Surrey Notes on Infinite Ergodic Theory

Roland Zweimüller
Fakultät für Mathematik, Universität Wien,
Nordbergstrasse 15, 1090 Wien, Austria
roland.zweimueller@univie.ac.at
http://homepage.univie.ac.at/roland.zweimueller

(slightly revised version)

April 8, 2009

The present notes contain material presented during a course on Infinite Ergodic Theory at the *LMS Graduate school on Ergodic Theory* at the University of Surrey, 16th-19th March 2009, plus a few additional bits of information.

A short course like this can hardly offer more than a glimpse through the keyhole. So, I have tried to arrange a little tableau, conveniently positioned behind the door, and to focus the lights on it. Needless to say, the choice of topics reflects my personal preferences and my view of the field. Let me assure you that beyond this little exposition, there is infinite space behind that door, inhabited by interesting (sometimes scary) creatures, and offering many mathematical challenges. Have fun!

1 What's up ? (basic examples and concepts)

Our setup and a very basic question. A *(deterministic, discrete time) dynamical system* is simply given by a map $T : X \to X$ acting on a *phase space* X. The consecutive images $T^n x$ of an *initial state* $x \in X$ represent the states of the system at later times $n \geq 0$, and our basic goal is to predict the long-term behaviour of such *orbits* $(T^n x)_{n \geq 0}$. Saying that a certain *event* occurs at time n means that $T^n x$ belongs to a specific subset $A \subseteq X$, and we will refer to A itself as the event.

It has become common knowledge that even apparently trivial maps T can lead to very complicated (chaotic) dynamics. Ergodic theory[1] can be seen as a *quantitative theory of dynamical systems*, enabling us to rigorously deal with such situations, where it is impossible to predict *when exactly* some relevant event A is going to take place. It still can, for example, tell us quite precisely *how often* A will occur for typical initial states, or *for how big a percentage* of them this event is going to take place at some definite instant n.

The canonical mathematical framework for such a quantitative approach is that of measure theory, and results of this flavour are most naturally interpreted in terms of probability. The state space will therefore come with a σ-algebra $\mathcal{A}$ of measurable subsets, and all sets, functions, and maps to appear below are understood to be measurable.

A rich quantitative theory is available for systems possessing an *invariant measure* $\mu : \mathcal{A} \to [0, \infty]$, meaning that $\mu \circ T^{-1} = \mu$. In fact, relevant systems often live on spaces so rich that they support many invariant measures for T, but these may live on parts of X we are not really interested in (e.g. on some countable subset). So, we usually focus on measures which are meaningful if we regard T as a model of some real-world system. For example, if X is part of a Euclidean space, measures *absolutely continuous* w.r.t. Lebesgue (i.e. possessing a *density* h such that $\mu(A) = \int_A h(x)\, dx$) are a prime choice.

To exclude pathological situations, all measures μ considered here will be *σ-finite*, that is, X can be represented as a countable union $X = \bigcup_{n \geq 1} X_n$ of subsets $X_n \in \mathcal{A}^+ := \{A \in \mathcal{A} : 0 < \mu(A) < \infty\}$ of finite positive measure[2].

On opening a textbook on ergodic theory, one often finds another standing assumption: invariant measures should be finite (and then w.l.o.g. normalized, $\mu(X) = 1$). In this framework a rich theory with many connections to other fields of mathematics has been developed over the years. (And, no doubt, this is where one should start learning ergodic theory.)

However, there do exist systems of interest (not necessarily too exotic), which happen to have an *infinite invariant measure*, $\mu(X) = \infty$. *Infinite Ergodic Theory* focuses on such creatures. As we will see in this course, they often behave in very strange ways, and fail to comply with rules forming the very basis of finite ergodic theory. But I also hope to convince you that despite their weird habits they are worth studying. And, there still are beautiful results waiting to be discovered, and interesting mathematical challenges to be met.

[1] More precisely, the facet we will be looking at.
[2] Most natural measures have this property.

In what follows, I will only focus on **the simplest quantitative question** of understanding the long-term behaviour of *occupation times*

$$\mathbf{S}_n(A) := \sum_{k=0}^{n-1} 1_A \circ T^k, \quad n \geq 1,$$

of sets $A \in \mathcal{A}$, which simply count the number of visits an orbit pays to A before time n. (Slightly more general, we can also look at *ergodic sums* $\mathbf{S}_n(f) := \sum_{k=0}^{n-1} f \circ T^k$ of measurable functions f.) We will soon see that attempts to answer this reveal strange phenomena and lead to unexpected results. Welcome to infinite ergodic theory!

Some nice examples. As the name suggests, the ergodic theory of dynamical systems has two faces. One (the ergodic theory bit) is pretty abstract, but naturally so, as it aims at understanding basic structures which govern the dynamics of many types of systems. The other (the dynamical systems bit) concentrates on specific classes of systems, and scrutinizes them, trying to unveil hidden structures which the abstract theory applies to. Of course, there are no sharp borders between the two, and indeed their interplay sometimes is the most exciting aspect.

This course will be more on the abstract side, but as a theory without convincing examples is (in my opinion) rather pointless, we first have a look at some very simple systems. (We just don't have enough time to discuss more impressive classes of examples. The point really is that even for the simple ones I'll present, matters already are fairly difficult.)

As a warm-up, THE most important finite measure preserving system:

Example 1 (Coin-tossing and Doubling map) *Well, in its favorite suit, it doesn't really look like a dynamical system. The fair coin-tossing process is an independent sequence $(C_n)_{n\geq 0}$ of random variables on some proba space $(\Omega, \mathcal{B}, \Pr)$ with $\Pr[C_n = 1] = \Pr[C_n = -1] = \frac{1}{2}$. If you have had a course on advanced probability, you know that there is a canonical way of constructing such a thing, by letting $\Omega := \{-1, 1\}^{\mathbb{N}} = \{\omega = (\omega_n)_{n\geq 0} : \omega_n = \pm 1\}$ be the space of all possible outcomes of the whole process, equipped with product σ-field $\mathcal{B}$ and product measure $\Pr := \bigotimes_{n\geq 0} \frac{1}{2}(\delta_{-1} + \delta_1)$. Then set $C_n := C \circ S^n$, $n \geq 0$, where $C(\omega) := \omega_0$ and $S : \Omega \to \Omega$ is the* shift, *$(S\omega)_n = \omega_{n+1}$. The fact that the process is* stationary *(i.e. $(C_n)_{n\geq 0}$ and $(C_{n+m})_{n\geq 0}$ have the same distribution for all $m \geq 0$) is equivalent to saying that S preserves the probability measure* Pr*. The formal model can thus be seen as a dynamical system, and the projection $\Pr \circ C$ of the invariant measure to the space $\{-1, 1\}$ of coin-states assigns the same mass $\frac{1}{2}$ to all of them.*

Another well-known representation, with $(\Omega, \mathcal{B}, \Pr) := ([0, 1], \mathcal{B}_{[0,1]}, \lambda)$ (the unit interval with Lebesgue measure), looks even more dynamical: Let $C := 2 \cdot 1_{[1/2,0]} - 1$, and $C_n := C \circ S^n$ where $Sx := 2x \bmod 1$, which preserves λ. This is a uniformly expanding piecewise smooth interval map*, and there is a large* folklore class *of such maps which form a very prominent family of finite-measure preserving systems which exhibit highly chaotic (i.e. as random as the coin-tossing game) behaviour.*

But then almost every mathematician also knows at least one infinite measure preserving system. Often without being aware of this fact. Here it is:

Example 2 (Coin-tossing Random Walk) *This is the random process* (Σ_n) *you obtain from* (C_n) *above as* $\Sigma_n := \sum_{k=1}^{n} C_n$*, i.e. starting at* $\Sigma_0 = 0$*, we toss a coin every second an hop, according to the outcome, one step to the left or to the right on the integer lattice* $\mathbb{Z}$*. Again, a canonical model is given by a shift transformation on the appropriate sequence space. (Do construct it as an exercise if you haven't seen it done.) This time, however, the invariant measure is not the proba measure describing our process, but an infinite measure* μ *which is the sum of all translates (to initial positions different from the origin* $0 \in \mathbb{Z}$*) of the latter! Projected to the lattice* $\mathbb{Z}$ *it simply gives the same mass (one, say) to all lattice points, which is, of course, most natural for a system with a translational symmetry.*

But now for something which really doesn't come from classical probability.

Example 3 (Boole's transformation) *is a simple map* T *on* $X := \mathbb{R}$*,*

$$T : X \to X, \quad Tx := x - \frac{1}{x}.$$

We claim that it preserves (the infinite) Lebesgue measure $\lambda =: \mu$*. To check this, it suffices to consider intervals* $A = [a, b]$*. Note that their preimages consist of two other intervals,* $T^{-1}[a, b] = [a_1, b_1] \cup [a_2, b_2]$*, one on the negative and one on the positive half-line. To compare the measure (i.e. length) of* $T^{-1}[a, b]$ *and* $[a, b]$ *we do not even have to explicitly calculate the preimages in this case! Simply note that* a_1 *and* a_2 *solve* $Tx = a$*, meaning that they are the roots of* $x^2 - ax - 1 = 0$*. Due to Vieta's formula (i.e. equating coefficients in* $x^2 - ax - 1 = (x - a_1)(x - a_2)$*) we have* $a_1 + a_2 = a$*. Analogously,* $b_1 + b_2 = b$*. But then* $\mu(T^{-1}[a, b]) = (b_1 - a_1) + (b_2 - a_2) = b - a = \mu([a, b])$*, as required. Cute, eh?*

Be warned that the reason for the invariant measure to be infinite is not *that* T *is defined on an infinite space in the first place. This is an illusion which depends on your choice of coordinates. Let us perform a change of variables, using the diffeomorphism* $\psi : (0, 1) \to \mathbb{R}$ *given by* $\psi(y) := \frac{1}{1-y} - \frac{1}{y}$*, and consider the representation of* T *in* y*-coordinates,* $\widetilde{T} := \psi^{-1} \circ T \circ \psi : (0, 1) \to (0, 1)$*. Explicit calculation gives*

$$\widetilde{T}y = \begin{cases} \frac{y(1-y)}{1-y-y^2} & \text{for } y \in (0, \frac{1}{2}), \\ 1 - \widetilde{T}(1 - y) & \text{for } y \in (\frac{1}{2}, 1). \end{cases}$$

This is an expanding map with two smooth branches. Very similar to the doubling map $x \mapsto 2x \bmod 1$*. However, one little thing makes all the difference: in contrast to the doubling map,* $\widetilde{T}$ *is* not uniformly expanding*. Instead, it has* indifferent (neutral) fixed points *at* $x = 0$ *and* $x = 1$*, which slow down orbits coming close to these points: the closer they get, the slower they move away again. In the present example this effect is so strong that orbits will spend most of their time in arbitrarily small neighborhoods of the fixed points, and this is what the infinite measure reflects: use* ψ^{-1} *to push the invariant measure* μ *from* $\mathbb{R}$ *to the interval* $(0, 1)$*, i.e. consider* $\widetilde{\mu}(\widetilde{A}) := \mu(\psi A)$ *which is invariant for* $\widetilde{T}$*.*

Again this can be calculated explicitly (exercise), and we find that

$$\widetilde{\mu}([c,d]) = \int_c^d \left(\frac{1}{(1-y)^2} + \frac{1}{y^2} \right) dy \qquad \textit{for } 0 < c < d < 1.$$

This density has non-integrable singularities at the bad points, giving infinite mass to any neighborhood.

Interval maps with indifferent fixed points*, similar to $\widetilde{T}$, form a prominent class of infinite measure preserving systems. While comparatively simple from the usual dynamical systems point of view, we already need to confront very serious issues if we wish to understand their finer ergodic properties. See [T1], [T2], or [Z1] for systematic studies of large classes of such maps.*

Example 4 (Parry-Daniels map) *Here is another nice example with $X = (0,1)$. Consider*

$$Tx := \begin{cases} \frac{x}{1-x} & \textit{for } x \in (0, \frac{1}{2}), \\ \frac{1-x}{x} & \textit{for } x \in (\frac{1}{2}, 1). \end{cases}$$

We leave it as an exercise to check that T preserves the infinite measure given by $\mu([a,b]) = \int_a^b \frac{dx}{x}$. This map has first been studied in [D] and [P].

Two basic properties. When studying some dynamical system, an obvious first step is trying to break it up into smaller bits, each of which can be studied separately. This is the case if we have an *invariant set*, that is, some $A \in \mathcal{A}$ for which $T^{-1}A = A$, because then $T^{-1}A^c = A^c$ as well, so that $TA \subseteq A$ and $TA^c \subseteq A^c$. An m.p.t. T on $(X, \mathcal{A}, \mu)$ is said to be *ergodic* if it does not possess non-trivial invariant sets, i.e. if $T^{-1}A = A$ implies $\mu(A) = 0$ or $\mu(A^c) = 0$. For our discussion, we will focus on such basic building blocks.

Having said at the beginning that we'd like to understand how often in large time intervals a certain event A occurs, i.e. how often the set A will be visited by typical orbits, we want to exclude, from the outset, systems which are trivial in that many points from A do not return at all. A m.p.t. T on $(X, \mathcal{A}, \mu)$ is called *conservative* (or *recurrent*) if, given any measurable set A, almost all points of A will eventually return to this set, that is, if[3]

$$A \subseteq \bigcup_{n \geq 1} T^{-n}A \pmod{\mu} \qquad \text{for all } A \in \mathcal{A} \text{ with } \mu(A) > 0.$$

The famous Poincaré Recurrence Theorem shows that in the case of a finite measure μ, this property is automatically fulfilled. (Don't worry, we'll re-prove it below.) Be warned that this is not the case if μ is infinite: The translation map $T : \mathbb{R} \to \mathbb{R}$, $Tx := x + 1$ obviously preserves the Lebesgue measure λ of sets, but it is also clear that no point of $W := (0,1]$ will ever return to this set. In fact, W simply wanders away under iteration of T.

We therefore need ways of checking conservativity. The following characterization of recurrence is very useful.

Proposition 1 (Characterizing Conservativity) *Let T be an m.p.t. on the σ-finite space $(X, \mathcal{A}, \mu)$, then each of the following conditions is equivalent to T being conservative:*

[3] A relation holds mod μ if it is true outside a set of μ-measure zero.

(i) *if $W \in \mathcal{A}$ is a* wandering set *for T (meaning that $W \cap T^{-n}W = \varnothing$ for $n \geq 1$), then necessarily $\mu(W) = 0$;*

(ii) *for all $A \in \mathcal{A}$, $\sum_{k\geq 1} 1_A \circ T^k \geq 1$ a.e. on A;*

(iii) *for all $A \in \mathcal{A}$, $\sum_{k\geq 1} 1_A \circ T^k = \infty$ a.e. on A;*

(iv) *if $B \in \mathcal{A}$ satisfies $B \supseteq T^{-1}B$, then necessarily $\mu(B \setminus T^{-1}B) = 0$.*

Proof. (iii)⇒(ii): Is obvious.

(ii)⇔conservativity: $\sum_{k\geq 1} 1_A \circ T^k(x)$ counts all visits of x to A.

(ii)⇒(i): If W is wandering, then $W \cap \bigcup_{n\geq 1} T^{-n}W = \varnothing$, whence $\mu(W) = \mu(W \setminus \bigcup_{n\geq 1} T^{-n}W)$, but this (due to (ii)) is zero.

(i)⇒(iv): Repeatedly applying T^{-1} to $B \supseteq T^{-1}B$, we get

$$B \supseteq T^{-1}B \supseteq T^{-2}B \supseteq T^{-3}B \supseteq \dots . \tag{1.1}$$

Therefore $W := B \setminus T^{-1}B$ is a wandering set (being disjoint from $T^{-1}B$, while each $T^{-n}W \subseteq T^{-n}B \subseteq T^{-1}B$). Hence $\mu(B \setminus T^{-1}B) = 0$.

(iv)⇒(iii): This is the tricky bit. Take any $A \in \mathcal{A}$, and observe that

$$A \setminus \left\{ \sum_{k\geq 1} 1_A \circ T^k = \infty \right\} \subseteq \left\{ 1 \leq \sum_{k\geq 0} 1_A \circ T^k < \infty \right\} =: B.$$

We need to show that the left-hand set has measure zero. Of course, this follows once we prove that $\mu(B) = 0$. By a curious coincidence, the letter B has already been used in (iv), and it is easy to see that indeed $B \supseteq T^{-1}B$ (i.e. if $Tx \in B$, then $x \in B$, since adding one step to the orbit cannot spoil the condition defining B). In view of (iv), we therefore see that

$$\mu(T^{-k}B \setminus T^{-(k+1)}B) = \mu(T^{-k}(B \setminus T^{-1}B)) = 0 \quad \text{for } k \geq 0.$$

But again we have a chain of inclusions, as in (1.1), so that, all in all,

$$B = \bigcup_{k\geq 0} (T^{-k}B \setminus T^{-(k+1)}B) \cup \bigcap_{k\geq 0} T^{-k}B \quad \text{(disjoint)}.$$

Now $\mu(\bigcup_{k\geq 0}(T^{-k}B \setminus T^{-(k+1)}B)) = \sum_{k\geq 0} \mu(T^{-k}B \setminus T^{-(k+1)}B) = 0$, and to verify our claim $\mu(B) = 0$, we need only observe that for our particular B,

$$\bigcap_{k\geq 0} T^{-k}B = \varnothing,$$

since each $x \in B$ has a maximal $k \geq 0$ for which $T^k x \in A$, which means that the orbit of $T^{k+1}x$ will never visit A, so that in particular $x \notin T^{-(k+1)}B$. ■

We now see at once that conservativity is automatic if $\mu(X) < \infty$: If, in that case, $B \in \mathcal{A}$ satisfies $B \supseteq T^{-1}B$, then $\mu(B \setminus T^{-1}B) = \mu(B) - \mu(T^{-1}B) = 0$ by invariance. (Note that the difference of measures does not make sense for infinite measure sets B.)

Let us finally point out that our two basic properties are also required to ensure that we can justly call μ THE invariant measure for T, i.e. that it is essentially unique (at least among absolutely continuous measures, i.e. those which do not hide on a set invisible to μ).

Proposition 2 (Uniqueness of μ) *Let T be a conservative ergodic m.p.t. on the σ-finite space $(X, \mathcal{A}, \mu)$. If ν is another T-invariant measure on $\mathcal{A}$, absolutely continuous w.r.t. μ (that is, $\mu(A) = 0$ implies $\nu(A) = 0$), then*

$$\nu = c \cdot \mu \quad \textit{for some } c \in (0, \infty).$$

This can be established using techniques introduced in the next section.

2 Wait and see (the powerful idea of inducing)

A simple idea ... Among the techniques which have proved useful in analyzing recurrent infinite m.p.t.s, one simple classical (cf. [Ka]) construction stands out. It enables us to view the big (infinite measure) system through a smaller (preferably finite measure) window. The basic idea is to fix some reference set $Y \in \mathcal{A}$, take points $x \in Y$, and just wait to see when, and where, they come back to Y. Since we are also going to use this for finding invariant measures for measurable maps on some $(X, \mathcal{A})$ in the first place, we define Y to be a *sweep-out set* if it is measurable and all orbits visit Y, i.e. if

$$\bigcup_{n \geq 1} T^{-n} Y = X. \tag{2.1}$$

We can then define the function

$$\varphi : X \to \mathbb{N} \quad \text{with} \quad \varphi(x) := \min\{n \geq 1 : T^n x \in Y\}, \tag{2.2}$$

called the *hitting time* of Y or, when restricted to this set, the *return time* of Y. That is, $\varphi(x)$ is the number of steps the orbit of x needs to (re-)enter Y. The position at which it enters Y then is $T^{\varphi(x)}x$, which defines the *first-return map* (or *induced map*) of T on Y,

$$T_Y : Y \to Y \quad \text{with} \quad T_Y\, x := T^{\varphi(x)} x. \tag{2.3}$$

T_Y thus is an accelerated version of T. In passing to the induced system we certainly lose the information what, precisely, happens during the successive *excursions* from Y. However, if we keep track of their lengths $\varphi \circ T_Y^j$, $j \geq 0$, we can, for example, reconstruct the occupation times $\mathbf{S}_n(Y)$ of Y (or, for that matter, of any subset $A \in Y \cap \mathcal{A}$).

In situations with a given invariant measure, it suffices to assume that (2.1) holds mod μ, and we will tacitly use this version when applicable. In this case, φ and T_Y are defined a.e. on X (resp Y), and everything works just as well[4].

[4] The proper formal framework suitable for either case is that of nonsingular transformations, which I have to skip here (see e.g. [A0], [T0]).

Here comes an easy but important identity, used in various formal arguments. As $\varphi < \infty$ we have $Y = \bigcup_{n\geq 1} Y \cap \{\varphi = n\}$ (disjoint), and since $T_Y = T^n$ on $Y \cap \{\varphi = n\}$, we find that

$$\begin{aligned} T_Y^{-1}(Y \cap A) &= \bigcup_{k\geq 1} Y \cap \{\varphi = k\} \cap T_Y^{-1}(Y \cap A) \\ &= \bigcup_{k\geq 1} Y \cap \{\varphi = k\} \cap T^{-k} A \qquad \text{for } A \subseteq X. \end{aligned} \tag{2.4}$$

... which is enormously useful. We will see various applications of this concept during this course. One very basic and classical way of using it is as follows: given a map T we'd like to analyze, try to find a good subset Y on which it induces a map T_Y which we can understand more easily (often meaning that it should belong to a class of systems which have already been studied earlier). Then use something like

Proposition 3 (Basic properties of T via T_Y) *Assume that Y is a sweep-out set for $T : X \to X$, for which T_Y is known to preserve some* finite *measure ν. Then,*

(i) *T has an invariant measure μ with $\mu \mid_Y = \nu$, given by*

$$\mu(A) := \sum_{n\geq 0} \nu\left(Y \cap \{\varphi > n\} \cap T^{-n} A\right), \qquad A \in \mathcal{A}; \tag{2.5}$$

(ii) *T is conservative on $(X, \mathcal{A}, \mu)$;*

(iii) *if T_Y is ergodic on $(Y, \mathcal{A} \cap Y, \mu \mid_{\mathcal{A}\cap Y})$, then T is ergodic on $(X, \mathcal{A}, \mu)$.*

Proof. (i) It is not hard to formally verify $\mu(T^{-1}A) = \mu(A)$ for $A \in \mathcal{A}$. We use the definition (2.5), and decompose

$$Y \cap \{\varphi > n\} = (Y \cap \{\varphi = n+1\}) \cup (Y \cap \{\varphi > n+1\}) \qquad \text{(disjoint)}$$

to obtain expressions of the type appearing in (2.5):

$$\begin{aligned} \mu(T^{-1}A) &= \sum_{n\geq 0} \nu\left(Y \cap \{\varphi > n\} \cap T^{-(n+1)} A\right) \\ &= \sum_{n\geq 0} \nu\left(Y \cap \{\varphi = n+1\} \cap T^{-(n+1)} A\right) \\ &\quad + \sum_{n\geq 0} \nu\left(Y \cap \{\varphi > n+1\} \cap T^{-(n+1)} A\right) \\ &= \sum_{n\geq 1} \nu\left(Y \cap \{\varphi = n\} \cap T^{-n} A\right) + \sum_{n\geq 1} \nu\left(Y \cap \{\varphi > n\} \cap T^{-n} A\right). \end{aligned}$$

Since the right-most sum is but the $\sum_{n\geq 1}$-part of (2.5), we only have to check that the other sum equals the missing $n = 0$ bit $\nu\left(Y \cap \{\varphi > 0\} \cap T^{-0} A\right) = \nu(Y \cap A)$. But in view of (2.4), we find

$$\begin{aligned} \sum_{n\geq 1} \nu\left(Y \cap \{\varphi = n\} \cap T^{-n} A\right) &= \nu\left(\bigcup_{n\geq 1} Y \cap \{\varphi = n\} \cap T^{-n} A\right) \\ &= \nu(T_Y^{-1}(Y \cap A)), \end{aligned}$$

and as, by assumption, ν is T_Y-invariant, our claim follows.

Finally we observe that, $\mu(A) = \nu(A)$ for $A \subseteq Y$ since $Y \cap \{\varphi > n\} \cap T^{-n}A = \varnothing$ for $n \geq 1$ in this case.

(ii) Note first that the definition of a sweep-out set implies a certain recurrence property for Y: applying T^{-N} to the identity (2.1), we see that $\bigcup_{n \geq N} T^{-n}Y = X$ for any $N \geq 1$, i.e. every orbit visits Y at arbitrarily late times, and hence infinitely often, that is,

$$\sum_{k \geq 0} 1_Y \circ T^k = \infty \qquad \text{on } X.$$

Now let W be a wandering set for T. Then $T^{-m}W \cap T^{-(m+n)}W = \varnothing$ for all $n \geq 1$ and $m \geq 0$. (Why?) Due to the disjointness of the $T^{-n}W$ we find, using $\mu(Y) = \nu(Y) < \infty$ and T-invariance of μ,

$$\begin{aligned}
\infty > \mu(Y) = \mu(T^{-n}Y) \quad &\geq \quad \mu\left(T^{-n}Y \cap \bigcup_{k=0}^{n} T^{-k}W\right) = \sum_{k=0}^{n} \mu\left(T^{-n}Y \cap T^{-k}W\right) \\
&= \quad \sum_{k=0}^{n} \mu\left(T^{-(n-k)}Y \cap W\right) = \int_W \left(\sum_{k=0}^{n} 1_Y \circ T^{n-k}\right) d\mu.
\end{aligned}$$

According to our introductory remark, $0 \leq g_n := \sum_{k=0}^{n} 1_Y \circ T^{n-k} = \sum_{k=0}^{n} 1_Y \circ T^k \nearrow \infty$ on X as $n \to \infty$. Therefore, the right-hand integrals can only remain bounded if $\mu(W) = 0$ (monotone convergence).

(iii) Turning to ergodicity, we observe that for any T-invariant set $A = T^{-1}A$, the intersection $Y \cap A$ is invariant for T_Y: due to (2.4), we get

$$\begin{aligned}
T_Y^{-1}(Y \cap A) \quad &= \quad \bigcup_{k \geq 1} Y \cap \{\varphi = k\} \cap T^{-k}A \\
&= \quad \bigcup_{k \geq 1} Y \cap \{\varphi = k\} \cap A \qquad = Y \cap A.
\end{aligned}$$

By assumption, T_Y is ergodic, so that $\mu(Y \cap A) = 0$ or $\mu(Y \cap A^c) = 0$. In the first case we can thus conclude (using $A = T^{-1}A$ again) that $\mu(T^{-n}Y \cap A) = \mu(T^{-n}(Y \cap A)) = \mu(Y \cap A) = 0$ for all $n \geq 1$. In view of (2.1) we then get $\mu(A) = \mu(\bigcup_{n \geq 1} T^{-n}Y \cap A) = 0$. Analogously, the 2nd case gives $\mu(A^c) = 0$. ∎

There is an equally useful converse to this proposition. We omit the proof (it is similar to the argument above).

Proposition 4 (Basic properties of T_Y via T) *Let T be a m.p. map on the σ-finite space $(X, \mathcal{A}, \mu)$, and Y a sweep-out set with $\mu(Y) < \infty$. Then*

- ***(i)*** *T_Y is measure-preserving on $(Y, \mathcal{A} \cap Y, \mu\,|_{\mathcal{A} \cap Y})$;*
- ***(ii)*** *T_Y is conservative (a joke, really);*
- ***(iii)*** *if T is ergodic on $(X, \mathcal{A}, \mu)$, then T_Y is ergodic on $(Y, \mathcal{A} \cap Y, \mu\,|_{\mathcal{A} \cap Y})$.*

As an easy exercise, show that conservative ergodic m.p. systems come with loads of sweep-out sets:

Remark 1 *Let T be a conservative ergodic m.p. map on the σ-finite space $(X, \mathcal{A}, \mu)$. Then every $Y \in \mathcal{A}^+$ is a sweep-out set.*

Just how infinite is μ ? (Sounds stupid, but is very important.) Let T be a conservative ergodic m.p. map on the σ-finite space $(X, \mathcal{A}, \mu)$, $\mu(X) = \infty$. The return-time function φ of a sweep-out set $Y \in \mathcal{A}^+$ enables us to make sense of this question. We shall see in the next section that $\mu(X) = \infty$ is equivalent to non-integrability of φ, $\int_Y \varphi \, d\mu = \infty$. The latter means that, starting in Y (with normalized measure $\mu_Y = \frac{\mu|_Y}{\mu(Y)}$), there is a high chance of having to wait a long time for the next visit. Precisely, $\mu(Y \cap \{\varphi > n\})/\mu(Y)$ is the probability of seeing an excursion of length larger than n, and non-integrability means that $\sum_{n \geq 0} \mu(Y \cap \{\varphi > n\}) = \infty$. Information on how fast this series diverges, or (equivalently) on how slowly the $\mu(Y \cap \{\varphi > n\})$ decrease to 0 thus quantifies *how small Y is in X* (under T), or how big X is (relative to Y).

Sometimes the asymptotics of $\mu(Y \cap \{\varphi > n\})$ can be determined using

Lemma 1 *If Y is a sweep-out set for the m.p. map T on the σ-finite space $(X, \mathcal{A}, \mu)$, $\mu(Y) < \infty$, then*

$$\mu(Y \cap \{\varphi > n\}) = \mu(Y^c \cap \{\varphi = n\}) \quad \textit{for } n \geq 1.$$

Proof. Observe first that for $n \geq 0$,

$$T^{-1}(Y^c \cap \{\varphi > n\}) = (Y^c \cap \{\varphi > n+1\}) \cup (Y \cap \{\varphi > n+1\}) \text{ (disjoint).} \tag{2.6}$$

Now take any $E \in \mathcal{A}$. This can be written as

$$E = (Y \cap \{\varphi > 0\} \cap E) \cup (Y^c \cap \{\varphi > 0\} \cap E) \text{ (disjoint),}$$

and repeated application of (2.6) yields the decomposition

$$T^{-n}E = \bigcup_{k=0}^{n} T^{-(n-k)}(Y \cap \{\varphi > k\} \cap T^{-k}E) \cup (Y^c \cap \{\varphi > n\} \cap T^{-n}E) \text{ (disjoint).}$$

If $E := T^{-1}Y$, this means that

$$T^{-(n+1)}Y = \bigcup_{k=0}^{n} T^{-(n-k)}(Y \cap \{\varphi = k+1\}) \cup (Y^c \cap \{\varphi > n+1\}) \text{ (disjoint),}$$

which by T-invariance of μ results in

$$\mu(Y) = \mu(Y \cap \{\varphi \leq n+1\}) + \mu(Y \cap \{\varphi = n+1\}), \quad n \geq 0,$$

as required. ■

Remark 2 *One important caveat: the order of* $(\mu(Y \cap \{\varphi > n\}))$ *really depends on the set Y! We won't discuss this in detail here, but we'll soon see analogous phenomena (meaning problems) on the level of occupation times. Still, understanding it* for certain types of sets Y *will be shown to be of utmost importance in Section 4 below.*

Let's have a look at some basic examples:

Example 5 (Simple Random Walk) *For the coin-tossing random walk on $\mathbb{Z}$, with Y corresponding to the origin, probabilists have known for a long time (see e.g. Chapter III of [F]) that*

$$\mu_Y(Y \cap \{\varphi > n\}) \sim \sqrt{\frac{2}{\pi n}} \qquad \text{as } n \to \infty.$$

Example 6 (Boole's transformation) *We return to the first "serious" infinite measure preserving dynamical system we have been introduced to, Boole's transformation T on the real line. So far, the only thing we know is that it preserves $\mu = \lambda$. What about other basic properties?*

Consider $Y := \left[-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right] =: [-x_0, x_0]$, the special property of x_0 being that it has period 2 for T. Obviously, $\mu(Y) = \sqrt{2} < \infty$, and a look at the graph of T shows that Y is a sweep-out set: Define $x_n > 0$, $n \geq 1$, by requiring that $Tx_n = x_{n-1}$, then $x_n \nearrow \infty$ and $\varphi = n$ on $\pm(x_{n-1}, x_n]$. We can thus appeal to Prop 4 to see that T_Y is m.p. on $(Y, \mathcal{A} \cap Y, \mu \mid_{\mathcal{A} \cap Y})$. By Prop 3 we can then conclude that T is conservative! Further analysis shows that T_Y is a uniformly expanding piecewise smooth "folklore map", and hence known to be ergodic. Using Prop 3 again, we then see that T is ergodic.

Now, how small a set in X is Y? By the preceding observation, we have $Y^c \cap \{\varphi = n\} = \pm(x_{n-1}, x_n]$, hence $\mu(Y \cap \{\varphi > n\}) = \mu(Y^c \cap \{\varphi = n\}) = 2(x_n - x_{n-1})$ by the lemma. According to the definition of the x_n, we have $x_{n-1} = x_n - \frac{1}{x_n}$, and hence

$$x_n^2 - x_{n-1}^2 = 2 - \frac{1}{x_n^2} \longrightarrow 2,$$

so that also

$$\frac{x_n^2}{n} = \frac{x_0^2}{n} + \frac{1}{n}\sum_{k=1}^{n}(x_n^2 - x_{n-1}^2) \longrightarrow 2,$$

which means $x_n \sim \sqrt{2n}$. This asymptotic relation can, in fact, be "differentiated" to give $x_n - x_{n-1} \sim \frac{1}{\sqrt{2n}}$ (exercise). Consequently,

$$\mu(Y \cap \{\varphi > n\}) \sim \sqrt{\frac{2}{n}} \qquad \text{as } n \to \infty.$$

We thus see that (at least for the reference sets Y we have chosen), the two examples have asymptotically proportional *tail probabilities* $\mu_Y(Y \cap \{\varphi > n\})$. But, of course, other asymptotic orders are possible.

Example 7 (Parry-Daniels map) *For this example, $Y := [\frac{1}{2}, 1)$ is easily seen to be a sweep-out set. As we know the invariant measure μ for T, it is clear that $\mu(Y) < \infty$. By Prop 4, it follows that T_Y preserves the finite measure $\mu \mid_Y$, and Prop 3 then proves that T is conservative. Again, closer inspection shows that T_Y is a folklore map and therefore ergodic. Hence so is T (Prop 3).*

To understand the asymptotics of the return time tails, consider the points $x_n := \frac{1}{n+1}$, $n \geq 1$, which satisfy $Tx_n = x_{n-1}$, so that $Y^c \cap \{\varphi = n\} = [x_{n+1}, x_n)$. Consequently, by the lemma,

$$\begin{aligned}
\mu(Y \cap \{\varphi > n\}) &= \mu(Y^c \cap \{\varphi = n\}) = \mu([x_{n+1}, x_n)) \\
&= \int_{\frac{1}{n+2}}^{\frac{1}{n+1}} \frac{dx}{x} = \log\left(1 + \frac{1}{n+1}\right) \sim \frac{1}{n} \qquad \text{as } n \to \infty.
\end{aligned}$$

3 Pointwise matters matter (ergodic theorems for infinite measures)

What does Birkhoff's Ergodic Theorem say about ∞ measures? After these preparatory sections we turn to the basic question we proposed to study, thus focussing on occupation times $\mathbf{S}_n(A)$ of sets $A \in \mathcal{A}$ or, more generally, on ergodic sums $\mathbf{S}_n(f)$ for measurable functions f. *To be definite, we will consider sets of positive finite measure,* $A \in \mathcal{A}^+ := \{B \in \mathcal{A} : 0 < \mu(B) < \infty\}$, *and* (generalizing 1_A) $f \in L_1^+(\mu) := \{g \in L_1(\mu) : g \geq 0 \text{ and } \int g \, d\mu > 0\}$. Throughout this section, T is a conservative ergodic m.p.t. on a σ-finite measure space $(X, \mathcal{A}, \mu)$. Observe that due to these basic assumptions we have, for A and f as above,

$$\mathbf{S}_n(A) \nearrow \infty \quad \text{resp.} \quad \mathbf{S}_n(f) \nearrow \infty \quad \text{a.e. on } X. \tag{3.1}$$

(Why not prove this as an exercise?) Recall that in the case of finite μ the pointwise ergodic theorem identifies the asymptotics of $\mathbf{S}_n(f)(x)$ for a.e. $x \in X$. We record this statement in a version which also explicates what the theorem says about infinite measure preserving situations:

Theorem 1 (Birkhoff's Pointwise Ergodic Theorem) *Let T be c.e.m.p. on the σ-finite measure space $(X, \mathcal{A}, \mu)$. If $\mu(X) < \infty$, then for all $A \in \mathcal{A}^+$ (and $f \in L_1^+(\mu)$),*

$$\frac{1}{n}\mathbf{S}_n(A) \longrightarrow \frac{\mu(A)}{\mu(X)} \quad \left[\textit{resp.} \quad \frac{1}{n}\mathbf{S}_n(f) \longrightarrow \frac{\int f \, d\mu}{\mu(X)}\right] \quad \textit{a.e. on } X. \tag{3.2}$$

If $\mu(X) = \infty$, then for all $A \in \mathcal{A}^+$ (and $f \in L_1^+(\mu)$),

$$\frac{1}{n}\mathbf{S}_n(A) \longrightarrow 0 \qquad \left[\textit{resp.} \quad \frac{1}{n}\mathbf{S}_n(f) \longrightarrow 0\right] \quad \textit{a.e. on } X. \tag{3.3}$$

Some proofs of the ergodic theorem automatically cover the infinite measure case (see e.g. Thm 1.14 of [W]). Since others don't, we will give an easy proof of (3.3) at the end of the present section.

Note that for *finite* μ the theorem tells us three things: It shows that the rate $\mathbf{S}_n(A)(x)$ at which the occupation times diverge is asymptotically the same for a.e. $x \in X$; it proves that this asymptotic rate depends on A only through the measure $\mu(A)$ of that set; and it explicitly identifies this typical pointwise rate as being proportional to n.

For *infinite* μ, however, the result only provides us with an upper bound for $\mathbf{S}_n(A)$, but we do not know how much slower than n the occupation times really increase to infinity at typical points. Neither does it clarify to what extent the asymptotics of $\mathbf{S}_n(A)(x)$ depends on x and A.

Trouble ahead. Can we do better? An optimal result providing the three bits of information we had for finite measures should explicitly identify *the correct rate*, i.e. a sequence $(a_n)_{n \geq 1}$ of positive normalizing constants, such that for all $A \in \mathcal{A}^+$,

$$\frac{1}{a_n}\mathbf{S}_n(A) \longrightarrow \mu(A) \quad \text{a.e. on } X. \tag{3.4}$$

This could then be regarded as an appropriate version of the ergodic theorem for infinite measure spaces. So, let's prove it! Perhaps we should start by checking that the asymptotics of $\mathbf{S}_n(A)(x)$ does not depend on x, and try to identify the proper rate. Hmmmmm...

Don't try too hard, you haven't got a chance! Any attempt to find the correct normalization is doomed to fail. *There simply is no correct rate.* The following slightly distressing result (see e.g. Theorem 2.4.2 of [A0]) provides a precise (but not the strongest possible) version of this statement:

Theorem 2 (Aaronson's Ergodic Theorem) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$, $\mu(X) = \infty$, and $(a_n)_{n\geq 1}$ be any sequence in $(0, \infty)$. Then for all $A \in \mathcal{A}^+$ (and $f \in L_1^+(\mu)$),*

$$\overline{\lim_{n\to\infty}} \frac{1}{a_n}\mathbf{S}_n(A) = \infty \quad \left[\textit{resp.} \quad \overline{\lim_{n\to\infty}} \frac{1}{a_n}\mathbf{S}_n(f) = \infty \right] \quad \textit{a.e. on } X,$$

or

$$\underline{\lim}_{n\to\infty} \frac{1}{a_n}\mathbf{S}_n(A) = 0 \quad \left[\textit{resp.} \quad \underline{\lim}_{n\to\infty} \frac{1}{a_n}\mathbf{S}_n(f) = 0 \right] \quad \textit{a.e. on } X.$$

That is, any potential normalizing sequence $(a_n)_{n\geq 1}$ either over- or underestimates the actual size of ergodic sums of $L_1^+(\mu)$-functions infinitely often.

This shows that the pointwise behaviour of ergodic sums for $L_1^+(\mu)$-functions, even of occupation times $\mathbf{S}_n(A)$ of the nicest possible sets of positive finite measure, is terribly complicated. The rate at which $\mathbf{S}_n(A)(x)$ diverges to ∞ depends in a serious way on the point x, and there is no set $B \in \mathcal{A}$ of positive measure on which the $\mathbf{S}_n(A)$ share a common order of magnitude, not to mention the same asymptotics, as $n \to \infty$.

So what, if anything, can we expect? Doesn't Aaronson's ergodic theorem just kill the field? Luckily, this is not the case. It only shows that the proper version of the ergodic theorem for infinite measure situations can't provide all the things we asked for in (3.4). In fact, the part which works was already discovered in the early years of ergodic theory, cf. [S], [H], shortly after Birkhoff's breakthrough in [B] (see also [N] and [Zu] for the history of this brilliant result). The ratio ergodic theorem states that (while crucially depending on the point) the pointwise asymptotics of the ergodic sums $\mathbf{S}_n(f)$ hardly depends on the function $f \in L_1^+(\mu)$: it only does so through its mean value $\int_X f\, d\mu$. We are going to prove

Theorem 3 (Hopf's Ratio Ergodic Theorem) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$, and $A, B \in \mathcal{A}^+$ (resp $f, g \in L_1^+(\mu)$). Then*

$$\frac{\mathbf{S}_n(A)}{\mathbf{S}_n(B)} \longrightarrow \frac{\mu(A)}{\mu(B)} \quad \left[\textit{resp.} \quad \frac{\mathbf{S}_n(f)}{\mathbf{S}_n(g)} \longrightarrow \frac{\int_X f\, d\mu}{\int_X g\, d\mu} \right] \quad \textit{a.e. on } X.$$

This certainly is an interesting fact, but why call a statement different from what we asked for in the beginning THE proper version? Well, the ratio ergodic theorem turns out to be tremendously useful a.e. in the theory, for example because it often allows us to replace a function $f \in L_1^+(\mu)$ we need to study by some more convenient tailor-made g which simplifies the question. In particular, as we shall see, it is an important tool for proving further results which are closer to our basic question.

Proving the Ratio Ergodic Theorem. The following proof of Theorem 3 is taken from [Z2] (see [KK] for yet another proof). It exploits the idea of inducing in a way which enables us to apply the finite measure ergodic theorem.

To prepare for the argument, we fix some $Y \in \mathcal{A}^+$, and induce on it to obtain a return map $T_Y = T^\varphi : Y \to Y$. According to Prop 4, T_Y is ergodic and m.p. on $(Y, Y \cap \mathcal{A}, \mu\mid_{Y\cap\mathcal{A}})$, and hence defines a dynamical system in its own right. For measurable functions $h : Y \to \mathbb{R}$ we denote *ergodic sums for the induced system* by

$$\mathbf{S}_m^Y(h) := \sum_{j=0}^{m-1} h \circ T_Y^j, \quad m \geq 1, \quad \text{on } Y. \tag{3.5}$$

The most important single example is given by

$$\varphi_m := \mathbf{S}_m^Y(\varphi) := \sum_{j=0}^{m-1} \varphi \circ T_Y^j \ . \tag{3.6}$$

Note that the general term inside, $\varphi \circ T_Y^j(x) = \varphi_j(x) - \varphi_{j-1}(x)$ (also true for $j = 1$ if $\varphi_0 := 0$), is the length of the jth excursion from Y of the orbit $(T^n x)_{n\geq 0}$ in the big system. Hence $\varphi_m(x)$ is just the time at which the *mth return* of x to Y takes place. Quite trivially,

$$\mathbf{S}_m^Y(Y) = m \quad \text{for } m \geq 1. \tag{3.7}$$

The idea of chopping up orbits of T, to obtain pieces corresponding to separate excursions is both simple and very useful, in particular if we collect the values of $f : X \to [0, \infty]$ observed during the first excursion and represent them via a single function, the *induced version of* f given by

$$f^Y : Y \to [0, \infty], \quad f^Y := \mathbf{S}_\varphi(f) = \sum_{j=0}^{\varphi-1} f \circ T^j. \tag{3.8}$$

This new function f^Y *is just* f *seen through the induced system*:

Lemma 2 (Induced functions and ergodic sums) *For measurable* $f \geq 0$,

$$\mathbf{S}_{\varphi_m}(f) = \mathbf{S}_m^Y(f^Y) \quad \textit{for } m \geq 1 \textit{ on } Y, \tag{3.9}$$

and

$$\int_X f \, d\mu = \int_Y f^Y \, d\mu. \tag{3.10}$$

Proof. (i) The orbit section $(T^k x)_{k=0,\ldots,\varphi_m(x)-1}$ which determines $\mathbf{S}_{\varphi_m}(f)$ consists of m complete excursions from Y. Simply chop it up into the corresponding subsections, to obtain

$$\begin{aligned}
\mathbf{S}_{\varphi_m}(f) &= \mathbf{S}_{\varphi_1}(f) + \mathbf{S}_{\varphi_2-\varphi_1}(f \circ T_Y) + \ldots + \mathbf{S}_{\varphi_m-\varphi_{m-1}}(f \circ T_Y^{m-1}) \\
&= \mathbf{S}_\varphi(f) + \mathbf{S}_{\varphi\circ T_Y}(f \circ T_Y) + \ldots + \mathbf{S}_{\varphi\circ T_Y^{m-1}}(f \circ T_Y^{m-1}) \\
&= \mathbf{S}_\varphi(f) + (\mathbf{S}_\varphi(f)) \circ T_Y + \ldots + (\mathbf{S}_\varphi(f)) \circ T_Y^{m-1} = \mathbf{S}_m^Y(f^Y).
\end{aligned}$$

(ii) We note that (3.10) is true for indicator functions 1_A, where it is equivalent to the magic formula (2.5) we saw earlier,

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

Simply write the bits involved as integrals, to get

$$\begin{aligned} \int_X 1_A \, d\mu &= \int_Y \left(\sum_{n\geq 0} 1_{Y\cap\{\varphi>n\}} \cdot 1_A \circ T^n \right) d\mu \\ &= \int_Y \left(\sum_{n=0}^{\varphi-1} 1_A \circ T^n \right) d\mu = \int_Y 1_A^Y \, d\mu. \end{aligned}$$

A routine argument from measure theory then shows that (3.10) holds for all measurable $f : X \to [0,\infty]$. (Easy exercise.) ■

As an immediate consequence, we see (taking $f := 1_X$ so that $f^Y = \varphi$) that

$$\mu(X) = \int_Y \varphi \, d\mu, \tag{3.11}$$

which is usually referred to as *Kac' formula*. In particular, we see that the T-invariant measure μ is infinite iff the return time function φ of any $Y \in \mathcal{A}^+$ is non-integrable. Inducing thus leads to a *duality between infinite measure preserving transformations and the study of non-integrable functions over finite measure preserving systems.* An explicit illustration of this scheme can be found at the end of the next section. We now proceed to the

Proof of Theorem 3. **(i)** Observe that it suffices to prove that for $f \in L_1^+(\mu)$,

$$\frac{\mathbf{S}_n(f)}{\mathbf{S}_n(1_Y)} \longrightarrow \frac{\int_X f\,d\mu}{\mu(Y)} \qquad \text{a.e. on } Y. \tag{3.12}$$

Indeed, as the set where $\mathbf{S}_n(f)/\mathbf{S}_n(1_Y) \longrightarrow \int_X f\,d\mu/\mu(Y)$ is T-invariant and contains Y, we then see that by ergodicity this convergence in fact holds a.e. on X. Applying the same to g, the assertion of the theorem follows at once.

(ii) To verify (3.12), we consider the return map T_Y which is an ergodic m.p.t. on the finite measure space $(Y, \mathcal{A}\cap Y, \mu \mid_{\mathcal{A}\cap Y})$. We can therefore apply Birkhoff's ergodic theorem to T_Y and f^Y (which is integrable by (3.10)), thus considering the ergodic sums $\mathbf{S}_m^Y(f^Y)$ of the induced system, to see (recalling (3.7) and (3.9)) that

$$\frac{\mathbf{S}_{\varphi_m}(f)}{\mathbf{S}_{\varphi_m}(1_Y)} = \frac{\mathbf{S}_m^Y(f^Y)}{m} \longrightarrow \frac{\int_Y f^Y\,d\mu}{\mu(Y)} = \frac{\int_X f\,d\mu}{\mu(Y)} \qquad \text{a.e. on } Y. \tag{3.13}$$

This proves (3.12) for a.e. $x \in Y$ along the subsequence of indices $n = \varphi_m(x)$.

(iii) To prove convergence of the full sequence, we need only observe that $\mathbf{S}_n(f)$ is non-decreasing in n since $f \geq 0$. Hence, if for any n we choose $m = m(n,x)$ such that $n \in \{\varphi_{m-1}+1, \ldots, \varphi_m\}$, we find (again using (3.7))

$$\frac{m-1}{m}\frac{\mathbf{S}_{m-1}^Y(f^Y)}{m-1} \leq \frac{\mathbf{S}_n(f)}{\mathbf{S}_n(1_Y)} \leq \frac{\mathbf{S}_m^Y(f^Y)}{m}\ ,$$

and (3.12) follows from (3.13) since $m(n,x) \to \infty$ as $n \to \infty$. ■

Having established the ratio ergodic theorem, we conclude this section with a quick proof of (3.3): by σ-finiteness of the space we have, for any $m \geq 1$, some $B_m \in \mathcal{A}^+$ with $\mu(B_m) \geq m$. Applying the ratio ergodic theorem to the pair A, B_m yields (since $\mathbf{S}_n(B_m) \leq n$)

$$0 \leq \overline{\lim_{n\to\infty}} \frac{\mathbf{S}_n(A)}{n} \leq \lim_{n\to\infty} \frac{\mathbf{S}_n(A)}{\mathbf{S}_n(B_m)} = \frac{\mu(A)}{\mu(B_m)} \quad \text{a.e. on } X.$$

Since m was arbitrary, and $\mu(B_m) \to \infty$, our claim (3.3) follows.

4 Distributions, too, do (capturing the order of $\mathbf{S}_n(A)$ in a weaker sense)

Another attempt: averaging over sets. Let T be a conservative ergodic m.p.t. on $(X, \mathcal{A}, \mu)$, and $A \in \mathcal{A}^+$. We still haven't achieved our goal of capturing the asymptotic size of occupation times $\mathbf{S}_n(A)$. Recall that by the ratio ergodic theorem the dependence of $\mathbf{S}_n(A)$ on the choice of A is very simple. However, the discussion above also showed that for each A the pointwise behaviour of the functions $\mathbf{S}_n(A)$ is awfully complicated, as no realization $(\mathbf{S}_n(A)(x))_{n\geq 1}$ captures the order (let alone the exact asymptotics) of $(\mathbf{S}_n(A))_{n\geq 1}$ on any set of positive measure.

So, is there anything we can do? One natural approach which might help us is to perhaps smoothen out the weird habits of individual points by *averaging* $\mathbf{S}_n(A)$ *over some set of positive measure*. (Note that, after all, Theorem 2 only tells us that a.e. $\mathbf{S}_n(A)(x)$ goes crazy infinitely often, but doesn't rule out the possibility of this only happening on very rare occasions, which could still be consistent with regular behaviour of an average.) We are thus lead to the idea of considering the following quantities: Given any $A \in \mathcal{A}^+$ we define (using the normalized restriction $\mu_A = \frac{\mu|_A}{\mu(A)}$)

$$a_n(A) := \int_A \mathbf{S}_n(A)\, d\mu_A, \quad \text{for } n \geq 1, \tag{4.1}$$

which, in probabilistic terms, is just the expectation of $\mathbf{S}_n(A)$ if we think of starting our system on A (and with initial distribution μ_A).

Having removed the dependence on individual points in this way, we've now got something which only depends on A. But then we already know by Theorem 3 that the asymptotics of $\mathbf{S}_n(A)$ depends on A only through $\mu(A)$. So we have found something encoding a "characteristic rate" of T, haven't we?

Weeeeell ... again it isn't true. We've hit another wall! In general we don't have good control of integrals of function sequences which we can control a.e., unless we have some extra information (e.g. monotone or dominated convergence). And, in fact, this inevitably causes problems in the present situation (cf. Thm 6.2 of [A2]):

Proposition 5 (Non-universality of $a_n(A)$) *Let T be a conservative ergodic m.p.t. on $(X, \mathcal{A}, \mu)$, $\mu(X) = \infty$. Then for every $A \in \mathcal{A}^+$ there is some $B \in \mathcal{A}^+$ for which*

$$a_n(A) = o(a_n(B)) \quad \text{as } n \to \infty.$$

This is getting pretty annoying! Why did I present something which still doesn't work? Of course, one aim was to emphasize once again that apparently simple things can become very tricky in the presence of an infinite invariant measure. (Note that in case $\mu(X) < \infty$ we have $a_n(A)/\mu(A) \sim a_n(B)/\mu(B) \sim n$ for all $A, B \in \mathcal{A}^+$ by the L_1-version of the ergodic thm.) However, there is another, even better reason: We'll see that *for special sets Y, and special systems T, the $a_n(Y)$ can really do what we want them to, but only in a still weaker sense.*

Transfer operator and very good sets. One extra assumption we need in order to really make sense of the $a_n(A)$ is formulated in terms of the *transfer operator* $\mathbf{P} : L_1(\mu) \to L_1(\mu)$ of the m.p. system $(X, \mathcal{A}, \mu, T)$. Recall that $\mathbf{P}$ describes the evolution of (proba-) densities under the action of T, that is, if u is the density of some proba measure ν w.r.t. μ, then $\mathbf{P}u$ is the density of the image measure $\nu \circ T^{-1}$. Formally, this is reflected in the duality[5] relation

$$\int_X f \cdot \mathbf{P}u \, d\mu = \int_X (f \circ T) \cdot u \, d\mu \quad \text{for } f \in L_\infty(\mu),\ u \in L_1(\mu), \tag{4.2}$$

which immediately explains the relevance of $\mathbf{P}$ for our present situation, as

$$a_n(A) = \frac{1}{\mu(A)} \sum_{k=0}^{n-1} \int (1_A \circ T) \cdot 1_A \, d\mu = \int_A \left(\sum_{k=0}^{n-1} \mathbf{P}^k 1_A \right) d\mu_A. \tag{4.3}$$

For transformations which locally expand on the state space X, $\mathbf{P}$ tends to locally smear out densities u by stretching their support. In such situations, $\mathbf{P}$ (in marked contrast to $\mathbf{U}f := f \circ T$) has a regularizing effect on u. For some classes of finite measure preserving dynamical systems, this makes $\mathbf{P}$ behave much better than $\mathbf{U}$ and enables an efficient analysis.

The last sentence remains true with "finite" replaced by "infinite". Here is a condition which turns out to be satisfied by all our examples. It asks for the existence of a particularly nice set: $Y \in \mathcal{A}^+$ is a *Darling-Kac (DK) set* if[6]

$$\frac{1}{a_n(Y)} \sum_{k=0}^{n-1} \mathbf{P}^k 1_Y \longrightarrow 1 \quad \text{uniformly (mod } \mu) \text{ on } Y. \tag{4.4}$$

Note that the existence of such sets defines a particular class of systems. In general DK-sets need not exist, but for some T they do. In this case the system automatically satisfies a pointwise ergodic theorem for the "dual" operator $\mathbf{P}$, of the form (3.4) we could not achieve for $\mathbf{U}$: the c.e.m.p.t. T then is *pointwise dual ergodic*, meaning that there are $a_n(T) > 0$ for which

$$\frac{1}{a_n(T)} \sum_{k=0}^{n-1} \mathbf{P}^k u \longrightarrow \int_X u \, d\mu \cdot 1_X \quad \text{a.e. on } X \text{ for } u \in L_1(\mu). \tag{4.5}$$

[5]Traditionally, $\mathbf{U}f := f \circ T$ is referred to as the *Koopman operator*, and $\mathbf{P}$ is often just referred to as its dual. Note, however, that $L_\infty(\mu)$ is the dual of $L_1(\mu)$, but not vice versa.

[6]Uniform convergence (mod μ) on Y means that there is some $Y' \subseteq Y$ with $\mu(Y \setminus Y') = 0$ on which the convergence is uniform.

This follows, with

$$a_n(T) := \frac{a_n(Y)}{\mu(Y)}, \tag{4.6}$$

since $\mathbf{P}$, too, satisfies a ratio ergodic theorem parallel to Thm 3:

Theorem 4 (Hurewicz' Ratio Ergodic Theorem) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$, and $A, B \in \mathcal{A}^+$ (resp $f, g \in L_1^+(\mu)$). Then*

$$\frac{\sum_{k=0}^{n-1} \mathbf{P}^k 1_A}{\sum_{k=0}^{n-1} \mathbf{P}^k 1_B} \longrightarrow \frac{\mu(A)}{\mu(B)} \quad \left[\textit{resp.} \quad \frac{\sum_{k=0}^{n-1} \mathbf{P}^k u}{\sum_{k=0}^{n-1} \mathbf{P}^k v} \longrightarrow \frac{\int_X u \, d\mu}{\int_X v \, d\mu} \right] \quad \textit{a.e. on } X.$$

We don't include a proof here, see e.g. §2.2 of [A0]. (Theorems 1, 3, and 4 are but special cases of a marvelously general ratio ergodic theorem for operators, known as the Chacon-Ornstein Theorem, cf. [Kr].) Finally, we can offer some good news about averaged occupation times. The following is an easy consequence of the definition of DK-sets and of Theorem 4:

Proposition 6 (Universality of $a_n(Y)$ for DK-sets Y) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$, and assume that Y, Y' are DK-sets. Then*

$$a_n(T) = \frac{a_n(Y)}{\mu(Y)} \sim \frac{a_n(Y')}{\mu(Y')} \qquad \textit{as } n \to \infty. \tag{4.7}$$

That is, at least the growth rare of averaged occupation times *of DK-sets*, encoded in the asymptotics of $a_n(T)$ is a meaningful concept. While, as pointed out earlier, (4.7) doesn't generalize to all $A \in \mathcal{A}^+$, we will see that it does capture the size of *all* $\mathbf{S}_n(A)$ in a weaker sense.

Calculating the asymptotics of $a_n(T)$. Regular variation. Before doing so, we first indicate how the asymptotics of $a_n(T)$ can be analyzed in our examples. The key to this is a non-trivial relation between the $a_n(T)$ and the tails $\mu(Y \cap \{\varphi > n\})$ of the return distribution of Y, which will enter the discussion via their partial sums $w_n(Y) := \sum_{k=0}^{n-1} \mu(Y \cap \{\varphi > n\})$, $n \geq 1$. (The sequence $(w_n(Y))$ is called the *wandering rate* of Y.) Observing that $w_n(Y) = \int_Y \min(n, \varphi) \, d\mu$ is the expectation of the excursion length φ truncated at n, suggests that its product with $a_n(Y)$, the expected number of visits to, and hence of excursions from Y, should be of order n.

This is made precise in the next proposition which we prove in order to illustrate the flavour of some arguments which are frequently used in the field. Determining the exact asymptotics of $a_n(Y)$ actually requires the $w_n(Y)$ to be *regularly varying*, meaning that there is some $\rho \in \mathbb{R}$ (the *index*) such that

$$w_n(Y) = n^\rho \cdot \ell(n),$$

with ℓ *slowly varying*, i.e. satisfying $\ell(cn)/\ell(n) \to 1$ as $n \to \infty$ for all $c > 0$. (For example, ℓ could be convergent in $(0, \infty)$, but also the logarithm is a suitable function.) The importance of this concept is due to the fact that it not only enables precise asymptotic analysis, but also turn out to be necessary for various desirable conclusions to hold. However, we don't have enough time to discuss this in detail here. (See Chapter 1 of the 490 pages treatise [BGT] for what, according to the authors, the "mathematician in the street" ought to know about regular variation.)

Proposition 7 (Asymptotics of $a_n(Y)$ via $w_n(Y)$) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$, and assume that Y is a DK-set. Then*[7]

$$\frac{n}{w_n(Y)} \lesssim \frac{a_n(Y)}{\mu(Y)} \lesssim \frac{2n}{w_n(Y)} \qquad \text{as } n \to \infty. \tag{4.8}$$

Moreover, if $(w_n(Y))$ is regularly varying of index $1-\alpha$ for some α (necessarily in $[0,1]$), then

$$\frac{a_n(Y)}{\mu(Y)} \sim \frac{1}{\Gamma(2-\alpha)\Gamma(1+\alpha)} \cdot \frac{n}{w_n(Y)} \qquad \text{as } n \to \infty. \tag{4.9}$$

Proof. (i) We are going to relate the two quantities to each other by taking orbit sections consisting of a certain number (the occupation time) of consecutive excursions and splitting off the last of these excursions (the distribution of which is encoded in the $\mu(Y \cap \{\varphi > n\})$). Formally, we decompose, for any $n \geq 0$,

$$Y^n := \bigcup_{k=0}^{n} T^{-k} Y = \bigcup_{k=0}^{n} T^{-k}(Y \cap \{\varphi > n-k\}) \qquad \text{(disjoint)}$$

(a point from Y^n belongs to the kth set on the right-hand side if k is the last instant $\leq n$ at which its orbit visits Y). Passing to indicator functions and integrating this identity over Y then gives (since $Y \cap Y^n = Y$)

$$\begin{aligned} \mu(Y) &= \sum_{k=0}^{n} \int_X 1_Y \cdot 1_{Y \cap \{\varphi > n-k\}} \circ T^k \, d\mu \qquad (4.10) \\ &= \sum_{k=0}^{n} \int_X \mathbf{P}^k 1_Y \cdot 1_{Y \cap \{\varphi > n-k\}} \, d\mu \\ &= \int_Y \left(\sum_{k=0}^{n} \mathbf{P}^k 1_Y \cdot 1_{Y \cap \{\varphi > n-k\}} \right) d\mu. \end{aligned}$$

The expression in brackets is a somewhat unhandy convolution. Still, we can use an elementary argument to validate the estimates (4.8). If we sum the identities (4.10) over $n \in \{0, \ldots, N\}$, we obtain the first and the last statement of

$$\begin{aligned} (N+1)\,\mu(Y) &= \int_Y \left(\sum_{n=0}^{N} \sum_{k=0}^{n} \mathbf{P}^k 1_Y \cdot 1_{Y \cap \{\varphi > n-k\}} \right) d\mu \\ &\leq \int_Y \left(\sum_{k=0}^{N} \mathbf{P}^k 1_Y \right) \left(\sum_{j=0}^{N} 1_{Y \cap \{\varphi > j\}} \right) d\mu \\ &\leq \int_Y \left(\sum_{n=0}^{2N} \sum_{k=0}^{n} \mathbf{P}^k 1_Y \cdot 1_{Y \cap \{\varphi > n-k\}} \right) d\mu = (2N+1)\,\mu(Y), \end{aligned}$$

while the two estimates in the middle are obtained by simply comparing for which pairs (k,j) the expression $\mathbf{P}^k 1_Y \cdot 1_{Y \cap \{\varphi > j\}}$ shows up in the respective expressions. As, by assumption, Y is a DK-set, we can understand the mysterious

[7] Here, $c_n \lesssim d_n$ means that $\overline{\lim}\, c_n/d_n \leq 1$.

bit in the middle, observing that

$$\int_Y \left(\sum_{k=0}^{N} \mathbf{P}^k 1_Y\right) \left(\sum_{j=0}^{N} 1_{Y\cap\{\varphi>j\}}\right) d\mu \;\sim\; a_{N+1}(Y) \cdot \int_Y \left(\sum_{j=0}^{n} 1_{Y\cap\{\varphi>j\}}\right) d\mu$$
$$= a_{N+1}(Y) \cdot w_{N+1}(Y),$$

and (4.8) follows.

(ii) The sharp asymptotic relation (4.9) requires some advanced analytic tools. To efficiently deal with the convolution in (4.10), we pass to the discrete Laplace transforms[8] (generating functions) of the sequences (in n) in that equation. This gives, for $s > 0$,

$$\mu(Y) \sum_{n\geq 0} e^{-ns} = \int_Y \sum_{n\geq 0} \left(\sum_{k=0}^{n} \mathbf{P}^k 1_Y \cdot 1_{Y\cap\{\varphi>n-k\}}\right) e^{-ns}\, d\mu \tag{4.11}$$
$$= \int_Y \left(\sum_{n\geq 0} \mathbf{P}^n 1_Y\, e^{-ns}\right) \left(\sum_{n\geq 0} 1_{Y\cap\{\varphi>n\}}\, e^{-ns}\right)$$

(all terms being non-negative, we are free to interchange summation and integration). Our assumption that Y should be DK provides good control of sums $\sum_{k=0}^{n-1} \mathbf{P}^k 1_Y$, but here we've got the transform $\sum_{n\geq 0} \mathbf{P}^n 1_Y\, e^{-ns}$ of the individual $\mathbf{P}^n 1_Y$ which are (really !!!) hard to understand. However, there is a neat little trick: note that

$$\sum_{n\geq 0} \mathbf{P}^n 1_Y\, e^{-ns} = (1-e^{-s}) \cdot \sum_{n\geq 0} \left(\sum_{k=0}^{n} \mathbf{P}^k 1_Y\right) e^{-ns}$$
$$\sim s \cdot \sum_{n\geq 0} a_n(Y)\, e^{-ns} \quad \text{uniformly on } Y \text{ as } s \searrow 0.$$

Substituting this, (4.11) becomes

$$\frac{\mu(Y)}{s} \sim s \cdot \int_Y \left(\sum_{n\geq 0} a_n(Y)\, e^{-ns}\right) \left(\sum_{n\geq 0} 1_{Y\cap\{\varphi>n\}}\, e^{-ns}\right)$$
$$= s \left(\sum_{n\geq 0} a_n(Y)\, e^{-ns}\right) \left(\sum_{n\geq 0} \mu(Y\cap\{\varphi>n\})\, e^{-ns}\right).$$

We have thus obtained an explicit asymptotic relation between the Laplace transforms of $(a_n(Y))$ and $(\mu(Y\cap\{\varphi>n\}))$. Now the condition of regular variation is exactly what we need in order to (twice) apply the following deep analytic result, which completes the proof. ■

Detailed studies of infinite measure preserving systems often require plenty of serious asymptotic (real) analysis. A cornerstone of this theory is the following result (cf. §1.7 of [BGT]).

[8] That is, $(b_n)_{n\geq 0}$ is encoded in $B(s) := \sum_{n\geq 0} b_n\, e^{-ns}$, $s > 0$. The rate at which $\sum_{k=0}^{n} b_k \to \infty$ is reflected in the behaviour of $B(s)$ as $s \searrow 0$.

Proposition 8 (Karamata's Tauberian Theorem for power series) *Let (b_n) be a sequence in $[0,\infty)$ such that for all $s>0$, $B(s) := \sum_{n\geq 0} b_n\, e^{-ns} < \infty$. Let ℓ be slowly varying and $\rho, \vartheta \in [0,\infty)$. Then*

$$B(s) \sim \vartheta \left(\frac{1}{s}\right)^{\rho} \ell\left(\frac{1}{s}\right) \qquad \text{as } s \searrow 0, \tag{4.12}$$

iff

$$\sum_{k=0}^{n-1} b_k \sim \frac{\vartheta}{\Gamma(\rho+1)}\, n^{\rho}\ell(n) \qquad \text{as } n \to \infty. \tag{4.13}$$

If (b_n) is eventually monotone and $\rho > 0$, then both are equivalent to

$$b_n \sim \frac{\vartheta\rho}{\Gamma(\rho+1)}\, n^{\rho-1}\ell(n) \qquad \text{as } n \to \infty. \tag{4.14}$$

The Darling-Kac Theorem for infinite m.p. T. We are finally in a position to state the main result of this section (generalizing [DK]). It states that in the presence of regular variation, $a_n(Y)$ (for Y a DK-set) exactly captures the asymptotics of *all* $\mathbf{S}_n(A)$ if we consider their distributions.

Theorem 5 (Aaronson's Darling-Kac Theorem) *Let T be a c.e.m.p.t. on the σ-finite measure space $(X, \mathcal{A}, \mu)$. Assume there is some DK-set $Y \in \mathcal{A}^+$. If*

$$(w_n(Y)) \text{ is regularly varying of index } 1-\alpha \tag{4.15}$$

(for some $\alpha \in [0,1]$), then for all $f \in L_1^+(\mu)$ and all $t > 0$,

$$\mu_Y\left[\frac{1}{a_n(T)}\mathbf{S}_n(f) \leq t\right] \longrightarrow \Pr\left[\mu(f)\cdot \mathcal{M}_\alpha \leq t\right] \qquad \text{as } n\to\infty. \tag{4.16}$$

(In fact, μ_Y may be replaced by any proba measure Q with $Q \ll \mu$.)

In here, $\mathcal{M}_\alpha$, $\alpha \in [0,1]$, denotes a non-negative real random variable distributed according to the *(normalized) Mittag-Leffler distribution of order* α, which can be characterized by its moments

$$\mathbb{E}\left[\mathcal{M}_\alpha^r\right] = r!\, \frac{(\Gamma(1+\alpha))^r}{\Gamma(1+r\alpha)}\ , \qquad r \geq 0.$$

For specific parameter values, there is a more explicit description: $\mathcal{M}_1 = 1$ (a constant random variable), $\mathcal{M}_{1/2} = |\mathcal{N}|$ (the absolute value of a standard Gaussian variable), and $\mathcal{M}_0 = \mathcal{E}$ (an exponentially distributed variable).

Example 8 *For both the coin-tossing random walk and Boole's transformation considered above, it can be checked that the reference sets Y given there are are DK, and we have observed that $\mu(Y \cap \{\varphi > n\}) \sim const/\sqrt{n}$. Hence $w_n \sim const \cdot \sqrt{n}$, and we are in the $\alpha = 1/2$ situation of the theorem. We thus see that, say for Boole's transformation,*

$$Q\left[\frac{\pi}{\sqrt{2n}}\mathbf{S}_n(A) \leq \lambda(A)\, t\right] \longrightarrow \frac{2}{\pi}\int_0^t e^{-\frac{y^2}{\pi}}\, dy, \qquad t \geq 0,$$

for $0 < \lambda(A) < \infty$, and Q any proba measure absolutely continuous w.r.t. λ.

Remark 3 *The existence of DK-sets is not a necessary condition for the conclusion of the theorem to hold. There are weaker conditions of the same flavour which suffice (e.g. in [A0] and [AZ]), and it is also possible to use other types of assumptions (see e.g. [TZ], [Z3]).*

Remark 4 *Still, some structural assumptions have to be made: There are conservative ergodic m.p.t.s for which*

$$\frac{1}{a_n(Y)}\mathbf{S}_n(A) \xrightarrow{\mu} 0 \quad \text{for all } A, Y \in \mathcal{A}^+,$$

meaning that the averages $a_n(Y)$ fail to capture the size of occupation times for any set A, even in the distributional sense (cf. Prop 3.3.4 of [A0]).

Duality. We finally mention what the duality between infinite spaces and non-integrable functions amounts to in the case of the DK-theorem. For simplicity we focus on the case $\alpha = 1$. This can be thought of as the threshold value, where the measure "has just become infinite". In fact, some properties of finite measure preserving systems remain true, in a weaker sense, for these *barely infinite invariant measures*. Specifically, while we know that an a.e. ergodic theorem of the form $\frac{1}{a_n}\mathbf{S}_n(A) \to \mu(A)$ is impossible, the $\alpha = 1$ case of the DK-theorem gives (in particular)

$$\frac{1}{a_n(T)}\mathbf{S}_n(Y) \xrightarrow{\mu} \mu(Y), \tag{4.17}$$

i.e. a weak law of large numbers for T. We now claim that this is equivalent to

$$\frac{1}{b_m(T)}\varphi_m \xrightarrow{\mu} \frac{1}{\mu(Y)}, \tag{4.18}$$

where $(b_m(T))$ is asymptotically inverse to $(a_n(T))$ (i.e. $a(b(m)) \sim b(a(m)) \sim m$) and regularly varying of index 1. In fact, this is a nice exercise, where you should assume the existence of such $b_m(T)$, which is guaranteed by the general theory of regular variation, and exploit the duality rule

$$\mathbf{S}_n(Y) \leq m \quad \text{iff} \quad \varphi_m \geq n \tag{4.19}$$

(which is pretty obvious once you read it aloud: the number of visits (including time 0) to Y before time n does not exceed m iff the mth return does not take place before time n.)

5 Back to Gauss (inducing used the other way)

A very simple map and its invariant measure. This section is devoted to a single example, $T : X \to X$ where $X := (0, \infty)$. It looks innocent enough:

$$Tx := \begin{cases} \frac{1}{x} - 1 & \text{for } x \in (0,1) =: Y, \\ x - 1 & \text{for } x \in (1, \infty) = Y^c. \end{cases} \tag{5.1}$$

Let's see if we can find a suitable measure μ which has a density h w.r.t. λ, i.e. can be expressed as $\mu(0,x] = \int_0^x h(t)\,dt =: H(x)$. Such a μ is T-invariant iff $\mu(0,x] = \mu(T^{-1}(0,x])$ for all $x > 0$. Straightforward calculation shows that $T^{-1}(0,x] = (\frac{1}{1+x}, 1+x]$, so that the condition for invariance can be expressed in terms of the distribution function H as

$$H(1+x) - H(x) = H\left(\frac{1}{1+x}\right) \qquad \text{for all } x > 0. \tag{5.2}$$

It is not hard to find solutions to this functional equation, if we observe that an additive operation outside should correspond to a multiplicative operation inside. As a first attempt we might try $\log x$, which does not solve (5.2), but already leads to something similar. Playing around a bit one finds that

$$H(x) = H_c(x) := c \cdot \log(1+x) \qquad \text{solves (5.2)}$$

(where $c > 0$ is any constant). Choosing $c := \frac{1}{\log 2}$, so that $H(1) = 1$, we have therefore found an invariant measure μ for T with $\mu(Y) = 1$, $\mu(X) = \lim_{x\to\infty} H(x) = \infty$, and density

$$h(x) = \frac{1}{\log 2}\frac{1}{1+x}, \qquad x > 0.$$

Well, we have seen this density before[9]! Its restriction to Y is the famous invariant density for the continued fraction (CF) map found by Gauss! In fact, it is easily seen that Y is a sweep-out set for T, so that we can induce on this subset and consider the return map $T_Y = T^{\varphi}$. By Prop 4 we know at once that $\mu\mid_Y$ (resp. $h\mid_Y$) is invariant for T_Y. This restriction being finite, we can conclude via Prop 3 that T is conservative.

Continuing with continued fractions. Let's have a closer look at the induced system. We find that $\varphi = n$ on $(\frac{1}{n+1}, \frac{1}{n})$, that is, φ coincides with the CF digit function d on Y! Moreover, T_Y turns out to be the *CF-map*,

$$T_Y x = \frac{1}{x} - \left\lfloor \frac{1}{x} \right\rfloor. \tag{5.3}$$

We have thus found Gauss' density by simple considerations about the infinite measure preserving map T! Note that T is not just some arbitrary m.p. transformation containing the CF-map, but that it has a very simple interpretation from the CF point of view, as it simply means to determine the fractional part of $\frac{1}{x}$ by successively subtracting 1 until we end up with a number in $(0,1)$,

$$\begin{aligned} x = \frac{1}{\frac{1}{x}} &= \frac{1}{1+\left(\frac{1}{x}-1\right)} = \frac{1}{2+\left(\frac{1}{x}-2\right)} = \ldots = \frac{1}{\mathrm{d}+\left(\frac{1}{x}-\mathrm{d}\right)} \\ &= \frac{1}{1+Tx} \quad = \frac{1}{2+T^2x} \quad = \ldots = \frac{1}{\mathrm{d}+T^{\mathrm{d}}x} \quad = \frac{1}{\mathrm{d}+T_Y x}. \end{aligned}$$

As T_Y is known to be ergodic, we can use Prop 3 to see that T is ergodic as well (which is equivalent to saying that the H_c are the only distribution functions solving (5.2), which can also be proven directly). But we can learn

[9] In Cor Kraaikamp's lectures on Ergodic Theory of Numbers.

even more about the CF-map T_Y by studying T: recall that the consecutive digits d_k of the CF-expansion of $x \in Y$,

$$x = \frac{1}{\mathrm{d}_1 + \frac{1}{\mathrm{d}_2 + \ldots}}$$

are given by $\mathrm{d}_j := \mathrm{d} \circ \mathrm{T}_Y^{j-1}$, $j \geq 1$, and hence $\mathrm{d}_1 + \ldots + \mathrm{d}_m = \varphi_m$. As $\mu(X) = \infty$ we know that $\int_Y \varphi \, d\mu = \int_Y \mathrm{d} \, d\mu = \infty$ (cf. Kac' formula (3.11) above). As a consequence,

$$\frac{\mathrm{d}_1 + \ldots + \mathrm{d}_m}{m} = \frac{\varphi_m}{m} = \frac{\mathbf{S}_m^Y(\varphi)}{m} \longrightarrow \infty \qquad \text{a.e. on } Y \tag{5.4}$$

(exercise!). We can, however, easily compute the tail behaviour of the return distribution. Since $Y \cap \{\varphi > n\} = (0, \frac{1}{n+1})$, we find that

$$\mu(Y \cap \{\varphi > n\}) = H\left(\frac{1}{n+1}\right) = \frac{1}{\log 2} \log\left(1 + \frac{1}{n+1}\right) \sim \frac{1}{\log 2} \frac{1}{n},$$

and therefore $w_n(Y) = \sum_{k=0}^{n-1} \mu(Y \cap \{\varphi > n\}) \sim \frac{1}{\log 2} \log n$, which is slowly varying. As Y can be shown to be a DK-set for T, we are in the situation of Theorem 5, with $\alpha = 1$, and conclude (computing the $a_n(T)$ via Prop 7) that

$$\frac{1}{\log 2} \frac{\log n}{n} \mathbf{S}_n(Y) \stackrel{\mu}{\longrightarrow} 1.$$

Dually (cf. the end of the preceding section), we have

$$\frac{\log 2}{m \log m} \varphi_m \stackrel{\mu}{\longrightarrow} 1.$$

Stated in terms of CF-digits, this gives a famous classical result (cf. [Kh]):

Proposition 9 (Khinchin's Weak Law for CF-digits) *The CF-digits* (d_j) *satisfy*

$$\frac{\mathrm{d}_1 + \ldots + \mathrm{d}_m}{m \log m} \stackrel{\mu}{\longrightarrow} \frac{1}{\log 2}. \tag{5.5}$$

Remark 5 *Well, you may not like it, but I can admit it now: We have seen* T *before. It is exactly the Parry-Daniels map we studied earlier, but in different coordinates (cf. [P]).*

6 Thinking big (is even harder)

Welcome to the jungle! We have seen that something as basic as the behaviour of occupation times $\mathbf{S}_n(A)$ of sets $A \in \mathcal{A}^+$ is astonishingly complicated as soon as we are in the realm of infinite ergodic theory. Perhaps you are even willing to regard it as interesting, and some of the results as beautiful. But then, you may ask, perhaps we have just been looking at the wrong sort of sets. The space is infinite after all, so why bother about tiny sets which our orbits are hardly ever going to visit? *Shouldn't we really focus on occupation times of infinite measure sets A? Perhaps these behave in a more satisfactory way?*

The opposite is true. There seem to be no limits to the ways in which infinite measure sets can fool our intuition! We conclude these notes with a few examples.

Example 9 *A very silly one is this: Take your favorite c.e.m.p. T on a σ-finite measure space $(X, \mathcal{A}, \mu)$, with $\mu(X) = \infty$. Fix any $Y \in \mathcal{A}^+$, and consider $A := Y^c$ which has infinite measure. Theorem 1 immediately shows that $\frac{1}{n}\mathbf{S}_n(A) \to 1$ a.e. since the small set Y is only visited with zero frequency as $n \to \infty$. Still, this convergence need not be uniform, for example if T is continuous near some fixed point x in the interior of Y which has μ-positive neighborhoods.*

A minimal requirement will be to ask for $\mu(A) = \mu(A^c) = \infty$.

Example 10 *But then, we could e.g. look at the coin-tossing random walk on the integers, and take A to be the infinite-measure set corresponding to the even integers. There is an obvious cyclic structure, by which even states are always followed by odd ones, and vice versa. Hence $\mid \mathbf{S}_n(A) - \frac{n}{2} \mid \leq 1$, i.e. in this proper infinite measure situation we get $\frac{1}{n}\mathbf{S}_n(A) \to \frac{1}{2}$ uniformly on X.*

That's not so bad, is it? But consider this beautiful example:

Example 11 *Take the coin-tossing random walk again, and A the set corresponding to the positive integers. In fact, we can also consider Boole's transformation and A the positive half-line. A pretty symmetric situation, right? So we are tempted to bet that again $\frac{1}{n}\mathbf{S}_n(A) \to \frac{1}{2}$ in some sense, and it is trivially true (by symmetry) that the expectation $\int_Y \frac{1}{n}\mathbf{S}_n(A)\, d\mu_Y \to \frac{1}{2}$ (where Y corresponds to the origin or the reference set we used for Boole). However, the pointwise behaviour once again is terrible, in that*

$$\varliminf_{n\to\infty} \frac{1}{n}\mathbf{S}_n(A) = 0 \quad \textit{and} \quad \varlimsup_{n\to\infty} \frac{1}{n}\mathbf{S}_n(A) = 1 \quad \textit{a.e. on } X. \tag{6.1}$$

Still, this could be due to very rarely occurring deviations. But it isn't. This "negative" result is accompanied by a very neat and counterintuitive distributional limit theorem, showing that, for a given large time horizon n, orbits are very likely to significantly favour one side. It is called the arcsine law,

$$Q\left(\left\{\frac{1}{n}\mathbf{S}_n(A) \leq t\right\}\right) \longrightarrow \frac{2}{\pi}\arcsin\sqrt{t}, \quad t \in [0,1], \tag{6.2}$$

where Q is any proba measure absolutely continuous w.r.t. λ.

Both statements (6.1) and (6.2) are but special cases of abstract limit theorems. The distinctive structure of this example is the following: the infinite measure sets A and A^c are *dynamically separated* by some (special !) finite measure set Y, i.e. T-orbits can't pass from one set to the other without visiting Y. In the case of the random walk, (6.2) is a classical result (cf. Chapter III of [F]). The first dynamical-systems version (which covers Boole's transformation) has been given in [T3]. Further generalizations can be found in [TZ] and [Z3]. In view of the Darling-Kac theorem above, you won't be surprised to learn that regular variation again plays a crucial role.

As for the pointwise statement (6.1), one might hope for a fairly general result like Theorem 2. However, the situation is quite delicate, and this conclusion depends on strong (!) mixing conditions for the return map T_Y. See [ATZ], where we also give a Markov-chain counterexample in which (6.1) fails, and another funny example which satisfies a weak law of large numbers, $\mathbf{S}_n(A)/c_n \stackrel{\mu}{\longrightarrow} 1/2$, but with $\underline{\lim}\, c_n/n = 0$ and $\overline{\lim}\, c_n/n = 1$.

References

[A0] J. Aaronson: *An Introduction to Infinite Ergodic Theory.* AMS 1997.

[A1] J. Aaronson: *An Overview of Infinite Ergodic Theory.* available at `http://www.math.tau.ac.il/~aaro/`

[A2] J. Aaronson: *Rational ergodicity and a metric invariant for Markov shifts*. Isr. J. Math. **27** (1977), 93-123.

[A3] J. Aaronson: *The asymptotic distributional behaviour of transformations preserving infinite measures*. Journal d'Analyse Mathématique **39** (1981), 203-234.

[ATZ] J. Aaronson, M. Thaler, R. Zweimüller: *Occupation times of sets of infinite measure for ergodic transformations*. Ergod. Th. & Dynam. Sys. **25** (2005), 959-976.

[AZ] J. Aaronson, R. Zweimüller: *Limit theory for some positive stationary processes*. preprint 2008.

[AW] R. L. Adler, B. Weiss: *The ergodic infinite measure preserving transformation of Boole*. Isr. J. Math. **16** (1973), 263-278.

[BGT] N. H. Bingham, C. M. Goldie, J. L. Teugels: *Regular Variation.* Cambridge University Press 1989.

[B] G.D. Birkhoff: *Proof of the ergodic theorem*. Proc. Nat. Acad. Sci. **17** (1931), 656-660.

[D] H.E. Daniels: *Processes generating permutation expansions*. Biometrika **49** (1962), 139-149.

[DK] D. A. Darling, M. Kac: *On occupation times for Markoff processes.* Trans. Amer. Math. Soc. **84** (1957), 444-458.

[F] W. Feller: *An introduction to probability theory and its applications, Vol. I.* 3^{rd} ed., Wiley, New York, 1968.

[H] E. Hopf: *Ergodentheorie.* Springer-Verlag Berlin 1937.

[Ka] S. Kakutani: *Induced measure preserving transformations.* Proc. Imp. Acad. Sci. Tokyo **19** (1943), 635-641.

[KK] T. Kamae, M. Keane: *A simple proof of the ratio ergodic theorem.* Osaka J. Math. **34** (1997), 653-657.

[Kh] A.Ya. Khinchin: *Metrische Kettenbruchprobleme.* Compositio Math. **1** (1935), 361-382.

[Kr] U. Krengel: *Ergodic Theorems.* de Gruyter 1984.

[N] J. von Neumann: *Proof of the quasi-ergodic hypothesis.* Proc. Nat. Acad. Sci. **18** (1932), 70-82.

[P] W. Parry: *Ergodic properties of some permutation processes.* Biometrika **49** (1962), 151-154.

[S] W. Stepanoff: *Sur une extension du theoreme ergodique.* Compos. Math. **3** (1936), 239-253.

[T0] M. Thaler: *Infinite Ergodic Theory.* course notes from "The Dynamic Odyssey", CIRM 2001. available at `http://www.sbg.ac.at/mat/staff/thaler/thaler_english.htm`

[T1] M. Thaler: *Estimates of the invariant densities of endomorphisms with indifferent fixed points.* Isr. J. Math. **37** (1980), 303-314.

[T2] M. Thaler: *Transformations on [0,1] with infinite invariant measures.* Isr. J. Math. **46** (1983), 67-96.

[T3] M. Thaler: *A limit theorem for sojourns near indifferent fixed points of one-dimensional maps.* Ergod. Th. & Dynam. Sys. **22** (2002), 1289-1312.

[TZ] M. Thaler, R. Zweimüller: *Distributional limit theorems in infinite ergodic theory.* Probab. Theory Relat. Fields **135** (2006), 15-52.

[W] P. Walters: *An Introduction to Ergodic Theory.* Springer 1982.

[Zu] J.D. Zund: *George David Birkhoff and John von Neumann: A Question of Priority and the Ergodic Theorems.* Historia Mathematica **29** (2002), 138-156.

[Z1] R. Zweimüller: *Ergodic properties of infinite measure preserving interval maps with indifferent fixed points.* Ergod. Th. & Dynam. Sys. **20** (2000), 1519-1549.

[Z2] R. Zweimüller: *Hopf's ratio ergodic theorem by inducing.* Colloq. Math. **101** (2004), 289-292.

[Z3] R. Zweimüller: *Infinite measure preserving transformations with compact first regeneration.* Journal d'Analyse Mathématique **103** (2007), 93-131.

Needless to say, the present notes overlap with some of the references above, most notably with [T0], and reflect ideas conceived through discussions with several colleagues. Therefore, if you find mistakes, blame them. - Ahm, no, better not. Really, all errors in here are mine.