The general asymptotic return-time process

(to appear in Israel J Math)

Roland Zweimüller

Abstract. We study limit processes for consecutive return-times of asymptotically rare events in general ergodic probability preserving systems. It is shown that in every ergodic system on a non-atomic space any non-negative stationary sequence with expectation not exceeding 1 occurs as a limit process. Moreover, the limiting behaviour is shown to be robust under small changes of the sets. We also determine the relation between asymptotic return-time processes and asymptotic hitting-time processes, and record some consequences.

1. Introduction

The asymptotic behaviour of return-times and hitting-times of small sets A in an ergodic probability-preserving dynamical system $(X, \mathcal{A}, \mu, T)$ has been studied in great detail and from various points of view. The purpose of the present article is to extend several results about distributional limits of *first* return-times and *first* hitting-times to the corresponding return- and hitting-time *processes*.

Throughout, $(X, \mathcal{A}, \mu)$ is a fixed probability space, and $T : X \to X$ is an ergodic μ-preserving map. Also, A, A_l and Y will always denote measurable sets of *positive* measure. By ergodicity and the Poincaré recurrence theorem, the measurable *(first) hitting time* function of A, $\varphi_A : X \to \overline{\mathbb{N}} := \{1, 2, \ldots, \infty\}$ with $\varphi_A(x) := \inf\{n \geq 1 : T^n x \in A\}$, is finite a.e. on X. When restricted to A it is called the *(first) return time* function of our set. Distinguishing between these two variants will be an important issue, and we sometimes use the notation $\widetilde{\varphi}_A := \varphi_A \mid_A$ when we wish to make the fact that we speak about the latter immediately recognizable.[1]

The return time satisfies Kac' formula $\int_A \varphi_A \, d\mu_A = 1/\mu(A)$, where $\mu_A(B) := \mu(A \cap B)/\mu(A)$, $B \in \mathcal{A}$. That is, when regarded as a random variable on the probability space $(A, \mathcal{A} \cap A, \mu_A)$, it has expectation $\mathbb{E}[\widetilde{\varphi}_A] = 1/\mu(A)$, and we will usually normalize our variable accordingly, thus passing to $\mu(A)\,\widetilde{\varphi}_A$. Of particular interest are the *(normalized) return-time distribution* of A, encoded in $\widetilde{F}_A(t) := \mu_A(\mu(A)\,\widetilde{\varphi}_A \leq t)$, as well as its *(normalized) hitting-time distribution* represented via $F_A(t) := \mu(\mu(A)\,\varphi_A \leq t)$, with $t \geq 0$.

Key words and phrases. Mathematics Subject Classification (2010): 28D05, 60F05, 60B12, 60G10.

[1] Often there is no logical necessity of writing $\widetilde{\varphi}_A$ rather than φ_A, but we choose to do so (at the cost of adding some redundancy or even arbitrariness) where it improves readability.

Consider a sequence $(A_l)_{l\geq 1}$ of *asymptotically rare events*, meaning that $A_l \in \mathcal{A}$ with $\mu(A_l) \to 0$. It is said to have *(asymptotic) return time statistics* given by a random variable $\widetilde{\varphi}$ which takes values in $[0,\infty]$, if $\widetilde{F}_{A_l} \Longrightarrow \widetilde{F}$, with $\widetilde{F}(t) := \Pr[\widetilde{\varphi} \leq t]$, $t \geq 0$. As usual, $\Longrightarrow$ indicates convergence at all continuity points of the (not necessarily normalized) limit function. Similarly, $(A_l)_{l\geq 1}$ has *(asymptotic) hitting time statistics* given by some $[0,\infty]$-valued variable φ if $F_{A_l} \Longrightarrow F$, with $F(t) := \Pr[\varphi \leq t]$, $t \geq 0$ (where $\Pr[\varphi = \infty]$ can be positive). To indicate more explicitly the role of the underlying probability measures, we shall express these asymptotic relations by writing $\mu(A_l)\,\widetilde{\varphi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\varphi}$ and $\mu(A_l)\,\varphi_{A_l} \overset{\mu}{\Longrightarrow} \varphi$, respectively.

There is an extensive literature about limit theorems of these types, with numerous articles studying specific classes of dynamical systems. Remarkably, some fundamental results regarding the general theory have only been obtained at a surprisingly late stage. The family of possible limit laws for return-times was determined in [**10**], which contains the following result. (Note that by normalization it is a priori clear that $\mathbb{E}[\widetilde{\varphi}] \leq 1$ for any return-time limit $\widetilde{\varphi}$.)

Theorem A (**Prescribing the asymptotic return-time statistics**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system on a non-atomic space, and let $\widetilde{\varphi}$ be any non-negative random variable with $\mathbb{E}[\widetilde{\varphi}] \leq 1$. Then there is some sequence $(A_l)_{l\geq 1}$ of asymptotically rare events such that*

$$\mu(A_l)\,\widetilde{\varphi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\varphi} \quad \text{as } l \to \infty. \tag{1.1}$$

There is a corresponding result for asymptotic hitting-time statistics, first established in [**9**]. The latter also follows from the main theorem of [**5**], which clarifies the relation between the two types of limits: The asymptotic hitting-time statistics is given by the integrated tail distribution of the return-time limit. The result reads

Theorem B (**Return-time statistics versus hitting-time statistics**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability-preserving system, and $(A_l)_{l\geq 1}$ a sequence of asymptotically rare events. Then*

$$\mu(A_l)\,\widetilde{\varphi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\varphi} \quad \text{for some random variable } \widetilde{\varphi} \text{ in } [0,\infty] \tag{1.2}$$

iff

$$\mu(A_l)\,\varphi_{A_l} \overset{\mu}{\Longrightarrow} \varphi \quad \text{for some random variable } \varphi \text{ in } [0,\infty]. \tag{1.3}$$

In this case, the sub-probability distribution functions F and $\widetilde{F}$ of φ and $\widetilde{\varphi}$ satisfy

$$\int_0^t [1 - \widetilde{F}(s)]\, ds = F(t) \quad \text{for } t \geq 0. \tag{1.4}$$

In particular, $\widetilde{F}$ is always a probability d.f. with $\int_0^\infty [1 - \widetilde{F}(s)]\, ds \leq 1$, and F may be degenerate, as $\mathbb{E}[\widetilde{\varphi}] = \Pr[\varphi < \infty]$. Through the integral equation (1.4) each of F and $\widetilde{F}$ uniquely determines the other.

The purpose of the present article is to extend these general results about one-dimensional limit laws to the level of processes.

The point at which the orbit of x first enters A is given by $T_A\, x := T^{\varphi_A(x)} x$, $x \in X$. This defines the (measurable) *first entrance map* $T_A : X \to A$. Restricting it to A, we obtain the standard *first return map* $T_A : A \to A$. It is a well-known classical result that the latter is a measure-preserving ergodic map on the probability space $(A, \mathcal{A} \cap A, \mu_A)$.

In the present paper we will focus, for sets A as above, on the random sequences of consecutive return- and hitting-times, that is, we are going to consider the sequences $\Phi_A : X \to [0,\infty)^{\mathbb{N}_0}$, $\mathbb{N}_0 = \{0,1,\ldots\}$, of functions given by

$$\Phi_A := (\varphi_A, \varphi_A \circ T_A, \varphi_A \circ T_A^2, \ldots) \quad \text{on } X. \tag{1.5}$$

When regarded as a random sequence defined on $(X, \mathcal{A}, \mu)$, we shall call Φ_A the *hitting-time process* of A. If we view it through μ_A, it is the *return-time process* of A. To emphasize that we deal with the latter, we sometimes use the notation $\widetilde{\Phi}_A := \Phi_A \mid_A$. Via a natural duality (detailed below), these processes are related to the counting processes which record the occupation times of A.

We are going to generalize Theorems A and B to convergence $\mu(A_l)\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi}$ and $\mu(A_l)\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi$ of the normalized variants of these processes, and discuss several related issues.

2. The general asymptotic return-time process

Finite-dimensional marginals and distributional convergence. Let $\Phi = (\varphi^{(j)})_{j\geq 0}$ be a random sequence in $[0,\infty]$ (that is, a random element of $[0,\infty]^{\mathbb{N}_0} = \{(s_j)_{j\geq 0} : s_j \in [0,\infty]\}$ equipped with its Polish product topology). We let $\Phi^{[d]} := (\varphi^{(0)}, \ldots, \varphi^{(d-1)})$ denote its initial piece of length d, $d \geq 1$. The (possibly degenerate) distribution function of the random vector $\Phi^{[d]}$ is $F^{[d]} : [0,\infty)^d \to [0,1]$, $F^{[d]}(t_0, \ldots, t_{d-1}) := \Pr[\Phi^{[d]} \leq (t_0, \ldots, t_{d-1})] := \Pr[\varphi^{(0)} \leq t_0, \ldots, \varphi^{(d-1)} \leq t_{d-1}]$.

Assuming that each Φ_l, $l \geq 1$, is a random sequence in $[0,\infty)$, we shall write

$$\Phi_l \Longrightarrow \Phi \quad \text{as } l \to \infty, \tag{2.1}$$

if all finite-dimensional distribution functions $F_l^{[d]} : [0,\infty)^d \to [0,1]$ converge weakly to the corresponding distribution functions $F^{[d]}$ of Φ, that is, $F_l^{[d]}(t_0, \ldots, t_{d-1}) \to F^{[d]}(t_0, \ldots, t_{d-1})$ at all continuity points $(t_0, \ldots, t_{d-1})$ of $F^{[d]}$. This way, the law of the limit process is uniquely determined when conditioned on $\{\Phi \in [0,\infty)^{\mathbb{N}_0}\}$[2]. Below, the Φ_l will be measurable maps defined on the same space $(X, \mathcal{A})$, but viewed through different probability measures ν_l on $(X, \mathcal{A})$, so that $F_l^{[d]}(t_0, \ldots, t_{d-1}) = \nu_l(\{\Phi_l^{[d]} \leq (t_0, \ldots, t_{d-1})\})$. We express this by writing $\Phi_l \overset{\nu_l}{\Longrightarrow} \Phi$. Specifically, when speaking about return-time processes, we take $\Phi_l = \widetilde{\Phi}_{A_l}$ and $\nu_l = \mu_{A_l}$. Following the convention for the univariate situation, the limit will then be denoted $\widetilde{\Phi}$,

$$\Phi_l \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} \quad \text{as } l \to \infty. \tag{2.2}$$

For a random sequence $\Phi = (\varphi^{(j)})_{j\geq 0}$ in $[0,\infty]$ its distribution on the sequence space will be denoted by law(Φ). Likewise, law$(\Phi^{[d]})$ is the law of $\Phi^{[d]}$ on $[0,\infty]^d$. Also, let $\sigma^m\Phi$ denote the *shifted* versions $(\varphi^{(j+m)})_{j\geq 0}$. Φ is *stationary* if $\sigma\Phi$ has the same law as Φ. If Φ is *a.s. finite* in that $\Pr[\varphi^{(j)} < \infty \text{ for all } j \geq 0] = 1$, then stationarity is a property of the $F^{[d]}$.

[2] As the variables in theses sequences are successive waiting times, details of the conditional law in the degenerate situation in which some diverge are, arguably, of lesser interest.

Associated point processes. Let $\mathbf{M} := \mathbf{M}_p([0,\infty))$ be the space of counting measures on $([0,\infty),\mathcal{B}_{[0,\infty)})$, that is, measures $\mathsf{n} : \mathcal{B}_{[0,\infty)} \to \mathbb{N}_0 \cup \{\infty\}$. Equipped with the topology of *vague convergence*, meaning that $\mathsf{n}_l \to \mathsf{n}$ in $\mathbf{M}$ iff $\mathsf{n}_l(f) \to \mathsf{n}(f)$ for every continuous function $f : [0,\infty) \to \mathbb{R}$ with compact support, $\mathbf{M}$ is a Polish space. A *point process* N on $[0,\infty)$ is a random element of $\mathbf{M}$. For a sequence $(\mathsf{N}_l)_{l\geq 1}$ of point processes defined on $(X,\mathcal{A})$ and viewed through the probability measures ν_l on $(X,\mathcal{A})$, distributional convergence will again be denoted by $\mathsf{N}_l \overset{\nu_l}{\Longrightarrow} \mathsf{N}$.

Specifically, we are interested in the point processes N_A given by the successive visits to a target set A: The *interarrival times* between consecutive visits to A are $\varphi_A\,, \varphi_A \circ T_A\,, \varphi_A \circ T_A^2\,, \ldots$ and the *arrival times* are the $\varphi_{A,k} := \sum_{j=0}^{k-1} \varphi_A \circ T_A^j$, $k \geq 1$. The number of visits to A within the time set $\mu(A)^{-1}B \in \mathcal{B}_{[0,\infty)}$ defines the point process $\mathsf{N}_A : X \to \mathbf{M}_p([0,\infty))$, $\mathsf{N}_A(B) := \sum_{k\geq 1} 1_B(\mu(A)\,\varphi_{A,k})$.

According to standard results (see e.g. [**8**]), convergence in law of the point processes is equivalent to convergence of finite tuples of interarrival times, hence

$$\mathsf{N}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \mathsf{N} \quad \text{iff} \quad \mu(A_l)\,\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} = (\widetilde{\varphi}^{(j)})_{j\geq 0}, \tag{2.3}$$

where the $\widetilde{\varphi}^{(j)}$, $j \geq 0$, are interarrival times of a locally finite process N.

We let φ_{Exp} denote a generic exponentially distributed random variable with distribution function $F_{\mathrm{Exp}}(t) := 1 - e^{-t}$, and $\Phi_{\mathrm{Exp}} = (\varphi_{\mathrm{Exp}}^{(0)}, \varphi_{\mathrm{Exp}}^{(1)}, \ldots)$ an iid sequence of such variables. In view of the duality above we take the liberty of referring to situations in which $\mu(A_l)\,\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \Phi_{\mathrm{Exp}}$ as exhibiting *Poisson asymptotics*, since in this case the corresponding N is a standard Poisson process.

The possible limits of return-time processes. A large body of work establishes Poisson asymptotics for natural sets in various types of dynamical systems. Beyond this, hardly anything appears to be known about the possible asymptotic dependence structure of asymptotic return-time processes.

Looking at the normalized processes $\mu(A_l)\widetilde{\Phi}_{A_l}$ for a sequence of asymptotically rare events, some easy properties shared by any limit process are fairly obvious:

Proposition 2.1 (**Stationarity and expectation of return time limits**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability preserving system, and $(A_l)_{l\geq 1}$ a sequence of asymptotically rare events. Assume that $\mu(A_l)\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi}$ for some random sequence $\widetilde{\Phi} = (\widetilde{\varphi}^{(j)})_{j\geq 0}$ in $[0,\infty]$. Then $\mathbb{E}[\widetilde{\varphi}^{(0)}] \leq 1$ and $\widetilde{\Phi}$ is stationary.*

Proof. The statement about the expectation is contained in Theorem A. Since, for every $A \in \mathcal{A}$ with $\mu(A) > 0$, $\widetilde{\Phi}_A$ is a stationary sequence on $(A, \mathcal{A}\cap A, \mu_A)$, it is straightforward that $\widetilde{\Phi}$ is stationary as well. □

We are going to show that any process satisfying the easy properties found above is in fact an asymptotic return-time process for every aperiodic ergodic system:

Theorem 2.1 (**Prescribing the asymptotic return-time process**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability preserving system on a non-atomic space, and let $\widetilde{\Phi} = (\widetilde{\varphi}^{(0)}, \widetilde{\varphi}^{(1)}, \ldots)$ be any non-negative stationary process with $\mathbb{E}[\widetilde{\varphi}^{(0)}] \leq 1$. Then there is some sequence $(A_l)_{l\geq 1}$ of asymptotically rare events in $\mathcal{A}$ such that*

$$\mu(A_l)\,\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} \quad \text{as } l \to \infty. \tag{2.4}$$

The proof of this result is given in Section 9 below.

REMARK 2.1 (**Possible limits for d-dimensional marginals**). *The final section of [**4**] raises the question of how to characterize all possible limits of the d-dimensional joint distributions of consecutive return-times. In view of Lemma 4.4 below, our Theorem implies that these are exactly the finite-dimensional marginals of the stationary sequences $\widetilde{\Phi}$ with $\mathbb{E}[\widetilde{\varphi}^{(0)}] \leq 1$.*

More specific types of sequences and robustness. Our definition of an asymptotically rare sequence (A_l) does not require the A_l to form a nested sequence of sets. This allows to apply the theory quite flexibly to a wide variety of situations. It is worth pointing out that the limit processes arising in this setup are, however, exactly the same as in the case of nested sequences.

PROPOSITION 2.2 (**Limit processes have realizations via decreasing sequences**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability-preserving system, and $(A_l)_{l\geq 1}$ a sequence of asymptotically rare events such that*

$$\mu(A_l)\,\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} \quad \text{as } l \to \infty. \tag{2.5}$$

Then the system also admits a decreasing sequence $(A'_j)_{j\geq 1}$ of asymptotically rare events for which

$$\mu(A'_j)\,\widetilde{\Phi}_{A'_j} \overset{\mu_{A'_j}}{\Longrightarrow} \widetilde{\Phi} \quad \text{as } l \to \infty. \tag{2.6}$$

Moreover, if $(G_i)_{i\geq 1}$ is another asymptotically rare sequence, then the sequence $(A'_j)_{j\geq 1}$ above can be chosen in such a way that each A'_j includes some G_i.

REMARK 2.2. *In its original form in [**10**], Theorem A is stated in the more specific setup of topological dynamical systems, and also asserts that the A_l can be taken to be neighborhoods of some particular point $x \in X$. That this statement remains true for processes is immediate from Proposition 2.2 if we take $(G_i)_{i\geq 1}$ to be a sequence of neighborhoods of some x with $\mu(G_i) \to 0$.*

The proposition will follow easily from a robustness property of asymptotic return-time processes, which is of interest in its own right: Sufficiently small changes of the sets don't change the limit processes. (Both results are established in Section 6 below.)

THEOREM 2.2 (**Robustness of return-time processes**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability-preserving system, and (A'_l), (A''_l) two sequences of asymptotically rare events. Suppose that*

$$\mu(A'_l \bigtriangleup A''_l) = o(\mu(A'_l)) \quad \text{as } l \to \infty. \tag{2.7}$$

Let $\widetilde{\Phi}$ be a random sequence in $[0, \infty]$, then

$$\mu(A'_l)\,\widetilde{\Phi}_{A'_l} \overset{\mu_{A'_l}}{\Longrightarrow} \widetilde{\Phi} \quad \text{iff} \quad \mu(A''_l)\,\widetilde{\Phi}_{A''_l} \overset{\mu_{A''_l}}{\Longrightarrow} \widetilde{\Phi}. \tag{2.8}$$

REMARK 2.3 (**Robustness of return-time statistics**). *This implies a corresponding statement for the first return time: Under the assumptions of the theorem, (2.7) ensures that $\mu(A'_l)\,\widetilde{\varphi}_{A'_l} \overset{\mu_{A'_l}}{\Longrightarrow} \widetilde{\varphi}$ iff $\mu(A''_l)\,\widetilde{\varphi}_{A''_l} \overset{\mu_{A''_l}}{\Longrightarrow} \widetilde{\varphi}$.*

3. Return-time processes versus hitting-time processes

Asymptotic hitting-time processes, Strong distributional convergence. Given a sequence $(A_l)_{l\geq 1}$ of asymptotically rare events for $(X,\mathcal{A},\mu,T)$, we are also interested in distributional convergence of their hitting-time processes, that is, in convergence statements of the form

$$\mu(A_l)\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi \quad \text{as } l\to\infty, \tag{3.1}$$

where Φ is a random sequence in $[0,\infty]$. One remarkable aspect of this type of limit theorem is that it always carries over to many other measures.

Let $\mathfrak{E}$ be a Polish space, $(R_l)_{l\geq 1}$ a sequence of measurable maps from $(X,\mathcal{A})$ into $(\mathfrak{E},\mathcal{B}_{\mathfrak{E}})$, and R some random element of $\mathfrak{E}$. Suppose that μ is some σ-finite measure on $(X,\mathcal{A})$. Then *strong distributional convergence* w.r.t. μ of $(R_l)_{l\geq 1}$ to R means that

$$R_l \overset{\nu}{\Longrightarrow} R \qquad \text{for all probability measures } \nu \ll \mu, \tag{3.2}$$

compare [**1**]. This type of convergence will be denoted by $R_l \overset{\mathcal{L}(\mu)}{\Longrightarrow} R$.

It is an interesting but sometimes neglected fact that various distributional limit theorems for ergodic processes automatically hold in this strong sense. A discussion of a natural and widely applicable sufficient condition is given in [**12**]. In particular, as a consequence of Corollary 6 of [**12**] we have

Proposition 3.1 (**Strong distributional convergence of hitting-time processes**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability-preserving system, and $\nu \ll \mu$ some probability measure. Let $(A_l)_{l\geq 1}$ be a sequence of asymptotically rare events. Then, for any random sequence Φ in $[0,\infty]$,*

$$\mu(A_l)\,\Phi_{A_l} \overset{\nu}{\Longrightarrow} \Phi \quad \textit{implies} \quad \mu(A_l)\,\Phi_{A_l} \overset{\mathcal{L}(\mu)}{\Longrightarrow} \Phi. \tag{3.3}$$

(The analogous statement for return-times is false.)

The duality. For any sequence $(A_l)_{l\geq 1}$ of asymptotically rare events, its return-time statistics and its hitting-time statistics are intimately related to each other via Theorem B. We are going to establish a similar result concerning the associated processes. (The proof of this result is given in Section 4 below.)

Theorem 3.1 (**Hitting-time process versus return-time process**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability-preserving system, and $(A_l)_{l\geq 1}$ a sequence of asymptotically rare events. Then*

$$\mu(A_l)\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi \quad \textit{for some random sequence } \Phi \textit{ in } [0,\infty] \tag{3.4}$$

iff

$$\mu(A_l)\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} \quad \textit{for some random sequence } \widetilde{\Phi} \textit{ in } [0,\infty]. \tag{3.5}$$

In this case, the sub-probability distribution functions $F^{[d]}$ and $\widetilde{F}^{[d]}$ of $\Phi^{[d]}$ and $\widetilde{\Phi}^{[d]}$ satisfy, for any $d\geq 0$ (where $\widetilde{F}^{[0]}:=1$) and $t_j\geq 0$,

$$\int_0^{t_0} [\widetilde{F}^{[d]}(t_1,\ldots,t_d) - \widetilde{F}^{[d+1]}(s,t_1,\ldots,t_d)]\,ds = F^{[d+1]}(t_0,t_1,\ldots,t_d). \tag{3.6}$$

Through (3.6), the families $\{F^{[d]}\}_{d\geq 1}$ and $\{\widetilde{F}^{[d]}\}_{d\geq 1}$ uniquely determine each other. Moreover, the shifted sequence $\sigma\Phi = (\varphi^{(j)})_{j\geq 1}$ is a.s. finite on $\{\varphi^{(0)}<\infty\}$.

REMARK 3.1. *Theorems 1.3 and 1.4 of* [**4**] *clarify the relation between convergence of the (one-dimensional) distribution of the k-th return-time and of the k-th hitting-time, respectively.*

One immediate consequence is that some features of return-process convergence automatically carry over to hitting-processes. In particular we have

COROLLARY 3.1 (**Robustness of hitting-time processes**). *The assertions of Theorem 2.2 and Proposition 2.2 remain true if convergence* $\mu(A'_j)\,\widetilde{\Phi}_{A'_j} \overset{\mu_{A'_j}}{\Longrightarrow} \widetilde{\Phi}$ *of return-processes is replaced by convergence* $\mu(A'_j)\,\Phi_{A'_j} \overset{\mu}{\Longrightarrow} \Phi$ *of hitting-processes.*

This duality principle can also be used in the other direction. (For an illustration in the one-dimensional setup of Theorem B see [**6**].)

Below we record a few more consequences of the theorem. The proofs are given in Section 5 below.

Comparing $\mathrm{law}(\sigma\Phi \mid \varphi^{(0)} < \infty)$ **and** $\mathrm{law}(\widetilde{\Phi})$**. Independent sequences.** In the hitting-time process $\Phi_A = (\varphi_A, \varphi_A \circ T_A, \varphi_A \circ T_A^2, \ldots)$ of a set A, only the first variable actually sees the whole space, since $\varphi_A \circ T_A^j = \varphi_A \mid_A \circ T_A^j$ for $j \geq 1$. It is therefore natural to ask whether the shifted version $\mu(A_l)\,\sigma\Phi_{A_l}$ (as a random process on $(X, \mathcal{A}, \mu)$) simply converges to the asymptotic return-process $\widetilde{\Phi}$. Theorem 3.1 shows that this is not necessarily the case, as it implies the easy

PROPOSITION 3.2 (**Characterizing the case** $\mathrm{law}(\sigma\Phi \mid \varphi^{(0)} < \infty) = \mathrm{law}(\widetilde{\Phi})$). *Let* $(X, \mathcal{A}, \mu, T)$ *be an ergodic probability-preserving system, and* $(A_l)_{l\geq 1}$ *a sequence of asymptotically rare events. Assume that (3.4) holds, and that* $\Pr[\varphi^{(0)} < \infty] > 0$. *Then* $\mathrm{law}(\sigma\Phi \mid \varphi^{(0)} < \infty) = \mathrm{law}(\widetilde{\Phi})$ *iff* $\widetilde{\Phi}$ *satisfies*

$$\mathbb{E}[\widetilde{\varphi}^{(0)}] = \mathbb{E}[\widetilde{\varphi}^{(0)} \mid \widetilde{\varphi}^{(1)}, \ldots, \widetilde{\varphi}^{(d)}] \quad \text{for } d \geq 1. \tag{3.7}$$

In particular, there are situations in which $\mathrm{law}(\sigma\Phi \mid \varphi^{(0)} < \infty) \neq \mathrm{law}(\widetilde{\Phi})$.

It is obvious from the proposition that $\mathrm{law}(\sigma\Phi \mid \varphi^{(0)} < \infty) = \mathrm{law}(\widetilde{\Phi})$ holds whenever $\widetilde{\Phi}$ is an independent sequence. It is worth recording that, in addition, Φ as a whole is independent in this case.

PROPOSITION 3.3 ($\Phi \mid \varphi^{(0)} < \infty$ **is independent iff** $\widetilde{\Phi}$ **is**). *Let* $(X, \mathcal{A}, \mu, T)$ *be an ergodic probability-preserving system, and* $(A_l)_{l\geq 1}$ *a sequence of asymptotically rare events. Assume that (3.4) holds, and that* $\Pr[\varphi^{(0)} < \infty] > 0$. *Then* Φ *conditioned on* $\{\varphi^{(0)} < \infty\}$ *is an independent sequence iff* $\widetilde{\Phi}$ *is an independent sequence.*

Characterizing Poisson asymptotics. Besides its intrinsic interest, Theorem B has been shown to be useful in proving convergence of return- and hitting time distributions. Indeed, it is straightforward to check that if F is a probability distribution function on $[0, \infty)$, then

$$\int_0^t [1 - F(s)]\, ds = F(t) \text{ for } t \geq 0 \quad \text{iff} \quad F = F_{\mathrm{Exp}}. \tag{3.8}$$

This easily leads to a method for proving convergence to an exponential law. It is not hard to see that $\widetilde{F}_{A_l}(t) - F_{A_l}(t) \to 0$ for all t from some dense subset of $(0, \infty)$, iff $\mu(A_l)\,\varphi_{A_l} \stackrel{\mu}{\Longrightarrow} \varphi_{\mathrm{Exp}}$ and $\mu(A_l)\,\varphi_{A_l} \stackrel{\mu_{A_l}}{\Longrightarrow} \varphi_{\mathrm{Exp}}$, compare [**7**]. We show that the same principle works for processes.

PROPOSITION 3.4 (**Characterizing Φ_{Exp}**). *Let Φ be some stationary random sequence in $[0,\infty)$. Then $\Phi = \Phi_{\mathrm{Exp}}$ iff the finite-dimensional marginals have distribution functions $F^{[d]}$ satisfying*

$$\int_0^{t_0} [F^{[d]}(t_1,\dots,t_d) - F^{[d+1]}(s,t_1,\dots,t_d)]\,ds = F^{[d+1]}(t_0,t_1,\dots,t_d) \tag{3.9}$$

whenever $d \geq 0$ and $t_j \geq 0$.

Just as in the one-dimensional case, this characterizes Poisson asymptotics.

THEOREM 3.2 (**Characterizing Poisson asymptotics**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability-preserving system, and $(A_l)_{l\geq 1}$ a sequence of asymptotically rare events. Let $\widetilde{F}_l^{[d]}$ and $F_l^{[d]}$ denote the distribution functions of $\Phi_{A_l}^{[d]}$ with respect to μ_{A_l} and μ, respectively. Then*

$$\text{for each } d \geq 0, \quad \widetilde{F}_l^{[d]} - F_l^{[d]} \longrightarrow 0 \quad \text{on a dense subset of } [0,\infty)^d \tag{3.10}$$

iff

$$\mu(A_l)\Phi_{A_l} \stackrel{\mu}{\Longrightarrow} \Phi_{\mathrm{Exp}} \tag{3.11}$$

iff

$$\mu(A_l)\Phi_{A_l} \stackrel{\mu_{A_l}}{\Longrightarrow} \Phi_{\mathrm{Exp}}. \tag{3.12}$$

4. Proof of Theorems 3.1 and 3.2

Our proof of Theorem 3.1 elaborates on an alternative proof of Theorem B given in [**2**]. As a first step we provide a lemma which compares conditional distributions of hitting- and return-times of a set A given an arbitrary conditioning event which takes place at (or after) time φ_A.

LEMMA 4.1 (**Conditional return- and hitting-time distributions**). *Let $(X,\mathcal{A},\mu,T)$ be an ergodic probability preserving system, and $A,B \in \mathcal{A}$ with $B \subseteq A$, and $\mu(B) > 0$. Then, for any $n \geq 1$,*

$$\mu(\{\varphi_A = n\} \cap T_A^{-1}B) = \mu(A \cap \{\varphi_A \geq n\} \cap T_A^{-1}B), \tag{4.1}$$

and for all $t > 0$,

$$\begin{aligned} \mu\big(\{\mu(A)\,\varphi_A \leq t\} \cap T_A^{-1}B\big) &\leq \int_0^t \mu_A\big(\{\mu(A)\,\varphi_A > s\} \cap T_A^{-1}B\big)\,ds \\ &\leq \mu\big(\{\mu(A)\,\varphi_A \leq t\} \cap T_A^{-1}B\big) + \mu(A). \end{aligned} \tag{4.2}$$

In particular,

$$\mu\big(T_A^{-1}B\big) \leq \int_0^\infty \mu_A\big(\{\mu(A)\,\varphi_A > s\} \cap T_A^{-1}B\big)\,ds \leq \mu\big(T_A^{-1}B\big) + \mu(A). \tag{4.3}$$

PROOF. **(i)** We let $\widehat{T}$ denote the transfer operator of T w.r.t. μ, so that $\int f \cdot \widehat{T}u\,d\mu = \int (f \circ T) \cdot u\,d\mu$ for $f \in L_\infty(\mu)$ and $u \in L_1(\mu)$. Note that for any $C \in \mathcal{A}$ we have $1_C = \widehat{T}1_{T^{-1}C}$ a.e. Writing $A(0) := A$ and $A(n) := A^c \cap \{\varphi_A = n\}$ for $n \geq 1$, we therefore see that $1_{A(n)} = \widehat{T} 1_{A \cap \{\varphi_A = n+1\}} + \widehat{T} 1_{A(n+1)}$ a.e. for $n \geq 0$. Repeated application yields

$$\widehat{T}^n 1_{\{\varphi_A = n\}} = \sum_{k \geq n} \widehat{T}^k 1_{A \cap \{\varphi_A = k\}} \quad \text{a.e.} \quad \text{for } n \geq 1, \tag{4.4}$$

since $\mu(A(n)) \searrow 0$ (compare e.g. equation (2.3) of [**11**]). This entails (4.1), as

$$\begin{aligned} \mu(\{\varphi_A = n\} \cap T^{-n} B) &= \int_B \widehat{T}^n 1_{\{\varphi_A = n\}} \, d\mu \\ &= \sum_{k \geq n} \int_B \widehat{T}^k 1_{A \cap \{\varphi_A = k\}} \, d\mu = \sum_{k \geq n} \mu(A \cap \{\varphi_A = k\} \cap T^{-n} B). \end{aligned}$$

(ii) Now take any $t > 0$. Using (4.1) and the fact that φ_A is integer-valued, we first obtain

$$\begin{aligned} \mu\big(\{\mu(A)\, \varphi_A \leq t\} \cap T_A^{-1} B\big) &= \sum_{n=1}^{\lfloor t/\mu(A) \rfloor} \mu(\{\varphi_A = n\} \cap T_A^{-1} B) \\ &= \sum_{n=1}^{\lfloor t/\mu(A) \rfloor} \mu(A \cap \{\varphi_A \geq n\} \cap T_A^{-1} B) \\ &= \int_0^{\lfloor t/\mu(A) \rfloor} \mu(A \cap \{\varphi_A \geq r\} \cap T_A^{-1} B) \, dr. \end{aligned}$$

On the other hand, an obvious change of variable yields

$$\begin{aligned} \int_0^t \mu_A\big(\{\mu(A)\, \varphi_A > s\} \cap T_A^{-1} B\big) \, ds &= \frac{1}{\mu(A)} \int_0^t \mu(A \cap \{\mu(A) \varphi_A \geq s\} \cap T_A^{-1} B) \, ds \\ &= \int_0^{t/\mu(A)} \mu(A \cap \{\varphi_A \geq r\} \cap T_A^{-1} B) \, dr. \end{aligned}$$

Since $0 \leq \int_{\lfloor t/\mu(A) \rfloor}^{t/\mu(A)} \mu(A \cap \{\varphi_A \geq r\}) \, dr \leq 1 \cdot \mu(A)$, the inequality (4.2) follows. Finally, letting $t \to \infty$, we obtain (4.3). □

Remark 4.1. *The preceding proof did not require the invariant measure to be finite. It applies without change whenever $(X, \mathcal{A}, \mu, T)$ is a conservative ergodic measure preserving system, and $A, B \in \mathcal{A}$ with $B \subseteq A$, and $0 < \mu(B) \leq \mu(A) < \infty$.*

To deal with integrated tail probabilities like those appearing in (4.2), we use an analytical observation. (The statement to follow is a bit stronger than what is actually used in the proof of the theorem.)

Lemma 4.2 (**Convergence of integrated tails**). *Let $(\widetilde{F}_l)_{l \geq 1}$ be a sequence of sub-probability distribution functions on $[0, \infty)$. Let $\widetilde{F}_l(\infty^-) := \lim_{t \to \infty} \widetilde{F}_l(t)$, and assume that $\widetilde{F}_l(\infty^-) \to \mathfrak{f} \in [0, 1]$ as $l \to \infty$.*

a) *If F is some non-decreasing function on $[0, \infty)$ such that*

$$\textstyle\int_0^t [\widetilde{F}_l(\infty^-) - \widetilde{F}_l(s)] \, ds \longrightarrow F(t) \quad \text{as } l \to \infty \tag{4.5}$$

on a set of points t which is dense in $(0, \infty)$, then F is continuous, and

$$\widetilde{F}_l \Longrightarrow \widetilde{F} \quad \text{as } l \to \infty \tag{4.6}$$

for a sub-probability distribution function $\widetilde{F}$ *which is uniquely characterized by*

$$\int_0^t [\mathfrak{f} - \widetilde{F}(s)]\,ds = F(t) \quad \text{for } t \geq 0. \tag{4.7}$$

Moreover, if F *is bounded, then* $\widetilde{F}(\infty^-) = \mathfrak{f}$.

b) *Conversely, if* $\widetilde{F}_l \Longrightarrow \widetilde{F}$ *for some sub-probability distribution function* $\widetilde{F}$*, then (4.5) holds for the continuous non-decreasing function* F *on* $[0, \infty)$ *given by (4.7).*

Proof. **a)** By Helly's selection theorem, every subsequence of $(\widetilde{F}_l)_{l\geq 1}$ has weak accumulation points. Let the sub-probability distribution function $\widetilde{F}$ be any of these, so that $\widetilde{F}_{l_j} \Longrightarrow \widetilde{F}$ as $j \to \infty$ for a suitable subsequence $l_j \nearrow \infty$ of indices. Since $0 \leq \widetilde{F}_l \leq 1$, dominated convergence ensures that, for every $t \geq 0$,

$$\int_0^t [\widetilde{F}_l(\infty^-) - \widetilde{F}_{l_j}(s)]\,ds \longrightarrow \int_0^t [\mathfrak{f} - \widetilde{F}(s)]\,ds \quad \text{as } j \to \infty.$$

We thus see, using continuity of the indefinite integral, and monotonicity of F, that

$$\int_0^t [\mathfrak{f} - \widetilde{F}(s)]\,ds = F(t) \quad \text{for } all\ t \geq 0. \tag{4.8}$$

In particular, F is continuous. Being right-continuous, $\widetilde{F}$ is uniquely determined by relation (4.8). Thus, all accumulation points $\widetilde{F}$ of $(\widetilde{F}_l)_{l\geq 1}$ coincide, which (in view of Helly's theorem) proves weak convergence $\widetilde{F}_l \Longrightarrow \widetilde{F}$ to this unique $\widetilde{F}$. If F is bounded, it is clear that $\mathfrak{f} - \widetilde{F}(s) \to 0$ as $s \to \infty$, hence $\widetilde{F}(\infty^-) = \mathfrak{f}$.

b) By dominated convergence again, $\widetilde{F}_l \Longrightarrow \widetilde{F}$ implies that for all $t \geq 0$,

$$\int_0^t [\widetilde{F}_l(\infty^-) - \widetilde{F}_l(s)]\,ds \longrightarrow \int_0^t [\mathfrak{f} - \widetilde{F}(s)]\,ds =: F(t) \in [0, \infty) \quad \text{as } l \to \infty,$$

and the function F thus defined is non-decreasing and continuous. □

The main step in the proof of the theorem relates the two limit processes in situations where both sequences converge. For the sake of clarity, we first recall an elementary fact.

Remark 4.2. *Let* Q *be a finite Borel measure on* $\mathbb{R}^{d+1}$ *with distribution function* $F^{[d+1]}$*, and let* $\pi_0, \pi_1, \ldots, \pi_d : \mathbb{R}^{d+1} \to \mathbb{R}$ *denote the canonical coordinate projections. Then, for each* $j \in \{0, 1, \ldots, d\}$*, the set* $D_j := \{t \in \mathbb{R} : Q[\pi_j = t] > 0\}$ *is countable, and* F *is continuous outside* $\bigcup_{j=0}^d \bigcup_{t\in D_j} \{\pi_j = t\}$.

Now, the crucial step of our argument is contained in

Lemma 4.3 (**Relating the two limit processes**). *Let* $(X, \mathcal{A}, \mu, T)$ *be an ergodic probability preserving system, and* $(A_l)_{l\geq 1}$ *a sequence of asymptotically rare events. Assume that*

$$\mu(A_l)\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi \quad \text{for some random sequence } \Phi \text{ in } [0, \infty], \text{ and} \tag{4.9}$$

$$\mu(A_l)\widetilde{\Phi}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi} \quad \text{for some random sequence } \widetilde{\Phi} \text{ in } [0, \infty]. \tag{4.10}$$

Then the sub-probability distribution functions $F^{[d]}$ *and* $\widetilde{F}^{[d]}$ *of* $\Phi^{[d]}$ *and* $\widetilde{\Phi}^{[d]}$*, respectively, satisfy, for any* $d \geq 0$ *and* $t_j \geq 0$,

$$\int_0^{t_0} [\widetilde{F}^{[d]}(t_1, \ldots, t_d) - \widetilde{F}^{[d+1]}(s, t_1, \ldots, t_d)]\,ds = F^{[d+1]}(t_0, t_1, \ldots, t_d), \tag{4.11}$$

where we let $F^{[d]} = \widetilde{F}^{[d]} := 1$. *Through this system of integral equations, each of the families* $\{F^{[d]}\}_{d\geq 0}$ *and* $\{\widetilde{F}^{[d]}\}_{d\geq 0}$ *uniquely determines the other.*

Proof. **(i)** Due to our assumptions (4.9) and (4.10), we have, for all $d \geq 1$,

$$(\Diamond_d) \qquad \mu(A_l)\Phi^{[d]}_{A_l} \overset{\mu}{\Longrightarrow} \Phi^{[d]} \quad \text{as } l \to \infty, \quad \text{and}$$

$$(\widetilde{\Diamond}_d) \qquad \mu(A_l)\widetilde{\Phi}^{[d]}_{A_l} \overset{\mu_{A_l}}{\Longrightarrow} \widetilde{\Phi}^{[d]} \quad \text{as } l \to \infty.$$

Note that Theorem B contains the $d = 0$ case of (4.11). From now on, we fix some $d \geq 1$. Our goal is to prove (4.11) in this case. In step (ii) below we will do so for $(t_1, \dots, t_d)$ from a large set $M \subseteq (0,\infty)^d$. To be definite, set $D_j := \{t \in (0,\infty) : \Pr[\pi_j \Phi = t] > 0\}$ and $\widetilde{D}_j := \{t \in (0,\infty) : \Pr[\pi_{j-1}\widetilde{\Phi} = t] > 0\}$, $\widetilde{D}_0 := \varnothing$. Let $M := \{(t_1, \dots, t_d) : t_j \notin D_j \cup \widetilde{D}_j \text{ for } 0 \leq j \leq d\}$ (so that we remove countably many hyperplanes).

Once (4.11) holds for $(t_0, t_1, \dots, t_d) \in [0,\infty) \times M$, we need only observe that this set is dense in $[0,\infty)^{d+1}$, and that both sides of (4.11) define right-continuous functions on this set. (For the integral this is an easy consequence of dominated convergence.) It is then clear that (4.11) holds for all $(t_0, t_1, \dots, t_d) \in [0,\infty)^{d+1}$.

Regarding uniqueness, it is then immediate from (4.11) that $F^{[d+1]}$ is determined by $\widetilde{F}^{[d]}$ and $\widetilde{F}^{[d+1]}$. Conversely, step (ii) will show that, for $(t_1, \dots, t_d) \in M$, $\widetilde{F}^{[d+1]}(t_0, t_1, \dots, t_d)$ is uniquely determined by $\widetilde{F}^{[d]}$ and $F^{[d+1]}$. By right-continuity, this extends to all $(t_0, t_1, \dots, t_d) \in [0,\infty)^{d+1}$. The uniqueness statement follows by an obvious induction, as Theorem B ensures that $\widetilde{F}^{[1]}$ is determined by $F^{[1]}$.

If $(t_0, t_1, \dots, t_d) \in D_0^c \times M$, then, according to Remark 4.2, $(t_0, t_1, \dots, t_d)$ is a continuity point for both $F^{[d+1]}$ and $\widetilde{F}^{[d+1]}$. Moreover, $(t_1, \dots, t_d)$ is a continuity point for $\widetilde{F}^{[d]}$, because $\widetilde{\Phi}$ is stationary by Proposition 2.1.

(ii) We fix any $(t_1, \dots, t_d) \in M$. Write, for $t \in [0,\infty)$,

$$F_l(t) := \mu(\mu(A_l)\Phi^{[d+1]}_{A_l} \leq (t, t_1, \dots, t_d)), \quad F(t) := F^{[d+1]}(t, t_1, \dots, t_d),$$
$$\widetilde{F}_l(t) := \mu_{A_l}(\mu(A_l)\widetilde{\Phi}^{[d+1]}_{A_l} \leq (t, t_1, \dots, t_d)), \quad \widetilde{F}(t) := \widetilde{F}^{[d+1]}(t, t_1, \dots, t_d),$$

which defines a family of sub-probability distribution functions on $[0,\infty)$. Since φ_{A_l} is finite a.e. and $\widetilde{\Phi}_{A_l}$ is a stationary sequence w.r.t. μ_{A_l}, we have

$$\widetilde{F}_l(\infty^-) := \lim_{t\to\infty} \widetilde{F}_l(t) = \mu_{A_l}(\mu(A_l)\widetilde{\Phi}^{[d]}_{A_l} \leq (t_1, \dots, t_d)). \tag{4.12}$$

Since $(t_1, \dots, t_d)$ is a continuity point of $\widetilde{F}^{[d]}$, $(\widetilde{\Diamond}_d)$ ensures

$$\widetilde{F}_l(\infty^-) \longrightarrow \widetilde{F}^{[d]}(t_1, \dots, t_d) =: \mathfrak{f} \in [0,1] \quad \text{as } l \to \infty. \tag{4.13}$$

Similarly, for any $t_0 \in D_0^c$, $(t_0, t_1, \dots, t_d)$ is a continuity point of both $F^{[d+1]}$ and $\widetilde{F}^{[d+1]}$, so that $(\Diamond_{d+1})$ and $(\widetilde{\Diamond}_{d+1})$ respectively imply

$$F_l(t_0) \longrightarrow F(t_0), \quad \widetilde{F}_l(t_0) \longrightarrow \widetilde{F}(t_0) \quad \text{for } t_0 \in D_0^c \text{ as } l \to \infty. \tag{4.14}$$

In particular, $F_l \Longrightarrow F$ and $\widetilde{F}_l \Longrightarrow \widetilde{F}$. Next, defining

$$B_l := A_l \cap \{\mu(A_l)\Phi^{[d]}_{A_l} \leq (t_1, \dots, t_d)\} = \{\mu(A_l)\widetilde{\Phi}^{[d]}_{A_l} \leq (t_1, \dots, t_d)\}, \quad l \geq 1, \tag{4.15}$$

we see that for any $t \in [0,\infty)$,

$$\{\mu(A_l)\Phi^{[d+1]}_{A_l} \leq (t, t_1, \dots, t_d)\} = \{\mu(A_l)\,\varphi_{A_l} \leq t\} \cap T^{-1}_{A_l} B_l, \tag{4.16}$$

with $T_{A_l}: X \to A_l$ the first entrance map of A_l. Using $\{\mu(A)\,\varphi_A > s\} \cap T_A^{-1}B = T_A^{-1}B \setminus [\{\mu(A)\,\varphi_A \leq s\} \cap T_A^{-1}B]$, restricting to A_l, and recalling (4.12) we get

$$\mu_{A_l}\big(\{\mu(A_l)\,\varphi_{A_l} > s\} \cap T_{A_l}^{-1}B_l\big) = \widetilde{F}_l(\infty^-) - \widetilde{F}_l(s). \tag{4.17}$$

Therefore, the crucial estimate (4.2) of Lemma 4.1 shows that

$$F_l(t) \leq \int_0^t [\widetilde{F}_l(\infty^-) - \widetilde{F}_l(s)]\,ds \leq F_l(t) + \mu(A_l) \quad \text{for } l \geq 1 \text{ and } t > 0. \tag{4.18}$$

In view of $\mu(A_l) \to 0$, (4.14) and (4.18) together prove that

$$\int_0^{t_0} [\widetilde{F}_l(\infty^-) - \widetilde{F}_l(s)]\,ds \longrightarrow F(t_0) \quad \text{as } l \to \infty \text{ for } t_0 \in D_0^c. \tag{4.19}$$

Since this applies for all t_0 in the dense set $D_0^c \subseteq (0,\infty)$, we can appeal to Lemma 4.2 a) to conclude that the weak limit $\widetilde{F}$ of $(\widetilde{F}_l)$ is the sub-probability d.f. unambiguously characterized by

$$\int_0^t [\mathfrak{f} - \widetilde{F}(s)]\,ds = F(t) \quad \text{for } t \geq 0. \tag{4.20}$$

In particular, this proves (4.11) for $(t_0, t_1, \ldots, t_d) \in [0,\infty) \times M$, as required. □

The next observation will enable us to reduce everything to the situation of Lemma 4.3.

LEMMA 4.4 (**Sequential precompactness**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system, and $(A_l)_{l \geq 1}$ any sequence in $\mathcal{A}$ with $\mu(A_l) > 0$ for all $l \geq 1$. Then there are indices $l_j \nearrow \infty$ and random elements $\Phi, \widetilde{\Phi}$ of $[0,\infty]^{\mathbb{N}_0}$ such that*

$$\mu(A_{l_j})\Phi_{A_{l_j}} \overset{\mu}{\Longrightarrow} \Phi \quad \textit{and} \quad \mu(A_{l_j})\widetilde{\Phi}_{A_{l_j}} \overset{\mu_{A_{l_j}}}{\Longrightarrow} \widetilde{\Phi} \quad \textit{as } j \to \infty. \tag{4.21}$$

PROOF. Given any integer $d \geq 1$, the d-dimensional version of Helly's selection theorem shows that any sequence $m_i \nearrow \infty$ of indices contains a further subsequence $l_j \nearrow \infty$, $l_j = m_{i_j}$, such that there are random elements $\Phi_*^{[d]}$ and $\widetilde{\Phi}_*^{[d]}$ of $[0,\infty]^{\mathbb{N}_0}$ for which

$$\mu(A_{l_j})\Phi_{A_{l_j}}^{[d]} \overset{\mu}{\Longrightarrow} \Phi_*^{[d]} \quad \text{and} \quad \mu(A_{l_j})\widetilde{\Phi}_{A_{l_j}}^{[d]} \overset{\mu_{A_{l_j}}}{\Longrightarrow} \widetilde{\Phi}_*^{[d]} \quad \text{as } j \to \infty. \tag{4.22}$$

(First, do it for the $\mu(A_l)\Phi_{A_l}^{[d]}$, and then apply it again to the subsequence thus obtained, to also take care of the $\mu(A_l)\widetilde{\Phi}_{A_l}^{[d]}$.)

Use this in a straightforward diagonalization argument, to provide $l_j \nearrow \infty$ such that (4.22) holds for all $d \geq 1$. But then the distributions of the $\Phi_*^{[d]}$ form a consistent family, so that by Kolmogorov's existence theorem there is some random element Φ of $[0,\infty]^{\mathbb{N}_0}$ satisfying $\Phi^{[d]} = \Phi_*^{[d]}$ in law for every $d \geq 1$, whence $\mu(A_{l_j})\Phi_{A_{l_j}} \overset{\mu}{\Longrightarrow} \Phi$. The same argument applies to the $\widetilde{\Phi}_*^{[d]}$. □

It is now easy to wrap things up.

Proof of Theorem 3.1. **(i)** We check that (3.4) is equivalent to (3.5), and that each implies (3.6). Assume (3.5), and let $\widetilde{F}^{[d]}$ be the distribution function of $\widetilde{\Phi}^{[d]}$, $d \geq 1$. Note first that by Lemmas 4.4 and 4.3 there is a random sequence Φ whose finite-dimensional distribution functions $F^{[d]}$ form the unique family $\{F^{[d]}\}$ related to $\{\widetilde{F}^{[d]}\}$ via (4.11). Now suppose, for a contradiction, that (3.4) fails. Then there is some $d \geq 1$ such that the $F_l^{[d]}(t_0, \ldots, t_{d-1}) := \mu(\mu(A_l)\Phi_{A_l}^{[d]} \leq (t_0, \ldots, t_{d-1}))$ fail to converge weakly to $F^{[d]}$. Hence, for some $\varepsilon > 0$ and $l_j \nearrow \infty$, we have $\mathrm{dist}^{[d]}(F_{l_j}^{[d]}, F^{[d]}) \geq \varepsilon$ for all $j \geq 1$, where $\mathrm{dist}^{[d]}$ denotes Lévy-distance. But then we can once again appeal to Lemmas 4.4 and 4.3 to obtain a further subsequence $m_i = l_{j_i} \nearrow \infty$ s.t. $F_{m_i}^{[d]} \Longrightarrow F^{[d]}$ nonetheless. This contradiction proves (3.4).

The same argument works if we start from (3.4).

(ii) To validate the finiteness assertion for $\sigma\Phi$, fix $d \geq 2$. Since $\widetilde{\Phi}$ is finite a.s. by Proposition 2.1, we have $\widetilde{F}^{[d]}(r, \ldots, r) \to 1$ and, for any $s \in [0, \infty)$, $\widetilde{F}^{[d+1]}(s, r, \ldots, r) \to \widetilde{F}^{[1]}(s)$ as $r \to \infty$. Due to (3.6) and dominated convergence,

$$\begin{aligned}
\Pr[\varphi^{(0)} < \infty, \ldots, \varphi^{(d)} < \infty] &= \lim_{t\to\infty} \lim_{r\to\infty} F^{[d+1]}(t, r, \ldots, r) \\
&= \lim_{t\to\infty} \lim_{r\to\infty} \textstyle\int_0^t [\widetilde{F}^{[d]}(r, \ldots, r) - \widetilde{F}^{[d+1]}(s, r, \ldots, r)]\, ds \\
&= \textstyle\int_0^\infty [1 - \widetilde{F}^{[1]}(s)]\, ds = \Pr[\varphi^{(0)} < \infty],
\end{aligned}$$

where the last step uses Theorem B. This concludes the proof. □

5. Proof of Propositions 3.2 - 3.4 and Theorem 3.2

Now for the (easy) proofs of the advertised consequences of Theorem 3.1.

Proof of Proposition 3.2. The characterization is immediate from (3.6). Elementary arguments confirm that stationary finite-state Markov chains $\widetilde{\Phi}$ typically violate (3.7). □

Proof of Proposition 3.3. **(i)** Suppose first that $\widetilde{\Phi}$ is an independent sequence (necessarily stationary and a.s. real-valued). Let $\widetilde{F} := \widetilde{F}^{[1]}$, then (3.6) immediately shows that

$$F^{[d+1]}(t_0, \ldots, t_d) = \left(\int_0^{t_0} [1 - \widetilde{F}(s)]\, ds\right) \widetilde{F}(t_1) \cdots \widetilde{F}(t_d). \tag{5.1}$$

In view of the product form of $F^{[d+1]}$, it is clear that Φ, when conditioned on $\{\varphi^{(0)} < \infty\}$, is (a.s. real-valued and) independent, too.

(ii) Suppose that Φ conditioned on $\{\varphi^{(0)} < \infty\}$ is an (a.s. real-valued) independent sequence. Then $F^{[d+1]}(t_0, \ldots, t_d) = (\int_0^\infty [1 - \widetilde{F}(s)]\, ds)\, F^{(0)}(t_0) \cdots F^{(d)}(t_d)$ for suitable probability distribution functions $F^{(j)}$. Note that (1.4) allows us to differentiate $F^{(0)}$, and gives $(F^{(0)})'(t_0) = (\int_0^\infty [1 - \widetilde{F}(s)]\, ds)^{-1}(1 - \widetilde{F}(t_0))$ for a.e. $t_0 \geq 0$. In view of (3.6), we have

$$\int_0^{t_0} [\widetilde{F}^{[d+1]}(\infty, t_1, \ldots, t_d) - \widetilde{F}^{[d+1]}(s, t_1, \ldots, t_d)]\, ds = \int_0^\infty [1 - \widetilde{F}(s)]\, ds \prod_{j=0}^{d} F^{(j)}(t_j).$$

Now fix any $t_1, \ldots, t_d \geq 0$. Differentiating, we see that for a.e. $t_0 \geq 0$,

$$\widetilde{F}^{[d+1]}(\infty, t_1, \ldots, t_d) - \widetilde{F}^{[d+1]}(t_0, t_1, \ldots, t_d) = [1 - \widetilde{F}(t_0)]\, F^{(1)}(t_1) \cdots F^{(d)}(t_d).$$

As an immediate consequence we see that the Borel probability measure on $[0, \infty)^{d+1}$ with distribution function $\widetilde{F}^{[d+1]}$ is a product measure. □

Proof of Proposition 3.4. Using (3.8) it is straightforward to see that Φ_{Exp} satisfies (3.9).

Conversely, assume (3.9). The $d = 0$ case shows, via (3.8), that the one-dimensional marginals of $\Phi(\varphi^{(0)}, \varphi^{(1)}, \ldots)$ are exponentially distributed. We are going to show that for every $d \geq 1$, $\varphi^{(0)}$ is independent of $(\varphi^{(1)}, \ldots, \varphi^{(d)})$. Due to stationarity this implies that Φ is iid.

Fix $d \geq 1$. We check that for any $(t_1, \ldots, t_d) \in [0, \infty)^d$ (henceforth fixed) for which $\Pr[(\varphi^{(1)}, \ldots, \varphi^{(d)}) \leq (t_1, \ldots, t_d)] > 0$, we have

$$\Pr[\varphi^{(0)} \leq t] = \Pr[\varphi^{(0)} \leq t \mid (\varphi^{(1)}, \ldots, \varphi^{(d)}) \leq (t_1, \ldots, t_d)], \tag{5.2}$$

for $t \geq 0$. Let $F(t)$ denote the right-hand side of (5.2), then (3.9) means that F satisfies $\int_0^t [1 - F(s)]\, ds = F(t)$. Hence, by (3.8), $F = F_{\mathrm{Exp}}$, and this is indeed the distribution function of $\varphi^{(0)}$. □

The auxiliary results developed in the previous section also lead to a natural

Proof of Theorem 3.2. Equivalence of (3.11) and (3.12) is immediate from Theorem 3.1 and Proposition 3.4. It is also straightforward that these imply (3.10).

We thus turn to the nontrivial statement that (3.10) implies (3.11) and (3.12). In view of Lemma 4.4 we need only check that each accumulation point Φ coincides with Φ_{Exp}. (Use the subsequence-in-subsequence argument of the previous proof.)

Now take any sequence $l_j \nearrow \infty$ of indices as in Lemma 4.4. As (A_l) is asymptotically rare, Lemma 4.3 applies to this subsequence, showing that (4.11) holds for all $d \geq 0$ and $t_j \geq 0$. But for every d we have

$$F_{l_j}^{[d]} \Longrightarrow F^{[d]} \quad \text{and} \quad \widetilde{F}_{l_j}^{[d]} \Longrightarrow \widetilde{F}^{[d]}.$$

Therefore, and since all these functions are right-continuous, (3.10) implies $\widetilde{F}^{[d]} = F^{[d]}$. This shows that $\Phi = \widetilde{\Phi}$ (in distribution). According to Lemma 2.1, this limit process is a.s. real-valued. Therefore, Proposition 3.4 applies to show that $\Phi = \Phi_{\mathrm{Exp}}$. □

6. Proof of Theorem 2.2 and Proposition 2.2

We will mainly work with the hitting-time process, and first establish an estimate for the finite-dimensional hitting-time distributions of small perturbations of a fixed set A.

Lemma 6.1 (**Robustness of individual d-dim hitting-time distributions**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability-preserving system and $A \in \mathcal{A}$, $\mu(A) > 0$. If $d \geq 1$, $N \geq 1$, and $B \in \mathcal{A}$ satisfy $\mu(B) < \mu(A)/dN^2$, then, for $t_j \in [0, N]$,*

$$F_A^{[d]}\left(\frac{N}{N+1}(t_0, \ldots, t_{d-1})\right) - \frac{1}{N} \leq F_{A\cup B}^{[d]}(t_0, \ldots, t_{d-1}) \leq F_A^{[d]}(t_0, \ldots, t_{d-1}) + \frac{1}{N}.$$

PROOF. Throughout, we fix d, N, and B as in the statement.

(i) We first show that for $t'_0, \ldots, t'_{d-1} \in [0, N]$,

$$\left\{\mu(A)\left(\Phi^{[d]}_{A\cup B} \wedge \Phi^{[d]}_{A}\right) \le (t'_0, \ldots, t'_{d-1})\right\} \cap \left\{\Phi^{[d]}_{A\cup B} \neq \Phi^{[d]}_{A}\right\} \subseteq C, \tag{6.1}$$

where

$$C := \{\mu(A)\varphi_B \le N\} \cup \bigcup_{j=1}^{d-1} \left(\{\mu(A)\varphi_A \le N\} \cap T^{-(j+1)}_{A\cup B} B\right).$$

To this end, take any $x \in \{\Phi^{[d]}_{A\cup B} \neq \Phi^{[d]}_{A}\}$. Then there is some $j \in \{0, \ldots, d-1\}$ such that

$$\varphi_{A\cup B}(x) = \varphi_A(x), \ldots, \varphi_{A\cup B}(T^{j-1}_{A\cup B}\, x) = \varphi_A(T^{j-1}_A x), \tag{6.2}$$

while $\varphi_{A\cup B}(T^{j}_{A\cup B}\, x) \neq \varphi_A(T^{j}_A\, x)$. Due to (6.2) we have $T^{j}_{A\cup B}\, x = T^{j}_A\, x$. Since $\varphi_{A\cup B} = \varphi_A \wedge \varphi_B \le \varphi_A$ we thus see that

$$\varphi_{A\cup B}(T^{j}_{A\cup B}\, x) = \varphi_{A\cup B}(T^{j}_A\, x) = \varphi_B(T^{j}_A\, x) < \varphi_A(T^{j}_A\, x). \tag{6.3}$$

In particular,

$$x \in T^{-(j+1)}_{A\cup B}\,(B \setminus A) \subseteq T^{-(j+1)}_{A\cup B}\,(B). \tag{6.4}$$

Now assume that also $x \in \{\mu(A)(\Phi^{[d]}_{A\cup B} \wedge \Phi^{[d]}_{A}) \le (t'_0, \ldots, t'_{d-1})\}$. As a consequence of (6.3) we then see that

$$\text{if } j = 0, \text{ then } x \in \{\mu(A)\varphi_B \le N\}, \tag{6.5}$$

while otherwise $\varphi_{A\cup B}(x) = \varphi_A(x)$, so that

$$\text{if } j > 0, \text{ then } x \in \{\mu(A)\varphi_A \le N\}. \tag{6.6}$$

Combining (6.4) with (6.5) and (6.6) shows that $x \in C$, as claimed.

(ii) Now fix $t_0, \ldots, t_{d-1} \in [0, N]$. In view of (6.1) we may control the event

$$E := \left\{\mu(A\cup B)\Phi^{[d]}_{A\cup B} \le (t_0, \ldots, t_{d-1})\right\}$$

as follows. First,

$$\begin{aligned} E &\subseteq \left(\{\mu(A)\Phi^{[d]}_{A\cup B} \le (t_0, \ldots, t_{d-1})\} \cap \{\Phi^{[d]}_{A\cup B} = \Phi^{[d]}_{A}\}\right) \\ &\quad \cup \left(\{\mu(A)\Phi^{[d]}_{A\cup B} \le (t_0, \ldots, t_{d-1})\} \cap \{\Phi^{[d]}_{A\cup B} \neq \Phi^{[d]}_{A}\}\right) \\ &\subseteq \{\mu(A)\Phi^{[d]}_{A} \le (t_0, \ldots, t_{d-1})\} \cup C, \end{aligned}$$

so that

$$F^{[d]}_{A\cup B}(t_0, \ldots, t_{d-1}) = \mu(E) \le F^{[d]}_{A}(t_0, \ldots, t_{d-1}) + \mu(C). \tag{6.7}$$

On the other hand,

$$\begin{aligned} E &\supseteq \left\{\mu(A)\Phi^{[d]}_{A} \le \frac{\mu(A)}{\mu(A\cup B)}\,(t_0, \ldots, t_{d-1})\right\} \cap \{\Phi^{[d]}_{A\cup B} = \Phi^{[d]}_{A}\} \\ &\supseteq \left\{\mu(A)\Phi^{[d]}_{A} \le \frac{\mu(A)}{\mu(A\cup B)}\,(t_0, \ldots, t_{d-1})\right\} \setminus C, \end{aligned}$$

and hence, since, by our condition, $\mu(A)/\mu(A\cup B) > N/(N+1)$,

$$F^{[d]}_{A\cup B}(t_0, \ldots, t_{d-1}) = \mu(E) \ge F^{[d]}_{A}\left(\frac{N}{N+1}\,(t_0, \ldots, t_{d-1})\right) - \mu(C). \tag{6.8}$$

(Note that although $\varphi_{A\cup B} \le \varphi_A$, its d-dimensional version $\Phi^{[d]}_{A\cup B} \le \Phi^{[d]}_A$ is *not* true in general.)

(iii) To conclude, we are going to estimate $\mu(C)$. Set $M := N/\mu(A)$, and observe first that $\{\varphi_B \le M\} = \bigcup_{k=1}^{\lfloor M\rfloor}\{\varphi_B = k\} \subseteq \bigcup_{k=1}^{\lfloor M\rfloor} T^{-k}B$. As T preserves μ, this gives

$$\mu(\varphi_B \le M) \le M\,\mu(B). \tag{6.9}$$

Note then that for any $j \ge 0$,

$$\begin{aligned}\{\varphi_A \le M\} \cap T^{-(j+1)}_{A\cup B}B &\subseteq \{\varphi_{A\cup B} \le M\} \cap T^{-(j+1)}_{A\cup B}B \\ &= \bigcup_{k=1}^{\lfloor M\rfloor}\{\varphi_{A\cup B} = k\} \cap T^{-1}_{A\cup B}\left(T^{-j}_{A\cup B}B\right) \\ &= \bigcup_{k=1}^{\lfloor M\rfloor}\{\varphi_{A\cup B} = k\} \cap T^{-k}\left(D_j\right),\end{aligned}$$

where $D_j := (A\cup B)\cap T^{-j}_{A\cup B}B$ (recall that $T_{A\cup B}$ is defined on all of X). Hence, $\mu(\{\varphi_A \le M\}\cap T^{-(j+1)}_{A\cup B}B) \le \sum_{k=1}^{\lfloor M\rfloor}\mu\left(T^{-k}\left(D_j\right)\right)$. But since $T_{A\cup B}$ preserves $\mu\mid_{A\cup B}$, we have $\mu(D_j) = \mu(B)$ for all $j \ge 0$, and therefore $\mu\left(T^{-k}\left(D_j\right)\right) = \mu(B)$ by T-invariance of μ. We thus end up with

$$\mu\left(\{\varphi_A \le M\} \cap T^{-(j+1)}_{A\cup B}B\right) \le M\,\mu(B). \tag{6.10}$$

Together, (6.9) and (6.10) yield $\mu(C) \le dM\,\mu(B) < 1/N$. When combined with (6.7) and (6.8), this proves the assertion of the lemma. □

The proof of the following auxiliary statement is a routine exercise.

Lemma 6.2. *For any fixed $d \ge 1$, let $F^{[d]}_l, G^{[d]}_l$, and $F^{[d]}$ be sub-probability distribution functions on $[0,\infty)^d$ such that for every $N \ge 1$ there is some l_N such that for $l \ge l_N$ we have*

$$F^{[d]}_l\left(\frac{N}{N+1}\,t\right) - \frac{1}{N} \le G^{[d]}_l(t) \le F^{[d]}_l(t) + \frac{1}{N} \qquad \text{for } t \in [0,N].$$

Then $F^{[d]}_l \Longrightarrow F^{[d]}$ iff $G^{[d]}_l \Longrightarrow F^{[d]}$.

We can then complete the proof of the theorem.

Proof of Theorem 2.2. **(i)** Consider $A_l, B_l \in \mathcal{A}$ with $\mu(B_l) = o(\mu(A_l))$. For any $d \ge 1$ let $F^{[d]}_l := F^{[d]}_{A_l}$ and $G^{[d]}_l := F^{[d]}_{A_l\cup B_l}$. Then Lemma 6.1 shows that $(F^{[d]}_l)$ and $(G^{[d]}_l)$ satisfy the assumption of Lemma 6.2. Therefore, if $F^{[d]}$ is any sub-probability distribution function on $[0,\infty)^d$, then $F^{[d]}_{A_l} \Longrightarrow F^{[d]}$ iff $F^{[d]}_{A_l\cup B_l} \Longrightarrow F^{[d]}$. As an immediate consequence, given some random sequence Φ in $[0,\infty]$, we have $\mu(A_l)\,\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi$ iff $\mu(A_l\cup B_l)\,\Phi_{A_l\cup B_l} \overset{\mu}{\Longrightarrow} \Phi$.

(ii) Turning to (A'_l) and (A''_l), let $A_l := A'_l \cap A''_l$. Then $A'_l = A_l \cup B'_l$ with $B'_l := A'_l \setminus A_l$, and $A''_l = A_l \cup B''_l$ with $B''_l := A''_l \setminus A_l$. Observe that $\mu(B'_l), \mu(B''_l) = o(\mu(A_l))$ as $l \to \infty$. Therefore step (i) ensures that both $\mu(A'_l)\,\Phi_{A'_l} \overset{\mu}{\Longrightarrow} \Phi$ and $\mu(A''_l)\,\Phi_{A''_l} \overset{\mu}{\Longrightarrow} \Phi$ are equivalent to $\mu(A_l)\,\Phi_{A_l} \overset{\mu}{\Longrightarrow} \Phi$.

(iii) If $\Pr[\widetilde{\varphi}^{(0)} > 0] > 0$, statement (2.8) follows from (ii) via Theorem 3.1. If $\widetilde{\varphi}^{(0)} = 0$ a.s., an simple direct argument proves (2.8). □

We conclude this section with the advertised application of this principle.

Proof of Proposition 2.2. **(i)** Passing to a subsequence (if necessary), we may assume w.l.o.g. that $\sum_{l\geq 1}\mu(A_l)<\infty$. We first construct a certain sequence $l_j \nearrow \infty$ of indices. Start with $l_1 := 1$. If l_j has been constructed, consider the sets

$$B_{j,k} := \bigcup_{l\geq l_j+k} A_l, \quad k\geq 1.$$

Then $\mu(B_{j,k}) \searrow 0$ as $k\to\infty$, and we choose $k_j \geq 1$ so large that $\mu(B_{j,k_j}) < \mu(A_{l_j})/j$. Now let $l_{j+1} := l_j + k_j > l_j$. Next, using the l_j, we define

$$A'_j := \bigcup_{i\geq j} A_{l_i}, \quad j\geq 1.$$

This is a nested sequence in $\mathcal{A}$ which satisfies

$$A'_j = A_{l_j} \cup A'_{j+1} \quad \text{and} \quad A'_{j+1} \subseteq B_{j,k_j} \quad \text{for } j\geq 1.$$

Due to our construction, $\mu(A'_{j+1}) < \mu(A_{l_j})/j$, so that $\mu(A'_j \triangle A_{l_j}) = o(\mu(A_{l_j}))$ as $j\to\infty$. Theorem 2.2 thus gives (2.6).

(ii) To validate the statement about the G_i, we may assume (replacing $(G_i)_{i\geq 1}$ by a suitable subsequence if necessary) that $\mu(G_l) = o(\mu(A_l))$. Set $A_l^* := A_l \cup G_l$, then $\mu(A_l \triangle A_l^*) = o(\mu(A_l))$, so that $\mu(A_l^*)\,\widetilde{\Phi}_{A_l^*} \overset{\mu_{A_l^*}}{\Longrightarrow} \widetilde{\Phi}$ by Theorem 2.2. Now repeat the construction of step (i), applying it to (A_l^*) instead of (A_l). □

7. Preparations for the proof of Theorem 2.1

Some approximation lemmas. We first record (without the very elementary proof) a lemma which compares the distribution of a random element of an arbitrary measurable space $(\Omega',\mathcal{F}')$ to its conditional distributions on certain subsets, in the sense of the total variation type norm $\|Q'\| := \sup_{F'\in\mathcal{F}'} |Q'(F')|$ for finite signed measures Q' on $\mathcal{F}'$.

Lemma 7.1 (**Elementary control of conditional distributions**). *Let $(\Omega,\mathcal{F},P)$ be a probability space, $\mathcal{F}_+ := \{F\in\mathcal{F} : P(F)>0\}$, and $\xi:\Omega\to\Omega'$ measurable.*
a) *If $\Omega_0\in\mathcal{F}_+$, then $\left\|P\circ\xi^{-1} - P_{\Omega_0}\circ\xi^{-1}\right\| \leq P(\Omega_0^c)$.*
b) *If $\Omega = \bigcup_n \Omega_n$ is a finite or countable disjoint union in $\mathcal{F}_+$, and Q' a probability measure on $\mathcal{F}'$, then*

$$\left\|P\circ\xi^{-1} - Q'\right\| \leq \sup_n \left\|P_{\Omega_n}\circ\xi^{-1} - Q'\right\|.$$

Next, we check that the prospective limit process in Theorem 2.1 can be approximated by discrete ergodic stationary sequences. In the following, $\mathrm{dist}^{[d]}$ denotes the Prokhorov-distance of Borel probabilities on $\mathbb{R}^d$. When applied to a random sequence, it refers to its d-dimensional marginals.

Lemma 7.2 (**Ergodic rational approximation of a stationary sequence**). *Let $\Theta = (\theta_j)_{j\geq 0}$ be a non-negative stationary process with $\mathbb{E}[\theta_j]\leq 1$. Given any $\varepsilon>0$ and $d\in\mathbb{N}$, there is some bounded stationary process $\Theta' = (\theta'_j)_{j\geq 0}$ with $\mathbb{E}[\theta'_j] = 1$ and values in some finite subset of $(0,\infty)\cap\mathbb{Q}$, such that $\mathrm{dist}^{[d]}(\Theta,\Theta') < \varepsilon$. Moreover, Θ' can be chosen to be ergodic.*

PROOF. **(i)** The following statement can be checked by routine arguments: Let $(\Omega, \mathcal{A}, P)$ be a non-atomic probability space, and $f : \Omega \to [0, \infty)$ measurable with $\int f \, dP \leq 1$. Then there exists a sequence $(f_n)_{n\geq 1}$ of measurable functions with $f_n \to f$ a.e. such that, for each $n \geq 1$, we have $\int f_n \, dP \leq 1$, and $f_n(\Omega)$ is a finite subset of $(0, \infty) \cap \mathbb{Q}$.

The process Θ has a representation as $\theta_j = f \circ S^j$, $j \geq 0$, where S is a measure-preserving map on some probability space $(\Omega, \mathcal{A}, P)$. The latter can w.l.o.g. be assumed to be non-atomic. (If necessary, replace it by $(\Omega \times (0,1], \mathcal{A} \otimes \mathcal{B}_{(0,1]}, P \otimes \lambda^1)$, use the map $(\omega, t) \mapsto (S\omega, t)$, and the observable $(\omega, t) \mapsto f(\omega)$.)

Apply the introductory observation to obtain a sequence (f_n) as specified there. Let $G_n := (f_n, f_n \circ S, \ldots, f_n \circ S^{d-1}) : \Omega \to \mathbb{R}^d$, and $G = (f, f \circ S, \ldots, f \circ S^{d-1})$. Then, $G_n \to G$ a.e. and hence also in distribution. Consequently, there is some $n' \geq 1$ such that $\mathrm{dist}^{[d]}(P \circ G^{-1}, P \circ G_{n'}^{-1}) < \varepsilon/2$. Defining $f' := f_{n'}$ and $\theta'_j := f' \circ S^j$, $j \geq 0$, thus gives a stationary sequence Θ' with the required properties (except, perhaps, ergodicity).

(ii) To show that one can ensure ergodicity, take Θ' as above. Adding some extra randomness, we find an $\varepsilon/2$-close sequence Θ'' which is also ergodic. To this end, define a family of stochastic kernels $\mathbf{P}_q : \Omega \times \mathcal{A} \to [0,1]$, $q \in [0,1]$, by letting $\mathbf{P}_q(\omega, A) := (1-q) 1_A(S\omega) + qP(A)$. Then each $\mathbf{P}_q$ preserves P. Each $\mathbf{P}_q$ defines a stationary Markov chain $(\xi_j^{[q]})_{j \geq 0}$ on $(\Omega, \mathcal{A})$ with $\mathrm{law}(\xi_0^{[q]}) = P$. For $q > 0$ the chain is clearly ergodic, while for $q = 0$ it is given by the process $(S^j)_{j\geq 0}$ on $(\Omega, \mathcal{A}, P)$ used to define Θ'.

Conditioned on a set of probability $1 - q^d$, the chain $(\xi_j^{[q]})_{j=0}^{d-1}$ has the same law as $(\xi_j^{[0]})_{j=0}^{d-1} = (S^j)_{j=0}^{d-1}$. Therefore, if $q > 0$ is sufficiently small, $\theta''_j := f'(\xi_j^{[q]})$, $j \geq 0$, defines an ergodic stationary sequence Θ'' with $\mathrm{dist}^{[d]}(\Theta', \Theta'') < \varepsilon/2$ and range contained in $f'(\Omega)$. □

Convergence of ergodic sums. We will also use the following supplement to the pointwise ergodic theorem.

LEMMA 7.3 (**Ergodic averages converging from below**). *Let $\Psi = (\psi_j)_{j\geq 0}$ be a non-negative ergodic stationary process with $\overline{\psi} := \mathbb{E}[\psi_j] < \infty$. Set $\xi_0 := 0$ and $\xi_r := \psi_0 + \ldots + \psi_{r-1}$, $r \geq 1$, and define $\xi_R^\diamondsuit := \max\{\xi_r : r \in \{0, \ldots, R\}$ and $\xi_r \leq R\overline{\psi}\}$, $R \geq 0$. Then,*

$$R^{-1} \xi_R^\diamondsuit \longrightarrow \overline{\psi} \quad \textit{a.s. and in } L_1 \quad \textit{as } R \to \infty. \tag{7.1}$$

PROOF. Obviously, $0 \leq R^{-1}\xi_R^\diamondsuit \leq \overline{\psi}$ for $R \geq 1$, so that the L_1-statement follows by dominated convergence once we prove the a.s.-statement.

Set $M_R := \max\{r \in \{0, \ldots, R\} : \xi_r \leq R\overline{\psi}\}$, $R \geq 0$, so that $\xi_R^\diamondsuit = \xi_{M_R}$, and note that $M_R \nearrow \infty$ a.s. as $R \to \infty$ (even if $\overline{\psi} = 0$). By the pointwise ergodic theorem, the event $E := \{R^{-1}\xi_R \to \overline{\psi}\} \cap \{M_R \nearrow \infty\}$ has full probability. Moreover, $E \subseteq \{R^{-1}\psi_R \to 0\}$. Fixing an arbitrary $\omega \in E$ we check that $R^{-1}\xi_R^\diamondsuit(\omega) \longrightarrow \overline{\psi}$.

Take $\varepsilon > 0$, and choose R_0 so large that $| R^{-1}\xi_R(\omega) - \overline{\psi} | < \varepsilon$ and $R^{-1}\psi_R(\omega) < \varepsilon$ for $R \geq R_0$. Then there is some $R_1 \geq R_0$ s.t. $M_R(\omega) \geq R_0$ for $R \geq R_1$. We claim

$$0 \leq \overline{\psi} - R^{-1}\xi_R^\diamondsuit(\omega) < \varepsilon \quad \text{for } R \geq R_1. \tag{7.2}$$

If $M_R(\omega) = R$, this is immediate from the definition of R_0. Otherwise, we have

$$\xi_R^\Diamond(\omega) = \xi_{M_R(\omega)}(\omega) \leq R\overline{\psi} < \xi_{M_R(\omega)+1}(\omega) = \xi_{M_R(\omega)}(\omega) + \psi_{M_R(\omega)+1}(\omega),$$

and therefore

$$\overline{\psi} - R^{-1}\xi_R^\Diamond(\omega) < R^{-1}\psi_{M_R(\omega)+1}(\omega) \leq (M_R(\omega)+1)^{-1}\psi_{M_R(\omega)+1}(\omega) < \varepsilon,$$

according to our definition of R_1. □

Rokhlin towers. When it comes to constructing sets with prescribed properties in an ergodic system, unleashing the Rokhlin Lemma often is a good start. Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system. A finite sequence $(X_l)_{l=0}^L$ of pairwise disjoint sets from $\mathcal{A}$ will be called a *Rokhlin tower (of height L)* if $\mu(X_L) > 0$ and $X_l = T^{-L+l}X_L$ for $0 \leq l \leq L$. In this case, each restriction $T : X_l \to X_{l+1}$, $0 \leq l < L$, is a μ-preserving map (not necessarily invertible). Points of X_l are interpreted to be on *level l*, which is made precise by introducing $\Lambda : X \to \{1, \ldots, L, \infty\}$ with $\Lambda := l$ on each X_l and $\Lambda := \infty$ otherwise. Then, $\Lambda \circ T = \Lambda + 1$ on $\{\Lambda < L\}$. We shall say that *x lies above x'* if $x = T^j x'$ for some $j \geq 0$ and $T^i x \notin X_0$ for $i \in \{1, \ldots, j\}$. If, in that case, $j > 0$, then *x lies strictly above x'*. Note that (a.e. point of) $X \setminus \bigcup_{l=0}^L X_l = \{\Lambda > L\}$ lies strictly above each X_l. The Rokhlin Lemma guarantees that an arbitrarily large proportion of the space can be represented as a Rokhlin tower. We shall use the following version (see e.g. Theorem 1.5.9 of [**1**]).

Lemma 7.4 (**The Rokhlin Lemma**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system on a non-atomic space. Then, for any $L \in \mathbb{N}$ and $\varepsilon > 0$ there exists a Rokhlin tower $(X_l)_{l=0}^L$ of height L with $\mu(X \setminus \bigcup_{l=0}^L X_l) < \varepsilon$.*

The simple transition structure $X_0 \to X_1 \to \ldots \to X_L$ on the tower often allows a good understanding of return times of subsets Y of $\bigcup_{l=0}^L X_l \subseteq X$. We now record some observations in this direction, henceforth focusing on sets $Y \in \mathcal{A}$ which satisfy

$$X_0 \subseteq Y \subseteq \bigcup\nolimits_{l=0}^L X_l. \tag{7.3}$$

Recall that $\varphi_Y : X \to \overline{\mathbb{N}}$ defines a first entrance map $T_Y : X \to Y$ on the whole space X (mod μ). Assuming (7.3) we can identify that part of X which (in the sense defined before) has no points of Y strictly above it as $T_Y^{-1}X_0$. The set $Y \cap T_Y^{-1}X_0$ can be viewed as the *roof* of Y. The following is a variant of Kac' formula. It characterizes the mean return time of Y when restricted to the roof.

Lemma 7.5 (**Kac on the roof**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system, $(X_l)_{l=0}^L$ a Rokhlin tower, and $Y \in \mathcal{A}$ with $X_0 \subseteq Y \subseteq \bigcup_{l=0}^L X_l$. Then*

$$\int_{Y \cap T_Y^{-1}X_0} \varphi_Y \, d\mu = \mu(T_Y^{-1}X_0). \tag{7.4}$$

Moreover, for $Y', Y \in \mathcal{A}$ with $X_0 \subseteq Y' \subseteq Y \subseteq \bigcup_{l=0}^L X_l$, we have

$$T_Y^{-1}X_0 \subseteq T_{Y'}^{-1}X_0. \tag{7.5}$$

Proof. Abbreviate $B := T_Y^{-1}X_0$ and $B' := Y \cap B$, and note that

$$\{\varphi_Y > n\} \cap T^{-n}B = \{\varphi_Y > n\} \cap B \quad \text{for } n \geq 0. \tag{7.6}$$

Recall the canonical representation (from the theory of induced transformations)

$$\mu(A) = \sum_{n\geq 0}\mu(Y \cap \{\varphi_Y > n\} \cap T^{-n}A) \quad \text{for } A \in \mathcal{A}.$$

Combining these we obtain

$$\mu(B) = \sum_{n\geq 0}\mu(\{\varphi_Y > n\} \cap B') = \int_{B'}\varphi_Y \, d\mu,$$

which is (7.4). Assertion (7.5) is straightforward. □

8. Small sets approximating a d-dimensional stationary distribution

The key step in our proof of Theorem 2.1 is contained in the following proposition. Take an arbitrary bounded ergodic $\mathbb{N}$-valued stationary process Ψ and any $d \in \mathbb{N}$. We claim that inside every aperiodic ergodic probability preserving system $(X, \mathcal{A}, \mu, T)$ there is a set Y whose d-dimensional return distribution imitates the d-dimensional marginal of Ψ up to an arbitrarily small error.

Theorem 8.1 (**Approximate embedding of a stationary sequence**). *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic probability preserving system on a nonatomic space, and $\Psi = (\psi_j)_{j\geq 0}$ any bounded ergodic $\mathbb{N}$-valued stationary process. Then, for any $d \in \mathbb{N}$ and $\varepsilon > 0$, there is some $Y \in \mathcal{A}$, $0 < \mu(Y) < 2/\overline{\psi}$, where $\overline{\psi} := \mathbb{E}[\psi_j]$, such that*

$$\text{dist}^{[d]}\left(\mu(Y)\,\widetilde{\Phi}_Y, \overline{\psi}^{\,-1}\,\Psi\right) < \varepsilon. \tag{8.1}$$

Proof. Throughout, we fix $d \geq 1$ and any $\varepsilon \in (0, 1)$. If $A \in \mathcal{A}$, $\mu(A) > 0$, and Ξ is any measurable map defined on X, we will write $\text{law}(\Xi\,|_A) := \mu_A \circ \Xi^{-1}$ for the law of the restricted map $\Xi\,|_A$ w.r.t. the conditional measure μ_A.

(i) We begin with some preparations. First, choose some $\delta > 0$ such that

$$\text{dist}^{[d]}(c\,\Psi, \overline{\psi}^{\,-1}\,\Psi) < \varepsilon/2 \quad \text{for } c \in (\overline{\psi} - \delta, \overline{\psi} + \delta). \tag{8.2}$$

Take any rational $\varepsilon' \in (0, \min(\overline{\psi}, \delta/24))$, and fix some $\varepsilon'' \in (0, \varepsilon/6)$ so small that

$$\varepsilon'/(1 - 2\varepsilon'') < \delta/4 \quad \text{and} \quad 2\varepsilon''\,\overline{\psi}/(1 - 2\varepsilon'') < \delta/4. \tag{8.3}$$

Write $\xi_r := \psi_0 + \ldots + \psi_{r-1}$. By the ergodic theorem, $R^{-1}\xi_R \to \overline{\psi}$ a.e., which shows that there is some $R_0 \geq 1$ such that

$$\Pr[R^{-1}\xi_R > \overline{\psi} + \varepsilon'] < \varepsilon'' \quad \text{for } R \geq R_0. \tag{8.4}$$

Moreover, according to Lemma 7.3, there is some $R_1 \geq R_0$ such that the variable $\xi_R^{\Diamond} := \max\{\xi_r : r \in \{0, \ldots, R\} \text{ and } \xi_r \leq R\,\overline{\psi}\}$ satisfies

$$\overline{\psi} - \varepsilon' < (R + 1)^{-1}\mathbb{E}[\xi_R^{\Diamond}] \quad \text{for } R \geq R_1. \tag{8.5}$$

Choose some integer $M > d/\varepsilon''$. Henceforth we fix an integer $R > \max(R_1, 1/\varepsilon', d(1-\varepsilon'')/\varepsilon'')$ for which $L := R\,(\overline{\psi} + \varepsilon')$ is an integer which satisfies $L > M\,\sup\psi_0$.

(ii) Appeal to the Rokhlin Lemma to obtain some Rokhlin tower $(X_l)_{l=0}^{L}$ with level function Λ, for which $\mu(X \setminus \bigcup_{l=0}^{L} X_l) = \mu(\{\Lambda > L\}) < \min(1/2, \delta/(64\,\overline{\psi}))$. Since $\varepsilon' < \overline{\psi}$, we have

$$(8\overline{\psi}\,R)^{-1} \leq (4L)^{-1} \leq (L + 1)^{-1}\mu(\{\Lambda \leq L\}) = \mu(X_0) \leq (L + 1)^{-1}. \tag{8.6}$$

Since μ is non-atomic, X_L admits a sequence $(\eta_r)_{r\geq 1}$ of measurable partitions $\eta_r = \{X_L(k_0, \ldots, k_{r-1}) : k_0, \ldots, k_{r-1} \in \mathbb{N}\}$ (where we allow some partitioning sets to be empty), with η_{r+1} refining η_r as $X_L(k_0,\ldots,k_{r-1}) = \bigcup_{k\in\mathbb{N}} X_L(k_0,\ldots,k_{r-1},k)$, and such that their distributions model the finite-dimensional marginals of the process Ψ in that $\mu_{X_L}(X_L(k_0,\ldots,k_{r-1})) = \Pr\left[\psi_0 = k_0,\ldots,\psi_{r-1} = k_{r-1}\right]$. As Ψ is bounded, each η_r is essentially finite (it only contains finitely many sets of positive measure). We define corresponding partitions of the other floors X_l, $0 \le l < L$, by letting $X_l(k_0,\ldots,k_{r-1}) := T^{-L+l} X_L(k_0,\ldots,k_{r-1})$ for $r, k_0,\ldots,k_{r-1} \in \mathbb{N}$. Then,

$$X_l(k_0,\ldots,k_{r-1}) = T^{-1} X_{l+1}(k_0,\ldots,k_{r-1}) \quad \text{for } 0 \le l < L, \tag{8.7}$$

and whenever $0 \le l \le L$ and $r, k_i \in \mathbb{N}$, we have

$$\mu_{X_l}(X_l(k_0,\ldots,k_{r-1})) = \Pr\left[\psi_0 = k_0,\ldots,\psi_{r-1} = k_{r-1}\right]. \tag{8.8}$$

Associated with these partitions are the indexing functions $\gamma_i : \{\Lambda \le L\} \to \mathbb{N}$, $i \ge 0$, given by $\gamma_i(x) := k$ if $x \in X_l(k_0,\ldots,k_{i-1},k)$ for suitable $l, k_0,\ldots,k_{i-1}$. Let $\Gamma := (\gamma_i)_{i\ge 0} : \{\Lambda \le L\} \to \mathbb{N}^{\mathbb{N}_0}$ denote the (product-measurable) sequence of the γ_i. Note that $\gamma_i \circ T = \gamma_i$ (and hence $\Gamma \circ T = \Gamma$) on $\{\Lambda < L\}$. Due to (8.8),

$$\mathrm{law}(\Gamma\,|_{X_l}) \text{ coincides with } \mathrm{law}(\Psi), \quad \text{for } 0 \le l \le L. \tag{8.9}$$

Let $\sigma^m\Gamma := (\gamma_{i+m})_{i\ge 0}$, $m \ge 0$, denote the shifted versions of Γ. As Ψ is stationary, so is each $\Gamma\,|_{X_l}$, and the preceding statement immediately generalizes to

$$\mathrm{law}(\sigma^m\Gamma\,|_{X_l}) \text{ coincides with } \mathrm{law}(\Psi), \quad \text{for } 0 \le l \le L \text{ and } m \ge 0. \tag{8.10}$$

(iii) Let $\Sigma_0 := 0$ and $\Sigma_r := \gamma_0 + \ldots + \gamma_{r-1}$ on $\{\Lambda \le L\}$, which satisfy $\Sigma_r \circ T = \Sigma_r$ on $\{\Lambda < L\}$. Next, we define pairwise disjoint sets $Y_0,\ldots,Y_R \subseteq \{\Lambda \le L\}$. Starting from $Y_0 := X_0$, the Y_r, $1 \le r \le R$, can be viewed as (parts of) the graphs of the sum functions Σ_r above Y_0. Formally, we set

$$\mathcal{Y}_r := \left\{X_{k_0+\ldots+k_{r-1}}(k_0,\ldots,k_{r-1}) : k_i \in \mathbb{N} \text{ with } k_0 + \ldots + k_{r-1} \le L\right\}, \tag{8.11}$$

$r \in \{0,\ldots,R\}$, which is a collection of pairwise disjoint sets, and let

$$Y_r := \bigcup \mathcal{Y}_r = \{\Sigma_r = \Lambda \le L\}. \tag{8.12}$$

Finally, we define $Y := \bigcup_{r=0}^{R} Y_r$, so that $x \in Y$ iff $\Sigma_r(x) = \Lambda(x) \le L$ for some $r \in \{0,\ldots,R\}$. For convenience, set $\rho(x) := r$ if $x \in Y_r$, which defines $\rho : Y \to \{0,\ldots,R\}$.

The set Y satisfies (7.3). Note that $Y_R \subseteq Y \cap T_Y^{-1} X_0$. We can identify the roof $Y \cap T_Y^{-1} X_0$ of Y, and the set $T_Y^{-1} X_0^c$ of points below the roof, in terms of $\Sigma^\blacklozenge : \{\Lambda \le L\} \to \{1,\ldots,L\}$ given by

$$\Sigma^\blacklozenge(x) := \max\{\Sigma_r(x) : r \in \{0,\ldots,R\} \text{ and } \Sigma_r(x) \le L\}.$$

By the corresponding property of the Σ_r we also have $\Sigma^\blacklozenge \circ T = \Sigma^\blacklozenge$ on $\{\Lambda < L\}$. We claim that

$$\{\Lambda \le L\} \cap T_Y^{-1} X_0 = \{\Sigma^\blacklozenge \le \Lambda \le L\}, \tag{8.13}$$

and

$$Y \cap T_Y^{-1} X_0 = \{\Sigma^\blacklozenge = \Lambda \le L\} = Y \cap \{\Sigma^\blacklozenge = \Lambda\}. \tag{8.14}$$

These follows from the definition of Y. To validate (8.13) we fix any $x \in \{\Lambda \le L\}$. Assume first that $\Sigma^\blacklozenge(x) > \Lambda(x)$. Then there is some $r \in \{1,\ldots,R\}$ for which $\Lambda(x) < \Sigma_r(x) \le L$. Letting $j := \Sigma_r(x) - \Lambda(x)$ we then have $0 < \Lambda(T^j x) = \Sigma_r(x) = \Sigma_r(T^j x)$, and hence $T^j x \in Y$. But since $Tx,\ldots,T^j x \in \{\Lambda > 0\}$, this implies $x \notin T_Y^{-1} X_0$. Conversely, suppose that $x \notin T_Y^{-1} X_0$. Let $j := \varphi_Y(x)$, then there is some $r \in \{1,\ldots,R\}$ and some $X_{k_0+\ldots+k_{r-1}}(k_0,\ldots,k_{r-1}) \in \mathcal{Y}_r$ which contains $T^j x$.

Also, as $\{\Lambda > L\} \subseteq T_Y^{-1}X_0$, we have $x, Tx, \ldots, T^{j-1}x \in \{\Lambda \leq L\}$. Therefore, by (8.7), $x \in X_{k_0+\ldots+k_{r-1}-j}(k_0, \ldots, k_{r-1})$, and $\Sigma^{\blacklozenge}(x) \geq \Sigma_r(x) = k_0 + \ldots + k_{r-1} = \Lambda(x) + j > \Lambda(x)$, as required. The formal proof of (8.14) is similar.

For points in Y which lie below the roof, φ_Y is easily understood because

$$\varphi_Y = \gamma_r \quad \text{on } Y_r \cap T_Y^{-1}X_0^c. \tag{8.15}$$

Indeed, let $Z = X_l(k_0, \ldots, k_{r-1}) \in \mathcal{Y}_r$, then $Z \cap \{\gamma_r = k\} = X_l(k_0, \ldots, k_{r-1}, k)$. If $x \in Z \cap \{\gamma_r = k\} \cap T_Y^{-1}X_0^c$, then $r < R$ and $\Sigma_{r+1}(x) = l + k \leq L$ since $\Sigma^{\blacklozenge}(x) > \Lambda(x) = l$ by (8.13). But then, $\Lambda(T^j x) = l + j$ for $1 \leq j \leq k$, and $k = \inf\{j \geq 1 : \Lambda(T^j x) = \Sigma_m(x) \text{ for some } m \in \{0, \ldots, R\}\} = \varphi_Y(x)$.

Recalling (8.7), we see that for $X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1}) \in \mathcal{Y}_r$, $1 \leq r \leq R$,

$$X_{k_0+\ldots+k_{r-2}}(k_0, \ldots, k_{r-1}) = T_Y^{-1}X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1}). \tag{8.16}$$

Moreover, it follows that $\rho \circ T_Y = \rho + 1$ on $Y \cap T_Y^{-1}X_0^c$.

(iv) As Y_0 is just the base $X_0 =: X_{0,0}$ of the Rokhlin tower, (8.9) identifies the law of Γ on Y_0 as being that of Ψ. For $r \in \{1, \ldots, R\}$, we can still relate $Y_r = Y \cap \{\rho = r\}$ to X_0 in order to understand the law of Γ on Y_r.

Repeated application of (8.16) shows that for $X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1}) \in \mathcal{Y}_r$,

$$X_0(k_0, \ldots, k_{r-1}) = T_Y^{-r}X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1}). \tag{8.17}$$

Combining these preimages by letting

$$X_{0,r} := \bigcup_{k_i \in \mathbb{N}: k_0+\ldots+k_{r-1} \leq L} X_0(k_0, \ldots, k_{r-1}) = X_0 \cap \{\Sigma_r \leq L\}, \tag{8.18}$$

we thus see that Y_r and $X_{0,r}$ are related via the μ_Y-preserving map T_Y^r,

$$X_{0,r} = T_Y^{-r}Y_r \quad \text{for } 0 \leq r \leq R. \tag{8.19}$$

As an immediate application, we record that $\mu(Y_r) \leq \mu(X_0)$ for all r, hence

$$\mu(Y) \leq (R+1)\mu(X_0) < 2/\overline{\psi} \tag{8.20}$$

(recall (8.6)). Moreover, if $0 \leq r \leq M$, then $\psi_0 + \ldots + \psi_{r-1} \leq M \sup \psi_0 < L$, and hence $X_{0,r} = X_0 \mod \mu$. Therefore, $M \leq \Sigma_M \leq L$, and

$$\mu(Y_r) = \mu(X_0) \quad \text{for } 0 \leq r \leq M. \tag{8.21}$$

Next, if we take any $Z \in \mathcal{A} \cap Y_r$ with $\mu(Z) > 0$, then $\Gamma \mid_Z$ has the same law as $\Gamma \mid_{T_Y^{-r}Z}$ since T_Y^r preserves μ_Y. This statement trivially contains the corresponding assertion for any shifted version of Γ. Hence we conclude that

$$\begin{aligned} &\text{if } 0 \leq r \leq R,\ m \geq 0, \text{ and } Z \in \mathcal{A} \cap Y_r \text{ satisfies } \mu(Z) > 0, \\ &\qquad \text{then } \mathrm{law}(\sigma^m \Gamma \mid_Z) = \mathrm{law}(\sigma^m \Gamma \mid_{T_Y^{-r}Z}). \end{aligned} \tag{8.22}$$

(v) On a large part $Y^{\heartsuit}$ of Y, its d-dimensional return time function $\widetilde{\Phi}_Y^{[d]}$ allows an explicit description in terms of Γ. Define, for $r \in \{0, \ldots, R-d\}$,

$$\mathcal{Y}_r^{\heartsuit} := \left\{X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1+d}) : k_i \in \mathbb{N} \text{ with } k_0 + \ldots + k_{r-1+d} \leq L\right\},$$

$Y_r^{\heartsuit} := \bigcup \mathcal{Y}_r^{\heartsuit} \subseteq Y_r$, and $Y^{\heartsuit} := \bigcup_{r=0}^{R-d} Y_r^{\heartsuit} \subseteq Y$. This is the set of points in Y for which the next d steps under T_Y do not lead out of $\{0 < \Lambda \leq L\}$. We claim that (8.15) and (8.16) imply

$$\widetilde{\Phi}_Y^{[d]} = (k_r, \ldots, k_{r-1+d}) \quad \text{on each } X_{k_0+\ldots+k_{r-1}}(k_0, \ldots, k_{r-1+d}) \in \mathcal{Y}_r^{\heartsuit}.$$

To validate this, note that, for $0 \leq j < d$, $T_Y^{-(d-j)} X_{k_0+\ldots+k_{r-1+d}}(k_0, \ldots, k_{r-1+d}) = X_{k_0+\ldots+k_{r-1+j}}(k_0, \ldots, k_{r-1+d}) \subseteq Y_{r-1+j} \cap T_Y^{-1} X_0^c$, and $\varphi_Y = k_{r-1+j}$ on this set.

We therefore have complete control of $\widetilde{\Phi}_Y^{[d]}$ on $Y^\heartsuit$ in that

$$\widetilde{\Phi}_Y^{[d]} = (\sigma^r \Gamma)^{[d]} \quad \text{on } Y_r^\heartsuit \quad \text{for } 0 \leq r \leq R-d. \tag{8.23}$$

Define $X_{0,r}^\heartsuit \subseteq X_{0,r}$ by $X_{0,r}^\heartsuit := T_Y^{-r} Y_r^\heartsuit = \bigcup_{k_i \in \mathbb{N}: k_0+\ldots+k_{r-1+d} \leq L} X_0(k_0, \ldots, k_{r-1+d})$. Combining (8.23) and (8.22) we see that

$$\text{law}(\widetilde{\Phi}_Y^{[d]} \mid_{Y_r^\heartsuit}) = \text{law}((\sigma^r \Gamma)^{[d]} \mid_{X_{0,r}^\heartsuit}), \quad \text{for } 0 \leq r \leq R-d. \tag{8.24}$$

(vi) We are now ready to show that the distribution of $\widetilde{\Phi}_Y^{[d]}$ on $Y^\heartsuit$ is close to the law of $\Psi^{[d]}$. First, since $X_0 \setminus X_{0,r}^\heartsuit = \bigcup_{k_i \in \mathbb{N}: k_0+\ldots+k_{r-1+d} > L} X_0(k_0, \ldots, k_{r-1+d})$ (disjoint), we need only recall (8.8) and (8.4) to see that

$$\begin{aligned} \mu_{X_0}(X_0 \setminus X_{0,r}^\heartsuit) &= \Pr[\psi_0 + \ldots + \psi_{r-1+d} > L] \\ &\leq \Pr[\xi_R > L] < \varepsilon'' \quad \text{for } 0 \leq r \leq R-d. \end{aligned} \tag{8.25}$$

Note then that, due to (8.24) and (8.10), we have

$$\text{law}(\widetilde{\Phi}_Y^{[d]} \mid_{Y_r^\heartsuit}) - \text{law}(\Psi^{[d]}) = \text{law}((\sigma^r \Gamma)^{[d]} \mid_{X_{0,r}^\heartsuit}) - \text{law}((\sigma^r \Gamma)^{[d]} \mid_{X_0}).$$

Therefore (8.25) allows us to apply Lemma 7.1 a) to get

$$\| \text{law}(\widetilde{\Phi}_Y^{[d]} \mid_{Y_r^\heartsuit}) - \text{law}(\Psi^{[d]}) \| < \varepsilon'' \quad \text{for } 0 \leq r \leq R-d,$$

and part b) of Lemma 7.1 then ensures that

$$\| \text{law}(\widetilde{\Phi}_Y^{[d]} \mid_{Y^\heartsuit}) - \text{law}(\Psi^{[d]}) \| < \varepsilon''. \tag{8.26}$$

For later use we record, as another consequence of (8.25), that

$$2^{-1}(R+1)\mu(X_0) \leq (1-2\varepsilon'')(R+1)\mu(X_0) \leq \mu(Y) \leq (R+1)\mu(X_0). \tag{8.27}$$

The upper bound is trivial since $\mu(Y_r) \leq \mu(X_0)$. For the lower bound, recall that $\mu(Y_r) = \mu(X_{0,r}) \geq \mu(X_{0,r}^\heartsuit)$, so that (8.25) ensures $\mu(Y_r) \geq (1-\varepsilon'')\mu(X_0)$ for $0 \leq r \leq R-d$, which proves that $\mu(Y) \geq (1-\varepsilon'')(R+1-d)\mu(X_0)$, and hence (8.27) since, by our choice of R, $(1-2\varepsilon'')(R+1) < (1-\varepsilon'')(R+1-d)$.

(vii) Next, we compare the law of $\widetilde{\Phi}_Y^{[d]}$ on $Y^\heartsuit$ to its law on Y by estimating the relative measure of $Y \setminus Y^\heartsuit$. Note first that since $\mu(Y_r) \leq \mu(X_0)$ for all r, we have

$$\mu(\textstyle\bigcup_{r=R-d+1}^R Y_r) \leq d\,\mu(X_0) \leq \varepsilon''\,\mu(Y),$$

where the second inequality uses (8.21) and the definition of M. Next, $\mu(Y_r \setminus Y_r^\heartsuit) = \mu(X_{0,r} \setminus X_{0,r}^\heartsuit)$ and (8.25) gives $\mu(X_{0,r}^\heartsuit) > (1-\varepsilon'')\,\mu(X_0) \geq (1-\varepsilon'')\,\mu(X_{0,r})$, that is, $\mu_{X_{0,r}}(X_{0,r} \setminus X_{0,r}^\heartsuit) < \varepsilon''$ for $0 \leq r \leq R-d$. Therefore,

$$\textstyle\sum_{r=0}^{R-d} \mu(Y_r \setminus Y_r^\heartsuit) \leq \varepsilon'' \sum_{r=0}^{R-d} \mu(Y_r) = \varepsilon''\,\mu(\bigcup_{r=0}^{R-d} Y_r) \leq \varepsilon''\,\mu(Y).$$

But $Y \setminus Y^\heartsuit = \bigcup_{r=R-d+1}^R Y_r \cup \bigcup_{r=0}^{R-d}(Y_r \setminus Y_r^\heartsuit)$. Therefore, combining the two estimates above yields

$$\mu_Y(Y \setminus Y^\heartsuit) < 2\varepsilon''. \tag{8.28}$$

Consequently, Lemma 7.1 a) shows that

$$\| \text{law}(\widetilde{\Phi}_Y^{[d]} \mid_{Y^\heartsuit}) - \text{law}(\widetilde{\Phi}_Y^{[d]} \mid_Y) \| < 2\varepsilon''. \tag{8.29}$$

Together with (8.26) the latter proves the encouraging estimate

$$\| \,\mathrm{law}(\widetilde{\Phi}_Y^{[d]} \mid_Y) - \mathrm{law}(\Psi^{[d]}) \,\| < 3\varepsilon'' < \varepsilon/2. \tag{8.30}$$

(viii) To pass to the appropriately scaled variants, note first that

$$\mathrm{dist}^{[d]}(\mu(Y)\,\widetilde{\Phi}_Y, \overline{\psi}^{\,-1}\,\Psi) \le \mathrm{dist}^{[d]}(\mu(Y)\,\widetilde{\Phi}_Y, \mu(Y)\,\Psi) + \mathrm{dist}^{[d]}\left(\mu(Y)\,\Psi, \overline{\psi}^{\,-1}\,\Psi\right),$$

where, due to $\mu(Y) \le 1$ and (8.30),

$$\begin{aligned}\mathrm{dist}^{[d]}(\mu(Y)\,\widetilde{\Phi}_Y, \mu(Y)\,\Psi) &\le \mathrm{dist}^{[d]}(\widetilde{\Phi}_Y, \Psi)\\ &\le \| \,\mathrm{law}(\widetilde{\Phi}_Y^{[d]}) - \mathrm{law}(\Psi^{[d]}) \,\| < \varepsilon/2.\end{aligned}$$

Therefore, once we check that $\mathrm{dist}^{[d]}\left(\mu(Y)\,\Psi, \overline{\psi}^{\,-1}\,\Psi\right) < \varepsilon/2$, the proof of the proposition will be complete. In view of (8.2) it suffices to show that $\mid \mu(Y)^{-1} - \overline{\psi} \mid < \delta$. But according to Kac' formula, $\mu(Y)^{-1} = \int_Y \varphi_Y \, d\mu_Y$. To complete the proof of the proposition, we are going to check that

$$\Big| \int_{Y \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu_Y - \overline{\psi} \Big| < \delta/2, \tag{8.31}$$

while

$$\int_{Y \cap T_Y^{-1} X_0} \varphi_Y \, d\mu_Y < \delta/2. \tag{8.32}$$

(ix) To get a better understanding of $\Sigma^{\blacklozenge}$, we define a related function $\Sigma^{\lozenge} : \{\Lambda \le L\} \to \{1, \ldots, L\}$ by letting $\Sigma^{\lozenge}(x) := \max\{\Sigma_r(x) : r \in \{0, \ldots, R\}$ and $\Sigma_r(x) \le R\overline{\psi}\}$. Note then that due to (8.9) the law of $\Sigma^{\lozenge} \mid_{X_l}$ coincides with that of $\xi_R^{\lozenge}$,

$$\mathrm{law}(\Sigma^{\lozenge} \mid_{X_l}) = \mathrm{law}(\xi_R^{\lozenge}) \quad \text{for } 0 \le l \le L. \tag{8.33}$$

This is clear since we can represent $\Sigma^{\lozenge}$ as a function $\Upsilon(\Gamma)$ of Γ, and $\xi_R^{\lozenge} = \Upsilon(\Psi)$ for the same measurable map $\Upsilon : \mathbb{N}^{\mathbb{N}_0} \to \mathbb{N}_0$. In particular, $\int_{X_0} \Sigma^{\lozenge} \, d\mu_{X_0} = \mathbb{E}[\xi_R^{\lozenge}]$.

Since $\Sigma^{\lozenge} \le \Sigma^{\blacklozenge} \le L$, (8.5) therefore ensures

$$\overline{\psi} - \varepsilon' < (R+1)^{-1} \int_{X_0} \Sigma^{\blacklozenge} \, d\mu_{X_0} < \overline{\psi} + \varepsilon'. \tag{8.34}$$

(x) We now show that

$$\int_{Y \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu = \int_{X_0} \Sigma^{\blacklozenge} \, d\mu. \tag{8.35}$$

To see this, note that due to (8.12) and (8.13), $Y_r \cap T_Y^{-1} X_0^c = \{\Lambda = \Sigma_r < \Sigma^{\blacklozenge}\}$. The T-invariance of the γ_i (and hence of Σ_r and $\Sigma^{\blacklozenge}$) on $\{\Lambda < L\}$ observed before implies that

$$\gamma_i \circ T_Y = \gamma_i, \; \Sigma_r \circ T_Y = \Sigma_r, \text{ and } \Sigma^{\blacklozenge} \circ T_Y = \Sigma^{\blacklozenge} \quad \text{on } Y \cap T_Y^{-1} X_0^c.$$

Therefore, $T_Y^{-r}(Y_r \cap T_Y^{-1} X_0^c) = T_Y^{-r}(Y_r \cap \{\Sigma_r < \Sigma^{\blacklozenge}\}) = X_{0,r} \cap T_Y^{-r}\{\Sigma_r < \Sigma^{\blacklozenge}\} = X_0 \cap \{\Sigma_r < \Sigma^{\blacklozenge}\}$. Since T_Y preserves μ_Y, we thus see that (8.15) entails

$$\int_{Y_r \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu = \int_{Y_r \cap T_Y^{-1} X_0^c} \gamma_r \, d\mu = \int_{X_0 \cap \{\Sigma_r < \Sigma^{\blacklozenge}\}} \gamma_r \, d\mu.$$

Our claim (8.35) follows, since

$$\begin{aligned}\int_{Y \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu &= \sum_{r=0}^R \int_{Y_r \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu = \sum_{r=0}^R \int_{X_0 \cap \{\Sigma_r < \Sigma^{\blacklozenge}\}} \gamma_r \, d\mu\\ &= \int_{X_0} \Big(\sum_{r=0}^R 1_{\{\Sigma_r < \Sigma^{\blacklozenge}\}} \gamma_r\Big) \, d\mu = \int_{X_0} \Sigma^{\blacklozenge} \, d\mu.\end{aligned}$$

When combined with (8.34) and (8.27), the identity (8.35) yields

$$\overline{\psi} - \varepsilon' < \int_{Y \cap T_Y^{-1} X_0^c} \varphi_Y \, d\mu_Y < (1 - 2\varepsilon'')^{-1} \, (\overline{\psi} + \varepsilon'). \tag{8.36}$$

Recalling (8.3), we see that the latter implies (8.31).

(xi) Turning to (8.32), we note that by Lemma 7.5,

$$\int_{Y\cap T_Y^{-1}X_0}\varphi_Y\,d\mu_Y=\mu_Y(\{\Lambda>L\}\cap T_Y^{-1}X_0)+\mu_Y(\{\Lambda\leq L\}\cap T_Y^{-1}X_0).$$

By construction of the Rokhlin tower (in particular (8.6)), and (8.27),

$$\begin{aligned}\mu_Y(\{\Lambda>L\}\cap T_Y^{-1}X_0)&=\mu_Y(\{\Lambda>L\})<\mu(Y)^{-1}\cdot\delta/(64\,\overline{\psi})\\&\leq 2/(R\,\mu(X_0))\cdot\delta/(64\,\overline{\psi})<\delta/4.\end{aligned}\tag{8.37}$$

By (8.13), T-invariance of $\Sigma^{\blacklozenge}$, and the fact that T is measure preserving,

$$\begin{aligned}\mu_Y(\{\Lambda\leq L\}\cap T_Y^{-1}X_0)&=\mu_Y(\{\Sigma^{\blacklozenge}\leq\Lambda\leq L\})=\textstyle\sum_{l=0}^{L}\mu_Y(X_l\cap\{\Sigma^{\blacklozenge}\leq l\})\\&=\textstyle\sum_{l=0}^{L}\mu_Y(X_0\cap\{\Sigma^{\blacklozenge}\leq l\})\leq\int_{X_0}(L+1-\Sigma^{\blacklozenge})\,d\mu_Y.\end{aligned}$$

Moreover, due to (8.34),

$$\begin{aligned}\textstyle\int_{X_0}(L+1-\Sigma^{\blacklozenge})\,d\mu_{X_0}&=R(\overline{\psi}+\varepsilon')-\textstyle\int_{X_0}\Sigma^{\blacklozenge}\,d\mu_{X_0}+1\\&<2R\,\varepsilon'+1<3R\,\varepsilon'.\end{aligned}$$

Combining these, we find via (8.27) that

$$\begin{aligned}\mu_Y(\{\Lambda\leq L\}\cap T_Y^{-1}X_0)&\leq\mu(X_0)/\mu(Y)\cdot\textstyle\int_{X_0}(L+1-\Sigma^{\blacklozenge})\,d\mu_{X_0}\\&<2/(R+1)\cdot 3R\,\varepsilon'<6\varepsilon'<\delta/4.\end{aligned}\tag{8.38}$$

Together with (8.37) this proves (8.32), and hence the proposition. □

9. Proof of Theorem 2.1

All the tools required for the proof of the main result are now ready.

Proof of Theorem 2.1. For every $l\geq 1$ use Lemma 7.2 to obtain a positive ergodic stationary sequence $\Theta_l=(\theta_{l,j})_{j\geq 0}$ with finite range in $(0,\infty)\cap\mathbb{Q}$, and such that $\mathbb{E}[\theta_{l,0}]=1$, while

$$\mathrm{dist}^{[l]}(\widetilde{\Phi},\Theta_l)<1/2l\quad\text{for } l\geq 1.$$

Choose integers $m_l>l$, $l\geq 1$, such that each $m_l\,\theta_{l,0}$ is integer-valued. Then each of the processes $\Psi_l=(\psi_{l,j})_{j\geq 0}:=m_l\Theta_l=(m_l\,\theta_{l,j})_{j\geq 0}$ satisfies the assumptions of Theorem 8.1, and $\overline{\psi}_l:=\mathbb{E}[\psi_{l,j}]=m_l$. Therefore, for every $l\geq 1$, the proposition provides some $A_l\in\mathcal{A}$ such that $0<\mu(A_l)<2/l$ and

$$\mathrm{dist}^{[l]}(\mu(A_l)\,\widetilde{\Phi}_{A_l},\Theta_l)=\mathrm{dist}^{[l]}(\mu(A_l)\,\widetilde{\Phi}_{A_l},\overline{\psi}_l^{\,-1}\,\Psi_l)<1/2l\quad\text{for } l\geq 1.$$

Hence, $(A_l)_{l\geq 1}$ is an asymptotically rare sequence in $\mathcal{A}$ which satisfies

$$\mathrm{dist}^{[l]}(\widetilde{\Phi},\mu(A_l)\,\widetilde{\Phi}_{A_l})<1/l\quad\text{for } l\geq 1.$$

The latter evidently implies $\mu(A_l)\,\widetilde{\Phi}_{A_l}\overset{\mu_{A_l}}{\Longrightarrow}\widetilde{\Phi}$, as required. □

References

[1] J. Aaronson: *An Introduction to Infinite Ergodic Theory.* AMS 1997.
[2] M. Abadi, B. Saussol: *Hitting and returning to rare events for all alpha-mixing processes.* Stochastic Process. Appl. **121** (2011), 314–323.
[3] P. Billingsley: *Convergence of probability measures* 2ed. Wiley 1999.
[4] V. Chaumoître, M. Kupsa: *k-limit laws of return and hitting times.* Discrete and Continuous Dynamical Systems **15** (2006), 73-86.
[5] N. Haydn, Y. Lacroix, S. Vaienti: *Hitting and return times in ergodic dynamical systems.* Ann. Probab. **33** (2005), 2043-2050.
[6] N. Haydn, N. Winterberg, R. Zweimüller: *Return-time statistics, Hitting-time statistics and Inducing.* In *Ergodic Theory, Open Dynamics, and Coherent Structures*, Springer Proceedings in Mathematics & Statistics **70** (2014), 163-171.
[7] M. Hirata, B. Saussol, S. Vaienti: *Statistics of Return Times: A General Framework and New Applications.* Commun. Math. Phys. **206** (1999), 33-55.
[8] A.F. Karr: *Point Processes and their Statistical Inference.* Marcel Dekker 1986.
[9] M. Kupsa, Y. Lacroix: *Asymptotics for hitting times.* Ann. Probab. **33** (2005), 610–619.
[10] Y. Lacroix: *Possible limit laws for entrance times of an ergodic aperiodic dynamical system.* Israel J. Math. **132** (2002), 253–263.
[11] M. Thaler, R. Zweimüller: *Distributional limit theorems in infinite ergodic theory.* Probab. Theory Relat. Fields **135** (2006), 15-52.
[12] R. Zweimüller: *Mixing limit theorems for ergodic transformations.* J. Theoret. Probab. **20** (2007), 1059–1071.

Fakultät für Mathematik, Uni Wien, Oskar-Morgenstern-Platz 1, 1090 Wien, A
E-mail address: rzweimue@member.ams.org
URL: http://www.mat.univie.ac.at/~zweimueller/