An L^1 Counting Problem in Ergodic Theory

Idris Assani,Department of Mathematics,
University of North Carolina at Chapel Hill,
Chapel Hill, North Carolina 27599, USA
email: assani@email.unc.edu
`www.math.unc.edu/Faculty/assani`

Zoltán Buczolich,* Department of Analysis, Eötvös Loránd
University, Pázmány Péter Sétány 1/c,1117 Budapest, Hungary
email: buczo@cs.elte.hu
`www.cs.elte.hu/∼buczo`

and

R. Daniel Mauldin,† Department of Mathematics,
University of North Texas, Denton, Texas 76203-1430, USA
email: mauldin@unt.edu
`www.math.unt.edu/∼mauldin`

March 25, 2004

---

*This work was completed while this author visited the Departments of Mathematics of University of North Texas and of University of North Carolina at Chapel Hill.

†Supported in part by NSF grant DMS 0100078

2000 Mathematics Subject Classification: Primary 37A05; Secondary 28D05, 47A35, 60F99.

Keywords: ergodic theorem, weak maximal inequality, return time theorems

Abstract

We solve the following counting problem for measure preserving transformations. For $f \in L^1_+(\mu)$, is it true that $\sup_n \dfrac{\mathbf{N}_n(f)(x)}{n} < \infty$, where

$$\mathbf{N}_n(f)(x) = \#\left\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\right\}?$$

One of the consequences is the nonvalidity of J. Bourgain's Return Time Theorem for pairs of (L^1, L^1) functions.

1 Introduction

Let $(X, \mathcal{B}, \mu)$ be a probability measure space, T an invertible measure preserving transformation on this space and $f \in L^1_+(\mu)$. Since $\frac{f(T^n x)}{n} \to 0$ a.e., the following function

$$\mathbf{N}_n(f)(x) = \#\left\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\right\}$$

is finite a.e. In this paper we consider the following

Counting Problem I. *Given $f \in L^1_+(\mu)$ do we have* $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} < \infty$, *$\mu$ a.e.?*

In [2] and [3] the maximal operator $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ was introduced and the pointwise convergence of $\frac{\mathbf{N}_n(f)(x)}{n}$ was studied. It was shown there that if $f \in L^p_+$ for $p > 1$, or $f \in L \log L$ and the transformation T is ergodic, then $\frac{\mathbf{N}_n(f)(x)}{n}$ converges a.e. to $\int f d\mu$. It was also shown that if T is not ergodic, then the limit is the conditional expectation of the function f with respect to the σ field of invariant sets for T. Hence, the limit is the same as the limit of the ergodic averages $\frac{1}{N}\sum_{n=1}^{N} f(T^n x)$. Now, the limit of the ergodic averages, by Birkhoff's pointwise ergodic theorem, exists for any function $f \in L^1(\mu)$. So, it is natural to ask whether $\frac{\mathbf{N}_n(f)(x)}{n}$ also converges a.e., when $f \in L^1(\mu)$. Another motivation for this question is given by the fact that for i.i.d. random variables $X_n \in L^1$ it was shown in [2] that

$$\frac{\#\{k : \frac{X_k(\omega)}{k} > \frac{1}{n}\}}{n}$$

converges a.e. to $E(X_1)$. In analogy with this fact, one might guess counting problem I has a positive answer whereas we show the answer is no. This counting problem was afterwards discussed by Jones, Rosenblatt and Wierdl in [11].

One can see by using the methods of [2], for instance, that the convergence for all functions $f \in L^1_+(\mu)$ will be guaranteed if one can answer the following equivalent maximal type inequality problem.

Counting Problem II. *Does there exist a finite positive constant C such that for all measure preserving systems and all $\lambda > 0$ and all $f \in L^1_+$:*

$$\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} \leq \frac{C}{\lambda}\|f\|_1?$$

In Theorem 4 we will be to show that this equivalent problem has a negative answer. This will imply our main result

Theorem 1. *In any nonatomic, ergodic system $(X, \mathcal{B}, \mu, T)$ there exists $f \in L^1_+$ such that $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} = \infty$ almost everywhere.*

We will also derive answers to some related problems. The first consequence, linked to the study of the maximal function
$\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, is what we call the return times for the tail (of the Cesaro averages).

Definition 1. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving system. The Return Times for the Tail Property holds in $L^r(\mu)$, $1 \leq r \leq \infty$ if for each $f \in L^r(\mu)$ we can find a set X_f of full measure such that for all $x \in X_f$ for all measure preserving systems $(Y, \mathcal{G}, \nu, S)$ and each $g \in L^1(\nu)$ the sequence $\frac{f(T^n x)\cdot g(S^n y)}{n}$ converges to zero for a.e. y.

Using Theorem 1 and results from [2] it is not difficult to show that the next corollary holds. The details of this verification can also be found in [5].

Corollary 2. *The Return Times for the Tail Property does not hold for $r = 1$.*

We observe that in [2] and [3] it was shown that the Return Times for the Tail Property holds in L^p for $1 < p \leq \infty$ and even in $L \log L$.

A second consequence is a solution to the (L^1, L^1) problem mentioned in [2], [4] and [14]. To explain this problem we need a few definitions.

Definition 2. A sequence of scalars a_n is said to be good universal for the pointwise ergodic theorem (resp. norm convergence) in L^r, $1 \leq r \leq \infty$ if for all dynamical systems $(Y, \mathcal{G}, \nu, S)$ the averages

$$\frac{1}{n} \sum_{k=1}^{n} a_k \cdot g(S^k y)$$

converge a.e. (resp. in $L^r(\nu)$ norm).

In [6], [7], and [8] J. Bourgain showed that given $f \in L^\infty(\mu)$ the sequence $f(T^n x)$ is μ a.e. good universal for the pointwise convergence in L^1. Using Hölder's inequality and the maximal inequality for the ergodic averages one can extend his result to the pairs (L^p, L^q) where $\frac{1}{p} + \frac{1}{q} = 1$. This was mentioned in [13]. Bourgain's Return Time Theorem strengthens Birkhoff's theorem on the product space when the functions, f and g, respect duality. That is, if the function $f \in L^p(\mu)$ for some $1 \leq p \leq \infty$, then the set of convergence obtained from the Return Times Theorem works for all functions $g \in L^q(\nu)$, where $\frac{1}{p} + \frac{1}{q} = 1$, hence it is a universal set. However, fixing f and g, the projection of the convergence set onto the first factor obtained by Birkhoff's theorem depends on both functions. A weakness of the Return Time Theorem is that it does not address the case of $f \in L^p$ and $g \in L^q$ when $\frac{1}{p} + \frac{1}{q} > 1$. In particular, it does not address the case where $f \in L^1$ and $g \in L^1$ while Birkhoff's theorem, on the other hand, guarantees convergence for $f \otimes g \in L^1 \times L^1$, $\mu \otimes \nu$-almost everywhere.

In [4] random stationary weights (i.i.d. random variables) were given for which one could go "beyond" the duality apparently imposed by the use of Hölder's inequality.

It was also shown that given $f \in L^1(\mu)$ the sequence $(f(T^n x))$ is μ-a.e. good universal for the L^1 norm. In [2] a Multiple Return Times Theorem for L^1 i.i.d. random variables was obtained while in [14] a Multiple Return Times theorem was proved for L^∞ stationary processes. The (L^1, L^1) problem was the following.

(L^1, L^1) **Problem.** *Given $f \in L^1(\mu)$, is the sequence $(f(T^n x))$, μ-a.e. good universal for the pointwise ergodic theorem in L^1?*

A consequence of Theorem 1 is the following solution to the (L^1, L^1) problem (see also [5])

Corollary 3. *Bourgain's Return Time Theorem does not hold for pairs of (L^1, L^1) functions.*

We also derive in Section 3 some consequences in $L^1(\mathbb{T})$ between the continuous analog of the maximal function $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, namely

$$A(f)(x) = \sup_t t \cdot m \left\{ 0 < y < x : \frac{f(x-y)}{y} > t \right\},$$

or, analogously,

$$A(f)(x) = \sup_t t \cdot m \left\{ 0 < y < x : \frac{f(y)}{x-y} > t \right\},$$

and the one sided Hardy–Littlewood maximal function.

2 Proof of Theorem 1

Before presenting the technical details of the proof of Theorem 1 we try to explain the main ideas behind the proof.

2.1 Heuristic explanation of the proof

The main technical aspect of the proof is contained in the following theorem which shows there is no constant C satisfying the requirements of counting problem II.

Theorem 4.

$$\sup_{(X,\mathcal{B},\mu,T)} \sup_{\|f\|_1=1} \sup_{\lambda>0} \lambda \cdot \mu \left\{ x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda \right\} = \infty.$$

To prove Theorem 4 we construct some periodic rotations T_p on $X = [0, 1)$ and functions ϕ_p with L^1 norm one such that

$$\sup_{(X,\mathcal{B},\mu,T_p)} \sup_{\lambda>0} \lambda \cdot \mu \left\{ x : \sup_n \frac{\mathbf{N}_n(\phi_p)(x)}{n} > \lambda \right\} = \infty. \tag{1}$$

From this one can conclude that for any given nonatomic ergodic system $(X, \mathcal{B}, \mu, T)$ we must have

$$\sup_{\|f\|_1=1} \sup_{\lambda>0} \lambda \cdot \mu \left\{ x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda \right\} = \infty.$$

(If we had a finite constant C for one nonatomic ergodic system then this constant would be the same for all measure preserving transformations). This in turn implies we can find in this system a function $f \in L^1_+$ for which $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} = \infty$ almost everywhere.

Now we explain how to establish Theorem 4. First observe that if the supports of two functions f_1 and f_2 are disjoint then

$$\mathbf{N}_n(f_1 + f_2)(x) = \mathbf{N}_n(f_1)(x) + \mathbf{N}_n(f_2(x)). \tag{2}$$

In the next lemmas when we build more and more complicated systems there will be some technical assumptions about the support of the functions constructed in order to ensure that we can use the above additivity of the counting function.

If one fixes n then for any $f \in L^1_+$ it is not difficult to see that

$$\mu \left\{ x : \frac{\mathbf{N}_n(f)(x)}{n} > \lambda \right\} \le \frac{\int_X f d\mu}{\lambda}. \tag{3}$$

Indeed, if $f(x) = c\mathbf{1}_A(x)$ for a constant c and a measurable set A then

$$\frac{\mathbf{N}_n(f)(x)}{n} = \frac{\#\{k : \frac{c\mathbf{1}_A(T^k x)}{k} > \frac{1}{n}\}}{n} = \frac{1}{n} \sum_{\{1 \le k < nc\}} \mathbf{1}_A(T^k x).$$

Hence,

$$\int_X \frac{\mathbf{N}_n(f)(x)}{n} d\mu(x) \le c\mu(A) = \int f d\mu. \tag{4}$$

Now by using (2) we can verify (4) for step functions and then by an approximation argument for arbitrary $f \in L^1_+$. By Markov's inequality one can deduce (3).

The bottom line of (3) is that we have L^1 limitations on the size of the set where we can make $\frac{\mathbf{N}_n(f)(x)}{n}$ large for a fixed n.

Now assume that A is measurable, $f(x) = \mathbf{1}_A(x)$ and the sets $T^{-k}A$ are disjoint for $k = 1, ..., n_0$. Then for any $x \in T^{-k}A$ we have $\mathbf{N}_n(f)(x) \geq 1$ for any $n \geq k+1$ and hence $\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n} \geq \frac{1}{k+1}$.

This means that if n_0 is sufficiently large then at the price of $\int f = \mu(A)$ we have a "harmonic series lower estimate" for the maximal function, namely $\mathbf{N}^*(f)(x) \geq 1/2$ on a set of measure $\mu(A)$, then $\mathbf{N}^*(f)(x) \geq 1/4$ on a set of measure $2\mu(A)$, and in general $\mathbf{N}^*(f)(x) \geq 1/2^l$ on a set of measure $2^{l-1}\mu(A)$. Hence if n_0 is so large that we have estimates of this type for $l = 1, ..., M$ then we have $\int \mathbf{N}^*(f)(x)d\mu(x) \geq \sum_{l=1}^{M} \frac{1}{2^l} 2^{l-1}\mu(A) = \frac{M}{2}\mu(A)$. This M will be called the gain constant of the construction.

We will use this type of construction with different functions f_h, $h = 1, ..., k$ which have different supports and will add these estimates together. The main difficulty is that for an x depending on h we have different "life times", n_h when $\frac{N_{n_h}(f_h)(x)}{n_h}$ is sufficiently large. In order to have a large value of the maximal function $\mathbf{N}^*(f)(x)$ for a given x we want to find a time n_x when $\frac{N_{n_x}(f_h)(x)}{n_x}$ is large simultaneously for many h's. To coordinate these "life times" for different x's and h's a concept, called *life function* is introduced in the course of the proof.

Finally, an argument based on probability theory is used. Identically distributed independent random variables, X_h are introduced with expected values, $\int X_h$ being equal to some constant times the gain constant, M, times $\int f_h$. The values of these X_h are distributed according to the above "harmonic series lower estimate" scheme. By using the controls established on the "life times" of our individual systems f_h, $h = 1, ..., k$ for every x we will have $N^*(\sum_{h=1}^{k} f_h)(x) \geq \sum_{h=1}^{k} X_h(x)$. Then an application of the weak law of large numbers together with the fact that we gain a factor of M coming from the gain constant will complete the proof of Theorem 4.

2.2 Main lemma

In the remaining subsections of Section 2, μ will denote the Lebesgue measure on $\mathbb{R}$. An interval I is a 2^{-R} grid interval if there is some $j \in \mathbb{Z}$ such that

$I = [j \cdot 2^{-R}, (j+1)2^{-R})$.

We say that a random variable $X : I \to \mathbb{R}$ is $(M-0.99)$-distributed on I if $X(x) \in \{0, 0.99, 0.99 \cdot 2^{-1}, ..., 0.99 \cdot 2^{-M+1}\}$ and $\mu(\{x : X(x) = 0.99 \cdot 2^{-l+1}\}) = 0.99 \cdot 2^{-M+l-1}\mu(I)$, for $l = 1, ..., M$. This is a slight abuse of terminology since the measure of the whole space is $\mu(I)$, but in our final application we will have $I = [0, 1)$ with measure one, so we will really talk about random variables.

The next lemma is the main lemma in the proof of Theorem 4. A system (T, f) satisfying this lemma will be referred to as a level k system.

We assume that a positive integer M, the gain constant, is fixed.

There are quite a few technical assumptions in the statement of this lemma. It is stated for a general 2^{-R} grid interval I_0, but we will use it in the end when $I_0 = [0, 1)$. The reason of the introduction of I_0 is that we prove this lemma by induction on k and during the induction steps we will have to put level k-systems onto some subintervals I_0 of $[0, 1)$. The induction argument explains why we state this lemma with constants explicitly depending on k. Finally, the constant D and the technical assumption (6) will ensure that in the induction steps we are adding together functions with disjoint support.

Lemma 5. *For any 2^{-R} grid interval I_0, positive integer $k \leq 2^D$, and any startup time $K_S^{(k)} > \max\{10, M\}$ there exists $J_0 > 0$ such that for all $J \geq J_0$ we can find a system (T, f) with the following properties. The transformation T is given by $T(x) = x + 2^{-J}$. We have independent $(M-0.99)$-distributed random variables X_h, $h = 1, ..., k$, on I_0 and an exit time $K_e^{(k)}$ such that for any $x \in I_0$ there exists an $n \in [2^{K_S^{(k)}}, 2^{K_e^{(k)}}]$, for which*

$$\frac{\mathbf{N}_n(f)(x)}{n} \geq \sum_{h=1}^{k} X_h(x). \tag{5}$$

Moreover, $f \geq 0$ is constant on the intervals of the form $[i \cdot 2^{-J}, (i+1)2^{-J})$, $\int_{I_0} f = k \cdot 2^{-M+1}\mu(I_0)$, $f(x) = 0 = X_h(x)$ for $x \notin I_0$, $h = 1, ..., k$. We also may require that

$$\textit{if } x \notin \bigcup_{l=0}^{k-1} \bigcup_{i \in \mathbb{Z}} [(i \cdot 2^D + l)2^{-J}, (i \cdot 2^D + l + 1) \cdot 2^{-J}), \textit{ then } f(x) = 0. \tag{6}$$

The proof of this lemma will be done in Subsection 2.5. In the next subsection we will see how this lemma can be used to prove Theorem 1.

2.3 Proof of Theorem 1

Proof. Suppose M and $0 < \gamma < 1$ are fixed. (One should think of γ as a number close to one.) We assume $X = I_0 = [0, 1)$ and consider independent $(M - 0.99)$-distributed random variables, X_h for $h = 1, ..., K$ for a sufficiently large K. Assume that u denotes the mean of these variables. An easy calculation shows that

$$u = \int X_h(x) d\mu(x) = \sum_{l=1}^{M} 0.99^2 \cdot 2^{-l+1} \cdot 2^{-M+l-1} > 0.9 \cdot M \cdot 2^{-M}. \tag{7}$$

By the weak law of large numbers

$$\mu \left\{ x : \left| \frac{1}{K} \sum_{h=1}^{K} X_h(x) - u \right| \geq \frac{u}{2} \right\} \to 0.$$

Therefore, we can choose K so large that

$$\mu \left\{ x : \frac{1}{K} \sum_{h=1}^{K} X_h(x) \geq \frac{u}{2} \right\} > \gamma.$$

Fix such a K. Then using (7) and letting

$$\Lambda = \left\{ x : \frac{1}{K} \sum_{h=1}^{K} X_h(x) \geq \frac{0.9}{2} \cdot M \cdot 2^{-M} \right\}$$

we have $\mu(\Lambda) > \gamma$.

Apply Lemma 5 with $k = K$, D chosen so that $K < 2^D$, with startup time $K_S^{(K)} > 2^M/K$, $I_0 = X = [0, 1)$.

Set $J = J_0$ and obtain a level K system (T, f). We have $\int f = K \cdot 2^{-M+1}$. We modify slightly the definition of T by saying that $T(x) = x + 2^{-J}$ modulo one. (Since in Lemma 5 f is supported on I_0 this modification of T does not decrease $\mathbf{N}_n(f)$.) By (5) and the choice of K and Λ we have for all $x \in \Lambda$

$$\frac{2^M}{K} \sum_{h=1}^{K} X_h(x) \geq \frac{0.9}{2} \cdot M.$$

Hence for any $x \in \Lambda$ we can find $n' \geq K_S^{(K)} > 2^M/K$ for which

$$\frac{2^M}{K}\frac{\mathbf{N}_{n'}(f)(x)}{n'} \geq \frac{0.9}{2}\cdot M. \tag{8}$$

Set $\phi = \frac{2^{M-1}}{K} f$. Then $\int \phi = 1$.

Now using the definition of $\mathbf{N}_{n'}$ we obtain

$$\frac{\mathbf{N}_{n'}(f)(x)}{n'} = \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n'}\}}{n'} = \frac{\#\{k : \frac{\phi(T^k x)}{k} > \frac{1}{n' 2^{-M+1} K}\}}{n'} <$$

$$< \frac{\#\{k : \frac{\phi(T^k x)}{k} > \frac{1}{\lfloor n' 2^{-M+1} K\rfloor + 1}\}}{\lfloor n' 2^{-M+1} K\rfloor + 1} \cdot \frac{n' 2^{-M+1} K + 1}{n'} =$$

let $n = \lfloor n' 2^{-M+1} K\rfloor + 1$ and using $n' > 2^M/K$, we can continue with

$$= \frac{\mathbf{N}_n(\phi)(x)}{n}\left(2^{-M+1}K + \frac{1}{n'}\right) < \frac{\mathbf{N}_n(\phi)(x)}{n}\cdot 2\frac{K}{2^{M-1}}.$$

By (8) for all x from Λ there exist n such that $\mathbf{N}_n(\phi)(x)/n > \frac{0.9}{8}M$.

Since M can be chosen as large, and γ as close to 1, as we wish we have established (1).

The density in the weak topology of conjugates of a single ergodic transformation [10] shows that if we had a finite constant C in the second form of the Counting Problem for one nonatomic ergodic transformation then we would have the same constant for all ergodic transformations. The ergodic decomposition would give then the same constant C for all measure preserving transformations, and the equation (1) shows that such a universal finite constant does not exist.

Thus if we consider a nonatomic ergodic dynamical system we have

$$\sup_{\|f\|_1=1}\sup_{\lambda>0}\lambda\cdot\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} = \infty.$$

From Theorem 4 in [2] we can conclude that there exists a function $f \in L^1_+$ such that $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ is not finite almost everywhere. As $\lim_n \frac{\mathbf{N}_n(f)(x)}{n} - \frac{\mathbf{N}_n(f)(Tx)}{n} = 0$, the function $\limsup_n \frac{\mathbf{N}_n(f)(x)}{n}$ is T invariant. The ergodicity of T implies that actually we have $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} = \infty$ almost everywhere.

□

2.4 Basic systems

As was mentioned in Subsection 2.2 we prove Lemma 5 by induction. This section is about the existence of the so called basic systems. These systems will serve on one hand as the first step of the definition of the level k systems and on the other hand these basic systems will be used during the induction when we want to build a level $k+1$ system, the $(k+1)$st "part" of this system will be created by Lemma 6.

As we pointed out in Subsection 2.1 the main difficulty is to coordinate the times when $\mathbf{N}_n(f(x))/n$ is large for a given x. We control when this happens by the use of life functions defined below.

A "life" function is a map $\nu : \mathbb{N} \to \mathbb{N}$ such that for each $N \in \mathbb{N}$, $\nu(N) > N$. Given a life function ν, a gain constant $M > 3$, and a startup time N_1 we choose a sequence $N_2, ..., N_M$ so that

$$N_l = 20 + \nu(N_{l-1}),\ l = 2, ..., M. \tag{9}$$

In the next lemma we obtain a "harmonic series lower estimate" for $\mathbf{N}_n(f)(x)/n$ in (10). On sets Γ_l of approximate measure $2^{-M+l-1}\mu(I)$ we will have lower estimate given by (10) where the life function provides an interval of n's when this estimate holds. In the statement of Lemma 6 we will also have technical assumptions about the support constants D and S to ensure that when we build more complicated systems then we work with functions of disjoint support.

Lemma 6. *Suppose that a gain constant $M > 3$, a life function ν, support constants D, $S < 2^D$, and a startup time $N_1 > \max\{10, M\}$ are given. Choose the sequence $N_2, ..., N_M$ based on M, ν, and N_1 satisfying* (9). *Given any 2^{-R} grid interval I, there exists a positive integer $J_0 > R$, disjoint subsets $\Gamma_1, ..., \Gamma_M$ of I, and for each integer $J \geq J_0 > R$ there is a simple function $f : \mathbb{R} \to [0, +\infty)$, such that $f(x) = 0$ for $x \notin I$ and if $T(x) = x + 2^{-J}$ then for all $l = 1, ..., M$,*

$$\frac{\mathbf{N}_n(f)(x)}{n} > 0.99 \cdot 2^{-l+1} \quad \text{when } 2^{N_l} \leq n \leq 2^{\nu(N_l)} \tag{10}$$

for all $x \in \Gamma_l$. Moreover, each set Γ_l consists of the union of intervals of the form $[i \cdot 2^{-J_0}, (i+1)2^{-J_0})$, $\mu(\Gamma_l) > 0.99 \cdot 2^{-M+l-1}\mu(I)$, and $\int_I f = 2^{-M+1}\mu(I)$. We can also require that $f(x) = 0$ for any x which is not in an interval of the form $[(i \cdot 2^D + S)2^{-J}, (i \cdot 2^D + S + 1)2^{-J})$ for some $i \in \mathbb{Z}$.

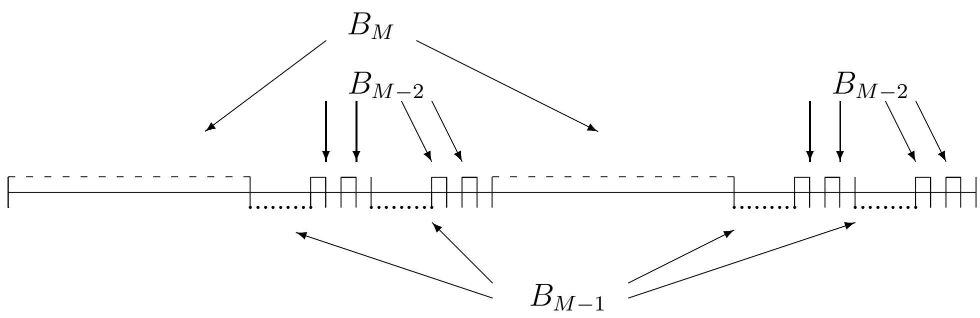

Figure 1: The sets B_M, B_{M-1}, and B_{M-2}

Proof. Set $h_0 = 2^{D+10}$ and choose J_0 such that

$$2^{10} 2^{\nu(N_M)} h_0 2^{-J_0} < 2^{-R}, \tag{11}$$

or equivalently, $J_0 > \nu(N_M) + D + 20 + R$. Now, let an integer $J \geq J_0$ be given. Set $h = h_0 \cdot 2^{J-J_0} = 2^{D+10+J-J_0}$. We shall first define a sequence of sets $B_M, B_{M-1}, \ldots, B_1$ each as the union of some intervals in a corresponding sequence of finer dyadic grids. To begin put

$$\begin{aligned} B_M =& I \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J} \cdot 2^{N_M} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_M} h) \\ =& I \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J_0} \cdot 2^{N_M} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_M} h_0). \end{aligned} \tag{12}$$

Thus, B_M consists of the intervals in the standard $2^{N_M+D-J_0}$ grid with even index, j and that are subsets of the interval I. Clearly, $\mu(B_M) = \mu(I)/2$. In Figure 1 we illustrate the manner in which the sets B_l, $l = M-2, M-1, M$, are located in I. Of course, in an illustration we cannot divide an interval into several thousand pieces, so in the figure the set B_M consists of two intervals of length $\mu(I)/4$, marked by dashed lines, B_{M-1} consists of four intervals of length $\mu(I)/16$, marked by dotted lines, B_{M-2} consists of eight intervals marked by solid lines. The complement of $B_M \cup B_{M-1} \cup B_{M-2}$ consists of eight "unmarked" intervals, each of the same length as the components of B_{M-2}.

In (12) the first expression for B_M is given for some computational purposes whereas the second expression shows that B_M does not depend on J but rather on J_0. The same is true for all the sets B_i to be defined now.

Assume that $l \in \{0, ..., M-3\}$ and $B_{M-l'}$ is given for all $l' \in \{0, ..., l\}$. Set

$$
\begin{aligned}
&B_{M-(l+1)} = \qquad (13)\\
&= (I \setminus \bigcup_{l'=0}^{l} B_{M-l'}) \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J} \cdot 2^{N_{M-l-1}} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_{M-l-1}} h) = \\
&= (I \setminus \bigcup_{l'=0}^{l} B_{M-l'}) \cap \bigcup_{j\in\mathbb{Z}} [2j \cdot 2^{-10-J_0} \cdot 2^{N_{M-l-1}} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_{M-l-1}} h_0).
\end{aligned}
$$

Thus, the set $B_{M-(l+1)}$ consists of the intervals with even index in the standard $2^{N_{M-l-1}+D-J_0}$ grid that are subsets of I and are not in $\cup_{i=M-l}^{M} B_i$.

Finally, if B_{M-l} is given for $l \in \{0, ..., M-2\}$, we set

$$
B_1 = B_{M-((M-2)+1)} = I \setminus \bigcup_{l=0}^{M-2} B_{M-l}.
$$

Returning to the illustration on Figure 1, if $M = 4$ then $B_M = B_4$ is marked by the dashed line, B_3 is by the dotted line, B_2 by the solid line, and B_1, the complement of the other three is the "unmarked" part of I.

Observe that $\mu(B_{M-l}) = \mu(I)/2^{l+1}$ holds for $l = 0, ..., M-2$ and $\mu(B_1) = \mu(I)/2^{M-1} > \mu(I)/2^M = \mu(I)/2^{(M-1)+1}$. The set B_1 is the union of some disjoint intervals of the form

$$
\begin{aligned}
&[(2j-1) \cdot 2^{-10-J} \cdot 2^{N_2} h, 2j \cdot 2^{-10-J} \cdot 2^{N_2} h) = \qquad (14)\\
&= [(2j-1) \cdot 2^{-10-J_0} \cdot 2^{N_2} h_0, 2j \cdot 2^{-10-J_0} \cdot 2^{N_2} h_0),
\end{aligned}
$$

while for any $l = 0, ..., M-2$ the set B_{M-l} is the union of some intervals of the form

$$
\begin{aligned}
&[2j \cdot 2^{-10-J} \cdot 2^{N_{M-l}} h, (2j+1) \cdot 2^{-10-J} \cdot 2^{N_{M-l}} h) = \qquad (15)\\
&= [2j \cdot 2^{-10-J_0} \cdot 2^{N_{M-l}} h_0, (2j+1) \cdot 2^{-10-J_0} \cdot 2^{N_{M-l}} h_0).
\end{aligned}
$$

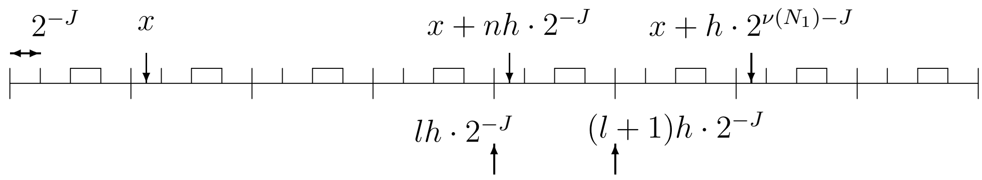

Figure 2: The definition of f in an interval I'

Our function f which depends on J will have value 0 on $\cup_{l=2}^{M} B_l$. To determine its values on B_1, consider one of the intervals making up B_1:

$$I' = [(2j-1) \cdot 2^{-10-J} \cdot 2^{N_2} h, 2j \cdot 2^{-10-J} \cdot 2^{N_2} h) \subset B_1.$$

For each $l \in \mathbb{Z}$ such that the interval $[lh \cdot 2^{-J}, (l+1)h \cdot 2^{-J}) \subset I'$ (and there are $\frac{2^{N_2}}{2^{10}}$ such l), choose exactly one l' such that $lh \leq l' < (l+1)h$, $l' \equiv S$ modulo 2^D and set $f(x) = h$ for $x \in [l' \cdot 2^{-J}, (l'+1) \cdot 2^{-J})$, otherwise we set $f(x) = 0$.

In Figure 2 one can see one interval I' being enlarged. Again we could not divide this interval in a drawing into several thousand subintervals, so in this illustration $h = 4$, and $S = 2$. One tiny interval is of length 2^{-J}, the tiny intervals marked by an extra solid line are the ones where $f = h$.

From the definition of f, we have $\int_{I'} f = \frac{h}{2^J} \cdot \frac{2^{N_2}}{2^{10}} = \mu(I')$. By summing this over all subintervals of B_1 of type I', we obtain $\int_I f = \int_{B_1} f = \mu(B_1) = \mu(I)/2^{M-1}$.

Suppose $2^{N_1} \leq n \leq 2^{\nu(N_1)}$, and

$$[x, x + h \cdot 2^{\nu(N_1)-J}) \subset I'. \tag{16}$$

Then $N_1 > 10$ implies $1000 \leq n$ and hence

$$\begin{aligned} \mathbf{N}_n(f)(x) =& \#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\} = \\ &= \#\{k : hn > k \text{ and } f(T^k x) = h\} > 0.99 \cdot \frac{nh}{h} = 0.99n, \end{aligned}$$

Of course, instead of 0.99 we could have used 0.999, but this is not of any consequence for our purposes.

Now, we define the sets Γ_i which do not depend on J from the sets B_i. To begin set

$$\begin{aligned}\Gamma_1 =&\{x \in B_1 : [x, x + h \cdot 2^{\nu(N_1)-J}) \subset B_1\} = \\ =&\{x \in B_1 : [x, x + h_0 \cdot 2^{\nu(N_1)-J_0}) \subset B_1\}.\end{aligned} \tag{17}$$

Again, the second expression here shows that Γ_1 does not depend on J since B_1 does not depend on J. For each interval I' making up B_1, by using (9), we have $|\Gamma_1 \cap I'| \geq |I'| - h \cdot 2^{\nu(N_1)-J} \geq |I'| \cdot (1 - 2^{-(N_2-\nu(N_1)-10)}) > 0.99|I'|$. So, $\mu(\Gamma_1) > 0.99 \cdot \mu(B_1)$.

Observe that for each $l = 1, ..., M-2$, the set $I \setminus \cup_{i=0}^{l-1} B_{M-i} = \cup_{i=1}^{M-l} B_i$ is the union of some intervals of the form

$$I'_{M-l} = [(2j-1) \cdot 2^{-10-J} \cdot 2^{N_{M-l+1}} h, 2j \cdot 2^{-10-J} \cdot 2^{N_{M-l+1}} h).$$

Also, the two sets B_{M-l} and $B_1 \cup \ldots \cup B_{M-l-1}$ are equally distributed in I'_{M-l} in the sense that if one takes the $2^{N_{M-l}} h / 2^{10+J}$ grid of the interval I'_{M-l}, then every evenly indexed interval is in B_{M-l} and the others are in $B_1 \cup \ldots \cup B_{M-l-1}$. In particular, $\mu(B_{M-l} \cap I'_{M-l}) = \mu(I'_{M-l})/2 = \mu(\cup_{i<M-l} B_i \cap I'_{M-l})$.

Finally, by induction one can also see that

$$\mu(B_1 \cap I'_{M-l}) = \mu(I'_{M-l})/2^{M-l-1}, \tag{18}$$

and, more generally, if $n \in [2^{N_{M-l}}, 2^{\nu(N_{M-l})}]$ and $[x, x + nh \cdot 2^{-J}) \subset I'_{M-l}$, then

$$\mu(B_1 \cap [x, x + nh \cdot 2^{-J})) > 0.995 nh \cdot 2^{-J} / 2^{M-l-1}. \tag{19}$$

Set

$$\begin{aligned}\Gamma_{M-l} &= \{x \in B_{M-l} : [x, x + h \cdot 2^{\nu(N_{M-l})-J}) \subset \bigcup_{l' \leq M-l} B_{l'}\} = \\ &= \{x \in B_{M-l} : [x, x + h_0 \cdot 2^{\nu(N_{M-l})-J_0}) \subset \bigcup_{l' \leq M-l} B_{l'}\}.\end{aligned} \tag{20}$$

Using (9) one can see that $\mu(\Gamma_{M-l}) > 0.99 \mu(B_{M-l}) \geq 0.99 \mu(I)/2^{l+1}$.

If $x \in \Gamma_{M-l}$ and I'_{M-l} is the subinterval of $\cup_{l'=1}^{M-l} B_{l'}$ containing x, then $x + jh \cdot 2^{-J} \in I'_{M-l}$ for all $0 \leq j \leq 2^{\nu(N_l)}$.

By using (19) and the definition of $f(x)$ we have

$$\mathbf{N}_n(f)(x) = \#\{k : hn > k \text{ and } f(T^k x) = h\} \geq 0.99 \frac{nh}{h \cdot 2^{M-l-1}} = 0.99 \frac{n}{2^{M-l-1}}.$$

From $N_2 > N_1 > 10$, (12), (13), (14), (15), (17), and (20) it follows that each Γ_l is the union of intervals of the form $[i \cdot 2^{-J_0}, (i+1) \cdot 2^{-J_0})$.

□

2.5 The proof of Lemma 5

Now we are ready to prove Lemma 5.

In this subsection the gain constant $M \in \mathbb{N}$ is fixed. During the proof of Lemma 5 at each step of the induction we want to use that lower level systems exist.

Next we define the life functions for all $k \leq 2^D$. Set $\nu_1(N) = N + 1$ for any $N \in \mathbb{N}$.

We proceed by induction, so assume that for $k \in \mathbb{N}$ we have already defined ν_k.

If some $N \in \mathbb{N}$ is given use $\nu = \nu_k$ and $N_1 = N_1^{(k)}(N) = N$ in (9) to determine the sequence $N_2^{(k)}(N), ..., N_M^{(k)}(N)$.

Put $\nu_{k+1}(N) = \nu_k(N_M^{(k)}(N)) > N$. The purpose of this choice is that in this way one interval $[2^N, 2^{\nu_{k+1}(N)}]$ contains the "complete lifespan" of a k system starting at time N.

Recall that by (10) for n's in $[2^N, 2^{\nu_{k+1}(N)}]$ we have a lower estimate of $\mathbf{N}_n(f)(x)/n$ when f comes from a level k system.

Proof. To define our level 1 systems we use Lemma 6 on I_0. We apply Lemma 6 with $\nu = \nu_1$, and $N_1 = K_S^{(1)}$. So, $K_e^{(1)} = \nu_1(N_M)$ will be the exit time. We choose our $(M - 0.99)$-distributed random variable the following way. For $l = 1, ..., M$ we select a measurable set $\widehat{\Gamma}_l \subset \Gamma_l$ such that $\mu(\widehat{\Gamma}_l) = 0.99 \cdot 2^{-M+l-1} \cdot 2^{-R}$. If $x \in \widehat{\Gamma}_l$ for some l then we set $X_1(x) = 0.99 \cdot 2^{-l+1}$ and $X_1(x) = 0$ otherwise. Viewed in this way Lemma 6 guarantees that level one systems exist.

We proceed by induction on k. Assume that level k systems exist and we need to verify the existence of level $k+1$ systems as long as $k + 1 \leq 2^D$. First, calling upon Lemma 6, we define a "mother" base system, this will be the $(k+1)$st function in our construction of a system with $k+1$ functions.

The "subsystems" of this "mother" system will be level k systems with different life intervals.

Given the startup constant $N_1 = N_{1,0} = K_S^{(k+1)} > \max\{10, M\}$ putting the life function ν_{k+1} into (9) defined at the beginning of this subsection determine the sequence $N_{2,0}, ..., N_{M,0}$, (the extra 0 in subscripts will refer to the "mother system"). We also put $N_{0,0} = N_1$, and set the support constant $S_0 = k$ for the mother system.

Next we apply Lemma 6 with $\nu = \nu_{k+1}$ to the 2^{-R} grid interval $I_0 = [j_0 \cdot 2^{-R}, (j_0+1) \cdot 2^{-R})$ we choose $J_{0,0}$ and disjoint subsets $\Gamma_{1,0}, ..., \Gamma_{M,0}$ of I_0 such that for each $l = 1, ..., M$, $\Gamma_{l,0}$ consists of the union of some intervals of the form $[i \cdot 2^{-J_{0,0}}, (i+1)2^{-J_{0,0}})$, and $\mu(\Gamma_{l,0}) > 0.99 \cdot 2^{-M+l-1} \cdot 2^{-R}$. For any $J \geq J_{0,0}$ we can choose a function $\phi_0 = f : I_0 \to \mathbb{R}$, such that if $T(x) = x + 2^{-J}$ then for all $l = 1, ..., M$,

$$\frac{\mathbf{N}_n(\phi_0)(x)}{n} > 0.99 \cdot 2^{-l+1}, \quad \text{when } 2^{N_{l,0}} \leq n \leq 2^{\nu_{k+1}(N_{l,0})}, \tag{21}$$

for all $x \in \Gamma_{l,0}$. Moreover, $\int_{I_0} \phi_0 = 2^{-M+1}\mu(I_0)$. Since $S_0 = k$, we also have $\phi_0(x) = 0$ for any x which is not in an interval of the form $[(i \cdot 2^M + k)2^{-J}, ((i \cdot 2^M + k + 1)2^{-J})$ for some $i \in \mathbb{Z}$.

Next, consider the intervals $I_j = [j_0 \cdot 2^{-R} + (j-1) \cdot 2^{-J_{0,0}}, j_0 \cdot 2^{-R} + j \cdot 2^{-J_{0,0}})$ for $j = 1, ..., 2^{J_{0,0}-R}$. Our level k "subsystems" will live on these intervals.

If $I_j \subset \cup_{l=1}^M \Gamma_{l,0}$ then there is a unique $l(j)$ such that $I_j \subset \Gamma_{l(j),0}$.

If $I_j \not\subset \cup_{l=1}^M \Gamma_{l,0}$ then $I_j \cap \cup_{l=1}^M \Gamma_{l,0} = \emptyset$, and in this case we set $l(j) = 0$.

This is the key step of our construction. From $I_j \subset \Gamma_{l(j),0}$ it follows that for the "mother system" we have an estimate (21) with $l = l(j)$ for all n's in $[2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}]$.

Now the choice of ν_{k+1} implies that we can put a level k system on the interval I_j with startup time $N_{l(j),0}$.

Therefore, for our $(k+1)$st, "mother" function, we can use (21) during the whole "lifetime" of the level k system which we put on I_j. This will imply that we will be able to choose X_{k+1} so that it is constant while the X_h's for $h = 1, ..., k$ take all possible values on I_j.

Hence we will be able to ensure that X_{k+1} will be independent from the X_h's when $h \leq k$. This is illustrated schematically on Figure 3 showing that X_{k+1}, the dashed line, stays constant while X_k, thick line, takes all its values on I_j. (X_k is drawn only on I_j.)

Now we turn to the details of the estimations. By our assumption on any I_j we can find level k systems. So, for each $j \in \{1, ..., 2^{J_{0,0}-R}\}$ choose

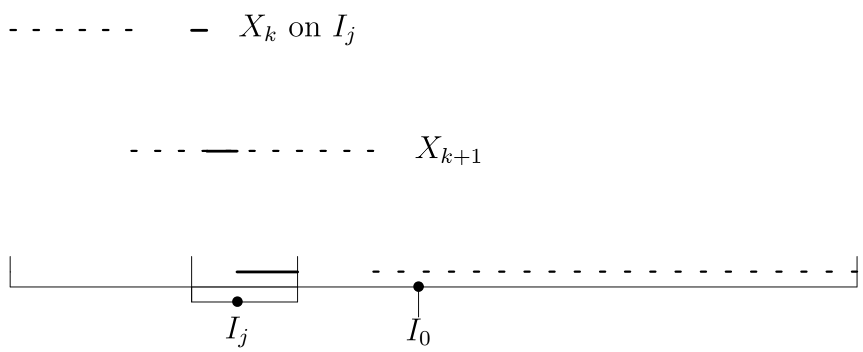

Figure 3: The scheme about the independence of the X_h's

a level k system on I_j with startup time $K_{S,j}^{(k)} = N_{l(j),0}$. Choose $J_{0,j}$ for each $j = 1, ..., 2^{J_{0,0}-R}$ according to our induction hypothesis. Set $J_0 = \max\{J_{0,j} : j = 0, ..., 2^{J_{0,0}-R}\}$ and choose a $J \geq J_0$. The transformation T will be given by $T(x) = x + 2^{-J}$. For this J choose ϕ_0 as was explained above, and by the induction hypothesis for any $j = 1, ..., 2^{J_{0,0}-R}$ choose $\phi_j = f$ and independent $(M - 0.99)$-distributed random variables $X_{h,j}$, $h = 1, ..., k$, on I_j, and an exit time $K_{e,j}^{(k)} = \nu_{k+1}(N_{l(j),0})$ such that for any $x \in I_j$ there exists an $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}]$, for which

$$\frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k} X_{h,j}(x). \tag{22}$$

Moreover, ϕ_j is constant on the intervals of the form $[i \cdot 2^{-J}, (i+1)2^{-J})$, $\int_{I_j} \phi_j = k \cdot 2^{-M+1}\mu(I_j)$, $\phi_j(x) = 0 = X_{h,j}(x)$, for $x \notin I_j$, $h = 1, ..., k$. We may also require that if

$$x \notin \bigcup_{l=0}^{k-1} \bigcup_{i \in \mathbb{Z}} [(i \cdot 2^D + l)2^{-J}, (i \cdot 2^D + l + 1) \cdot 2^{-J})$$

then $\phi_j(x) = 0$. This last property implies that the support of ϕ_0 is disjoint from the support of any ϕ_j, $j = 1, ..., 2^{J_{0,0}-R}$. Since ϕ_j is supported on I_j, we see that the supports of the functions ϕ_j are also disjoint.

Set $f = \sum_{j=0}^{2^{J_{0,0}-R}} \phi_j$. Then, using the fact that the supports are disjoint, we have $\mathbf{N}_n(f)(x) = \sum_{j=0}^{2^{J_{0,0}-R}} \mathbf{N}_n(\phi_j)(x)$. We also calculate

$$\int_{I_0} f = \int_{I_0} \phi_0 + \sum_{j=1}^{2^{J_{0,0}-R}} \int_{I_j} \phi_j =$$

$$= 2^{-M+1}\mu(I_0) + k \cdot 2^{-M+1} \sum_{j=1}^{2^{J_{0,0}-R}} \mu(I_j) = (k+1)2^{-M+1}\mu(I_0).$$

For $h = 1, ..., k$, set $X_h = X'_h = \sum_{j=1}^{2^{J_{0,0}-R}} X_{h,j}$. Let $X'_{k+1}(x) = 0.99 \cdot 2^{-l+1}$ if $x \in \Gamma_{l,0}$, otherwise set $X'_{k+1} = 0$. Since X'_{k+1} is constant on the intervals I_j, one can also see that the functions $X'_h(x)$, $h = 1, ..., k+1$ are independent. The functions $X'_h(x)$ are $(M-0.99)$-distributed on I_0 for $h = 1, ..., k$. The function $X'_{k+1}(x)$ is not $(M-0.99)$-distributed, but is $(M-0.99)$-superdistributed. By this we mean that $\mu(\{x : X'_{k+1}(x) = 0.99 \cdot 2^{-l+1}\}) \geq 0.99 \cdot 2^{-M+l-1}\mu(I_0)$, for any $l = 1, ..., M$. But we can and do choose $X_{k+1} \leq X'_{k+1}$ such that X_{k+1} is $(M-0.99)$-distributed on I_0 and the system $X_h(x)$, $h = 1, ..., k+1$ is independent.

If $x \in I_j \subset \Gamma_{l(j),0}$, then

$$\frac{\mathbf{N}_n(\phi_0)(x)}{n} > 0.99 \cdot 2^{-l(j)+1} = X'_{k+1}(x) \geq X_{k+1}(x),$$

when $2^{N_{l(j),0}} \leq n \leq 2^{\nu_{k+1}(N_{l(j),0})}$.

For these same x, by our induction hypothesis, there exists $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}]$ for which

$$\frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k} X_{h,j}(x) = \sum_{h=1}^{k} X_h(x).$$

Therefore, there exists $n \in [2^{N_{l(j),0}}, 2^{\nu_{k+1}(N_{l(j),0})}] \subset [2^{K_S^{(k+1)}}, 2^{\nu_{k+1}(N_{M,0})}]$ for which

$$\frac{\mathbf{N}_n(f)(x)}{n} = \sum_{j=0}^{2^{J_{0,0}-R}} \frac{\mathbf{N}_n(\phi_j)(x)}{n} \geq \sum_{h=1}^{k+1} X_h(x).$$

This also shows that the exit time $K_e^{(k+1)}$ can be chosen to be $\nu_{k+1}(N_{M,0})$.

□

3 The counting problem and Birkhoff's theorem

Theorem 1 also helps to refine connections between Birkhoff's pointwise ergodic theorem and the counting problem. It provides an example of a maximal operator which is of restricted weak type (1,1) but does not satisfy a weak type (1,1) inequality. However, this operator coincides with the one sided Hardy–Littlewood maximal function on characteristic functions of measurable sets. Let us see how and why.

One way to prove Birkhoff's pointwise ergodic theorem is via the maximal inequality

$$\mu\left\{x : \sup_N \frac{1}{N}\sum_{n=1}^{N} |f|(T^n x) > \lambda\right\} \leq \frac{1}{\lambda}\|f\|_1.$$

It turns out (see [9] for instance) that this maximal inequality is equivalent to the weak type (1,1) inequality for the Hardy–Littlewood maximal function on $\mathbb{T}$, the unit circle, that we identify with the interval $[-\frac{1}{2}, \frac{1}{2})$,

$$H(f)(x) = \sup_{t>0} \frac{1}{t}\int_0^t |f(x-y)|dy.$$

The following maximal function was introduced by the first author

$$A(f)(x) = \sup_{\lambda>0} \lambda \cdot m\left\{0 < y < x : \frac{|f(x-y)|}{y} > \lambda\right\}.$$

The interest in the operator A lies in the following results

1. It was used in [12] to give the details of the fact that the return time for the tail in all L^p spaces $1 < p \leq \infty$ is equivalent to the validity of Birkhoff's theorem in all L^r spaces for $1 < r \leq \infty$. In other words, the finiteness of $\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ shown in [2] is equivalent to Birkhoff's theorem in L^p for $1 < p \leq \infty$.

2. If one considers the characteristic function of a measurable set B, then simple computations show that

$$A(\mathbf{1}_B)(x) = H(\mathbf{1}_B)(x). \qquad (23)$$

Thus the operator A satisfies a restricted weak type (1,1) inequality in the sense that we have for all $\lambda > 0$

$$m\{x : A(\mathbf{1}_B)(x) > \lambda\} \leq \frac{1}{\lambda} m(B)$$

i.e. a weak type (1,1) inequality for characteristic functions of measurable sets. (See also [15] or [9] for instance for more on restricted weak type inequalities.)

3. The operator A can be viewed as a continuous analog of the counting function studied in the previous sections. Furthermore, we have the following lemma.

Lemma 7. *Given p, $1 \leq p \leq \infty$ the following statements are equivalent*

(a) There exists a finite constant C such that for all $\lambda > 0$ and $(a_n) \in l^p(\mathbb{Z})$

$$\#\left\{ i \in \mathbb{Z} : \sup_n \left(\frac{\#\{k > 0 : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n} \right) > \lambda \right\} \leq \frac{C}{\lambda^p} \|(a_n)\|_p^p. \tag{24}$$

(b) There exists a finite constant C such that for all $f \in L^p(\mathbb{T})$ and $\lambda > 0$ we have

$$m\{x : A(f)(x) > \lambda\} \leq \frac{C}{\lambda^p} \int |f|^p dm.$$

(c) We can find a finite constant C such that for all $f \in L^p_+(\mu)$ for all measure preserving systems $(X, \mathcal{B}, \mu, T)$

$$\mu \left\{ x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda \right\} \leq \frac{C}{\lambda^p} \int |f|^p d\mu$$

Proof. The proof uses known methods in ergodic theory such as transference or Rohlin's tower lemma. Details of such computations can be seen in [12]. So we only sketch some of them. We remark that (a) is equivalent to the following inequality.

There exists a finite constant C such that for all $\lambda > 0$, $(a_n) \in l^p(\mathbb{Z})$, positive integers K and I,

$$\#\left\{i \in [-I, I] : \sup_{n \leq K} \left(\frac{\#\{k > 0 : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n} \right) > \lambda \right\} \leq \frac{C}{\lambda^p} \|(a_n)\|_p^p. \tag{25}$$

In order to prove that (a) and (b) are equivalent we use step functions of the form $f = \sum_{j=-I}^{I-1} a_j \mathbf{1}_{I_j}$ where $a_j \in \mathbb{R}$ and $a_j = 0$ for $|j| > I$. The interval I_i equals the dyadic interval $[\frac{i}{2^I}, \frac{i+1}{2^I})$.

To show that (a) and (c) are equivalent we use Rohlin's tower lemma where the tower is symmetric and of height $2J + 1$. Rohlin's lemma tells us that for any $\epsilon > 0$ and $J \in \mathbb{N}$ we can find disjoint sets $T^{-i}B$ for $-J \leq i \leq J$, such that the tower $\cup_{i=-J}^{J} T^{-i}(B)$ has total measure greater than $1 - \epsilon$. We take a function $f = \sum_{i=-J}^{J} a_i \mathbf{1}_{T^i B}$ and note that

$$\begin{aligned}\frac{\mathbf{N}_n(f)(x)}{n} &= \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n} \\ &\geq \sum_{i=-J}^{J} \mathbf{1}_{T^i B}(x) \frac{\#\{k \leq J - |i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n}.\end{aligned}$$

Thus, the inequality

$$\mu\left\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda\right\} \leq \frac{C}{\lambda^p} \int |f|^p d\mu$$

implies

$$\begin{aligned}\sum_{i=-J}^{J} \mu\left\{x \in T^i B : \sup_n \left(\frac{\#\{k \leq J - |i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n} \right) > \lambda \right\} \quad (26) \\ \leq \frac{C}{\lambda^p} \mu(B) \sum_{i=-J}^{J} |a_i|^p.\end{aligned}$$

As (26) equals

$$\mu(B) \cdot \#\left\{-J \leq i \leq J : \sup_n \left(\frac{\#\{k \leq J - |i| : \frac{a_{k+i}}{k} > \frac{1}{n}\}}{n} \right) > \lambda \right\}$$

we have

$$
\begin{aligned}
&\#\left\{i \in \mathbb{Z} : \sup_{0<n\le K} \left(\frac{\#\{k>0 : \frac{a_{k+i}}{k} \ge \frac{1}{n}\}}{n} \right) > \lambda \right\} \\
&\le \lim_J \#\left\{-J \le i \le J : \sup_{0<n\le K} \left(\frac{\#\{k \le J-|i| : \frac{a_{k+i}}{k} \ge \frac{1}{n}\}}{n} \right) > \lambda \right\}.
\end{aligned}
$$

□

So Theorem 1 gives us the following contribution to the problem of characterizing operators for which a restricted weak type (1,1) inequality implies a weak type (1,1) inequality. (See [9] for more on this problem.) The operator A does not satisfy a weak type (1,1) inequality. It is shown in [9] that if an operator is generated by convolutions, then a restricted weak type (1,1) inequality implies a weak type (1,1) inequality. Such is the case of the Hilbert transform and the Hardy–Littlewood maximal function. It is shown in [1] that the situation is different in the discrete case.
Next we list some of the properties of the operator A.

Theorem 8. *The operator A defined on $\mathbb{T}$ by the formula*

$$
A(f)(x) = \sup_{\lambda>0} \lambda \cdot m \left\{ 0 < y < x : \frac{|f(x-y)|}{y} > \lambda \right\}
$$

has the following properties

1. *It coincides with the one sided Hardy–Littlewood maximal function when f is the characteristic function of a measurable set on $\mathbb{T}$ hence it satisfies a restricted weak type (1,1) inequality.*

2. *It maps functions in L^p to functions in weak L^p.*

3. *There exists a positive function $f \in L^1(\mathbb{T})$ such that $A(f)(x) \not< \infty$ for a.e. x in $\mathbb{T}$.*

Proof. Statements (1) and (2) follow from Lemma 7.

The last statement is a consequence of Theorem 1. The arguments developped in [2] (cf. Theorem 4) indicate that if we had $A(f)(x) < \infty$ for a.e. x then we would have a weak type (1,1) inequality for A. By Lemma 7 this would imply a weak type (1,1) inequality for $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$, a conclusion that we disproved in Theorem 1. □

Acknowledgment

We thank the referee and Y. Katznelson for their comments that helped improve the presentation of our paper.

References

[1] M. Akcoglu, J. Baxter, A. Bellow, R. Jones, "On restricted weak type $(1, 1)$; the discrete case," *Israel J. Math.*, 124 (2001), 285–297

[2] I. Assani, "Strong Laws for weighted sums of iid random variables," *Duke Math J.*, 88, 2, (1997), 217-246.

[3] I. Assani, "Convergence of the p-Series for stationary sequences," *New York J. Math.,* 3A, (1997), 15-30.

[4] I. Assani, "A weighted pointwise ergodic theorem," *Ann. Inst. Henri Poincare,* 34, 1, (1998), 139-150.

[5] I. Assani, Z. Buczolich and R. D. Mauldin "Counting and convergence in Ergodic Theory," Preprint 2004, submitted.

[6] J. Bourgain, "Return Time sequences of dynamical systems," IHES, Preprint, (1988).

[7] J. Bourgain, "Temps de retour pour des systemes dynamiques," *C.R. Acad. Sci. Paris,* t. 306, Série I, (1988), 483-485.

[8] J. Bourgain, "Pointwise ergodic theorems for arithmetic sets," With an appendix by the author, Harry Fürstenberg, Yitzhak Katznelson and Donald S. Ornstein, *Inst. Hautes Études Sci. Publ. Math.* No. 69 (1989), 5–45.

[9] C. Bennett and R. Sharpley, *Interpolation of operators,* Academic Press, New York, 1988.

[10] P. Halmos, *Lectures on ergodic theory,* Chelsea Publishing Co., New York 1960.

[11] R. Jones, J. Rosenblatt and M. Wierdl, "Counting in Ergodic Theory," *Cand. J. Math.*, 51, (1999), 996-1019.

[12] K. Noonan, *Return Times for the tail and Birkhoff's theorem,* Master's thesis, UNC Chapel Hill. Dec. 2002.

[13] D. Rudolph, "A joining proof of Bourgain's return time theorem," *Erg. Th. and Dyn. Syst.*, 14, (1994), 197-203.

[14] D. Rudolph, "Fully generic sequences and a multiple term return times theorem", *Invent. Math.* , 131, (1998), 199-228.

[15] E. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces,* Princeton University Press, 1971.