cap T^{-i}A) - \mathbb{P}(\{W^{i,+}_m = a^+\} \cap T^{-i}A)\mathbb{P}(W^{i,-}_m = a^-)$$

and we want to estimate $\sum_a f(a+1)R_{2,1}(a^-, a^+)$. For simplicity we assume that $t \ll m$ is small and put

$$\sigma_{a^-,a^+} = \operatorname{sgn}\left(\mathbb{P}(\{W^{i,\pm}_m = a^\pm\} \cap T^{-i}A) - \mathbb{P}(\{W^{i,+}_m = a^+\} \cap T^{-i}A)\mathbb{P}(W^{i,-}_m = a^-)\right).$$

Then by the previous lemma

$$\begin{aligned}
R_{2,1}^+ &= \sum_a \sum_{\substack{a^-+a^+=a\\ \sigma_{a^-,a^+}=+1}} f(a+1)R_{2,1}(a^-,a^+)\\
&\le \sum_a \sum_{\substack{a^-+a^+=a\\ \sigma_{a^-,a^+}=+1}} \frac{2+t}{a+1}R_{2,1}(a^-,a^+)\\
&\le \sum_{k=0}^{\lg m}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{m2^{-k}}\sum_{\substack{a^-\in[0,m2^{-k})\\ \sigma_{a^-,a^+}=+1}}\Big(\mathbb{P}(\{W_m^{i,\pm}=a^\pm\}\cap T^{-i}A)-\mathbb{P}(\{W_m^{i,+}=a^+\}\cap T^{-i}A)\mathbb{P}(W_m^{i,-}=a^-)\Big)
\end{aligned}$$

where we used exponential progression to approximate the function $f(a+1)$, i.e. subdivided into classes where $a\in[m2^{-k-1},m2^{-k})$. Thus

$$\begin{aligned}
R_{2,1}^+ &\le \sum_{k=0}^{\lg m}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{m2^{-k}}\Bigg(\mathbb{P}(\{W_m^{i,+}=a^+\}\cup\bigcup_{a^-,\sigma_{a^-,a^+}=+1}\{W_m^{i,-}=a^-\}\cap T^{-i}A)\\
&\qquad -\mathbb{P}(\{W_m^{i,+}=a^+\}\cap T^{-i}A)\mathbb{P}(\bigcup_{a^-,\sigma_{a^-,a^+}=+1}\{W_m^{i,-}=a^-\})\Bigg)\\
&\le \sum_{k=0}^{\lg m}\frac{2+t}{m2^{-k}}\sum_{a^+=0}^{m2^{-k}}\alpha(\Delta-n)
\end{aligned}$$

where we used that

$$\{W_m^{i,+}=a^+\}\cap T^{-i}A\in T^{-i-\Delta}\sigma(\mathcal{A}^{m-i+n}),\qquad \{W_m^{i,-}=a^-\}\in\sigma(\mathcal{A}^{i-\Delta+n}).$$

Hence ($n\ll\Delta$)

$$R_{2,1}^+\le c_1(2+t)\sum_{k=0}^{\lg m}\alpha(\Delta)\le c_2(2+t)\alpha(\Delta)\lg m.$$

In a similar way one estimates the terms with $\sigma_{a^-,a^+}=-1$ to get $R_{2,1}^-\le c_2(2+t)\alpha(\Delta)\lg m$. Consequently $R_{2,1}\le 2c_2(2+t)\alpha(\Delta)\lg m$.

Similarly one estimates

$$\begin{aligned}
R_{2,2} &= \sum_a f(a+1)\sum_{a^-+a^+=a}\Big(\mathbb{P}(\{W_m^{i,+}=a^+\}\cap T^{-i}A)-\mathbb{P}(W_m^{i,+}=a^+)\mu(A)\Big)\,\mathbb{P}(W_m^{i,-}=a^-)\\
&\le \alpha(\Delta)(2+t)\lg m
\end{aligned}$$

and

$$R_{2,3}\le\alpha(2\Delta)(2+t)\lg m.$$

These three estimates combined yield

$$R_2\le c_3(2+t)\alpha(\Delta)\lg m$$

and everything combined proves the statement in the proposition. ∎

Proof of Theorem 18. For the proof of the theorem we get

$$\begin{aligned}|\mathbb{P}(W_m \in E) - \nu_t(E)| &\leq \sum_{i=1}^{m} \mu(A) \sum_{a=0}^{m} f(a+1)\,(\mu(A) + \xi_{a,i}) \\ &\leq \sum_a f(a+1)\mu(A) + \sum_a f(a+1) \frac{\xi_{a,i}}{\mu(A)} \\ &\leq c_1 \mu(A) \lg m + c_2 \inf_{\Delta} \left(\frac{\alpha(\Delta)}{\mu(A)} + \Delta\mu(A) + \mathbb{P}_A(\tau_A \leq \Delta) \right) \lg m.\end{aligned}$$

∎

4.3. **ϕ-mixing measures.** Let μ be polynomially ϕ-mixing so that $\phi(k) \lesssim k^{-p}$. In order to estimate the third term inside the brackets of Theorem 18 we get for $A = A_n(x)$ an n-cylinder at some point x that

$$\mathbb{P}(\tau_A < \Delta) \leq \sum_{j=r_A}^{\Delta} \mu(A_n(x) \cap T^{-j} A_n(x))$$

where the individual terms in the sum are for a right ϕ-mixing measure μ bounded by

$$\mu(A_n(x) \cap T^{-j} A_n(x)) \leq \left(\mu(A_{\frac{j}{2} \wedge n}(x)) + \phi(j/2) \right) \mu(A_n(x)).$$

Hence, if we choose $\Delta \sim \mu(A_n(x))^{-\omega}$ for any $\omega \in (0,1)$ then

$$\mathbb{P}(\tau_A < \Delta) \leq \mu(A_n(x)) \sum_{j=r_A}^{\Delta} \left(\mu(A_{\frac{j}{2} \wedge n}(x) + \phi(j/2) \right)$$

and consequently

$$\mathbb{P}_A(\tau_A < \Delta) \leq c_1 \left(\vartheta^{r_A/2} + r_A^{-(p-1)} + \Delta\mu(A_n(x)) \right) \lesssim r_A^{-(p-1)}$$

since $\mu(A_w(x)) \lesssim \vartheta^w$ for some $\vartheta < 1$. If x is not a periodic point, then $r_{A_n(x)} \to \infty$ as $n \to \infty$.

The first term inside the brackets of Theorem 18 converges to zero as n goes to infinity because $\Delta\mu(A_n(x)) \lesssim \mu(A_n(x))^{1-\omega}$ and ω is positive.

The second term is estimated thusly

$$\frac{\alpha(\Delta)}{\mu(A_n(x))} \leq \frac{\phi(\Delta)}{\mu(A_n(x))} \lesssim \frac{\Delta^{-p}}{\mu(A_n(x))} \lesssim \mu(A_n(x))^{p\omega - 1}$$

which goes to zero as $n \to \infty$ since we can choose $\omega < 1$ close enough to 1 so that $\omega p > 1$.

Since the same argument can be made for left ϕ-mixing measures, we obtain the following result.

Theorem 20. *Let μ be a left or right ϕ-mixing measure so that $\phi(k) \lesssim k^{-p}$ for some $p > 2$. Then for all x not periodic we obtain*

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

as $n \to \infty$ *provided* $\log \mu(A_n(x)) \lesssim n$ *and* $r_{A_n(x)} \gtrsim n$.

References

[1] M Abadi: Exponential Approximation for Hitting Times in Mixing Stochastic Processes; Mathematical Physics Electronic Journal **7** (2001).

[2] M Abadi: Instantes de ocorrência de eventos raros em processos misturadores; PhD thesis 2001, Universidade de São Paulo.

[3] M Abadi: Sharp error terms and necessary conditions for exponential hitting times in mixing processes; Ann. Prob. **32** (2004), 243–264.

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

[5] M Abadi: Poisson approximations via Chen-Stein for non-Markov processes; *In and Out of Equilibrium 2* V Sidoravicius and M E Vares (editors), 2008, pp1–19.

[6] M Abadi and B Saussol: Hitting and returning into rare events for all alpha-mixing processes; Stoch. Proc. Appl. **121** (2011), 314–323.

[7] M Abadi and N Vergne: Sharp errors for point-wise Poisson approximations in mixing processes; Nonlinearity **21** (2008), 2871–2885.

[8] R Arratia, L Goldstein and L Gordon: Poisson approximation and the Chen-Stein method. With comments and a rejoinder by the authors; Stat. Sci. **5**, 403–434.

[9] A. D. Barbour, Louis H.Y. Chen, *Stein's Method and Applications, Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore*, Vol. **5** 2005

[10] R Bowen: *Equilibrium States for Anosov Diffeomorphism*; Springer Lecture Notes 470, Springer, New York/Berlin 1975. Second revised edition Springer-Verlag, Berlin, 2008.

[11] H Bruin, B Saussol, S Troubetzkoy and S Vaienti: Return time statistics via inducing; Ergod. Th. & Dynam. Syst. **23**, 991–1013 (2003).

[12] V Chamoître and M Kupsa: k-limit laws of return and hitting times; Discrete and Continuous Dynamical Systems **15** (2006), 73–86.

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

[14] M Denker, M Gordin and A Sharova: A Poisson limit theorem for toral automorphisms; Illinois J. Math. **48(1)** (2004), 1–20.

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

[16] A Galves and B Schmitt: Inequalities for hitting times in mixing dynamical systems; Random Comput. Dynam. **5** (1997), 337–347.

[17] N Haydn: A note on the limiting entry and return times distributions for induced maps; available at `http://arxiv.org/abs/1208.6059`.

[18] N Haydn, Y Lacroix and S Vaienti: Hitting and Return Times in Ergodic Dynamical Systems: Ann. of Probab. 33 (2005), 2043–2050

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

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

[21] N Haydn, N Winterberg and R Zweimüller: Return-time statistics, Hitting-time statistics and Inducing; in *Ergodic Theory, Open Dynamics and Coherent Structures*, Bahsoun, Bose and Froyland editors, Springer Proceedings in Mathematics & Statistics Vol. 70, 2014

[22] N Haydn and S Vaienti: The limiting distribution and error terms for return times of dynamical systems; Disc. Cont. Dyn. Syst. **10** (2004) 589–616.

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

[24] N Haydn and F Yang: A Derivation of the Poisson Law for Returns of Smooth Maps with Certain Geometrical Properties; Contemporary Mathematics Proceedings in memoriam Chernov 2017

[25] N Haydn and F Yang: Entry times distribution for dynamical balls on metric spaces; J. Stat. Phys. (2017) DOI: 10.1007/s10955-017-1745-7.

[26] N Haydn and F Yang: Entry times distribution for mixing systems; J. Stat. Phys. (2016) DOI 10.1007/s10955-016-1487-y.

[27] M Hirata: Poisson law for Axiom A diffeomorphisms; Ergod. Th. & Dynam. Syst. **13** (1993), 533–556.

[28] M Hirata: Poisson law for the dynamical systems with the "self-mixing" conditions; *Dynamical Systems and Chaos*, Vol. 1 (Worlds Sci. Publishing, River Edge, New York (1995), 87–96.

[29] M Hirata, B Saussol and S Vaienti: Statistics of return times: a general framework and new applications; Comm. Math. Phys. **206** (1999), 33–55.

[30] M Kac: On the notion of recurrence in discrete stochastic processes; Bull. A.M.S. **53** (1947),1002–1010.

[31] Y Kifer, A Rapaport: Poisson and compound Poisson approximations in conventional and nonconventional setups, Probab. Th. Related Fields **160** (2014), 797–831.

[32] M Kupsa and Y Lacroix: Asymptotics for hitting times, Ann. of Probab. **33(3)** (2005), 610–614.

[33] Y Lacroix: Possible limit laws for entrance times of an ergodic aperiodic dynamical system; Israel J. Math. **132** (2002), 253–264.

[34] B Pitskel: Poisson law for Markov chains; Ergod. Th. & Dynam. Syst. **11** (1991), 501–513.

[35] H Poincaré: Les méthodes nouvelles de la mécanique céleste; vol. 3, Gauthiers-Villars, Paris 1899.

[36] L-S Young: Statistical properties of dynamical systems with some hyperbolicity; Annals of Math. **7** (1998), 585–650.

[37] L-S Young: Recurrence time and rate of mixing; Israel J. of Math. **110** (1999), 153–188.