Counting and convergence 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 5, 2004

[*]This author received travel support from the Hungarian National Foundation for Scientific Research Grant T032042 to present the results of this paper at the 32nd Winter School in Abstract Analysis, 2004.

[†]Supported in part by NSF grant DMS 0100078

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

Abstract

We give a survey of some L^1 open problems in Ergodic Theory that we hope will also be readable to people who are not necessarily experts in this field. Some emphasis is put on past and recent results on the return times theorem and the recently solved (L^1, L^1) problem. Part of this paper was presented in a series of three talks by the second author at the 32nd Winter School on Abstract Analysis, 2004 January.

1 Irrational Rotations

In his talk at this same Winter School D. Fremlin talked about equidistributed sequences. For a sequence (x_n) the arithmetic average $A_n(f) = \frac{1}{n}\sum_{k=1}^{n} f(x_k)$ was considered and it was investigated whether $A_n(f)$ converges to $\int f$. The most common source of these sequences is coming from Ergodic Theory. One considers $x_n = x + n\alpha$ modulo 1 and it is a consequence of Birkhoff's Ergodic Theorem that for any irrational α for (Lebesgue) almost every x the sequence $A_n(f)$ converges almost everywhere to $\int f$. That is, using the notation $\mathbb{T}$ for the unit circle and considering the transformation $T : \mathbb{T} \to \mathbb{T}$, with $T(x) = x + \alpha$ we have $A_n(f)(x) = \frac{1}{n}\sum_{k=1}^{n} f(T^k x) \to \int_{\mathbb{T}} f$ when $f \in L^1(\mathbb{T})$.

If we do not assume that $f \in L^1(\mathbb{T})$ strange things can happen. From the main result of [15] it follows that if α and β are independent over the rationals, $T(x) = x+\alpha$ and $S(x) = x+\beta$ on $\mathbb{T}$ then there exists a measurable $f : \mathbb{T} \to \mathbb{R}$ for which $\frac{1}{n}\sum_{k=1}^{n} f(T^k x) \to 1$, and $\frac{1}{n}\sum_{k=1}^{n} f(S^k x) \to 0$ for almost every $x \in \mathbb{T}$. Of course, if $f \in L^1$ this cannot happen, since both limits should equal the integral of f.

This leads to the question asked during the problem session of this Winter School by the second author. Assume that f is a measurable real function defined on the unit circle, $\mathbb{T}$. The rotation set of f is

$$R_f = \{\alpha \in \mathbb{T} : \frac{1}{n}\sum_{k=1}^{n} f(x + k\alpha) \text{ converges almost everywhere}\}.$$

By Birkhoff's ergodic theorem if $f \in L^1(\mathbb{T})$ then R_f equals $\mathbb{T}$. It is shown in [14] that if R_f is of positive Lebesgue measure then $f \in L^1(\mathbb{T})$ and hence $R_f =$

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

$\mathbb{T}$, on the other hand given a sequence of rationally independent irrationals one can find $f \notin L^1(\mathbb{T})$ such that R_f contains this sequence. Hence R_f can be dense for a non-L^1 function. In [29] it was shown that there exists non-integrable f for which R_f is c-dense. This means that R_f is of cardinality continuum in every non-empty open interval.

Nothing is known about the possible Hausdorff dimension of R_f for a non-L^1 function. If one can show that $dim_H(R_f) > 0$ for a measurable $f \notin L^1(\mathbb{T})$ then the second author offers a prize of one bottle of Hungarian wine (Egri bikavér). For an example with $dim_H(R_f) = 1$ two bottles are offered. For a "complete characterization" of R_f three bottles are offered. (In case at some of the Czech conferences Hungarian wine needs to be converted into Czech beer there is an exchange rate of 1 bottle of wine = 3 beers.)

The above results show that in Ergodic Theory the big difference is between the class of L^1 functions and non-L^1 ones.

In the sequel we will see that sometimes even the L^1 functions are bad.

2 Classical Results everyone should know

The first important tool is the Banach-principle, see [17] and Chapter V of [23].

We assume that $(\Omega, \mathcal{B}, \mu)$ is a probability space, and μ is a non atomic measure. We will use the notation $L(\Omega)(\neq L^1(\Omega))$ for the linear space of equivalence classes of measurable real functions on Ω, this class should not be confused with $L^1(\Omega)$.

Assume that $(\mathcal{S}, \rho)$ is a metric space.

Definition 1. We say that the operator $A : \mathcal{S} \to L(X)$ is *continuous in measure* if whenever the sequence $\gamma_i \in \mathcal{S}$ satisfies $\lim_{i\to\infty} \rho(\gamma_i, \gamma) = 0$ for a $\gamma \in \mathcal{S}$ then for all $\epsilon > 0$ we have $\lim_{i\to\infty} \mu\{x : |A\gamma_i(x) - A\gamma(x)| \geq \epsilon\} = 0$.

Given $A_n : \mathcal{S} \to L(X)$ we denote the corresponding maximal operator by $A^*\gamma(x) = \sup\{|A_n\gamma(x)| : n \in \mathbb{N}\}$.

Theorem 1. (Banach Principle) *Suppose $\mathcal{S}$ is a Banach space in the norm $||.||$, and $A_n : \mathcal{S} \to L(X)$ is a sequence of linear maps which are continuous in measure. Assume that for all $f \in \mathcal{S}$,*

$$A^*f(x) = \sup_n |A_n f(x)| \text{ is finite } \mu\text{-almost everywhere.} \tag{1}$$

Then the set of functions $f \in \mathcal{S}$ such that $\lim_{n\to\infty} A_n f(x)$ exists μ-almost everywhere forms a norm-closed subspace of $\mathcal{S}$.

There are many applications of this principle outside Ergodic Theory but to understand the way it is used in Ergodic Theory one should think for example of $\mathcal{S} = L^p(\Omega)$ and assume that $T : \Omega \to \Omega$ is a measure preserving transformation. Then consider the operator $A_n f = \frac{1}{N}\sum_1^N f(T^n x)$. It is not difficult to verify that A_n is continuous in measure. To verify almost everywhere convergence is usually a difficult task. By using the Banach principle this task can be split into two parts. First one shows that a maximal inequality, (1) holds and then it is enough to verify convergence of A_n for a dense set in L^p, say for continuous, or for bounded functions.

Therefore, we have sufficient motivation for studying maximal inequalities. Next we recall the Stein-Sawyer result about maximal inequalities of weak type see [26], [27].

We work again in a probability space $(\Omega, \mathcal{B}, \mu)$ and our sequence of operators is $A_n : L^p(\Omega) \to L(\Omega)$. We again assume that each A_n is continuous in measure which in this special case means that from $||f_i - f||_p \to 0$ it follows that $A_n f_i(x) \to A_n f(x)$ in measure. We use the notation for the maximal operator $A^* f(x) = A^*(x, f) = \sup_n |A_n f(x)|$ for $f \in L^p(\Omega)$.

Assume that T is a measure preserving transformation of Ω. The sequence A_n *commutes* with T as for each n we have $A_n(f \circ T)(x) = A_n(f, Tx)$, μ almost everywhere. This means in particular that $A^*(Tx, f) \le A^*(x, f(Tx))$ μ almost everywhere for all $f \in L^p(\Omega)$.

Assume $\mathcal{T}$ is a collection of measure preserving transformations on Ω. The set $X \subset \Omega$ is *fixed by* $\mathcal{T}$ if for all $T \in \mathcal{T}$ we have $X = T^{-1}(X)$. We say that $\mathcal{T}$ is an *ergodic family* on Ω when from $X \subset \Omega$ is fixed by $\mathcal{T}$ it follows that $\mu(X) = 0$, or $\mu(X) = 1$. The sequence of operators (A_n) is *distributive* if each A_n commutes with every member of some ergodic family on Ω.

To think of an example one can take a sequence of operators commuting with one fixed ergodic transformation, T on Ω. To be more specific if T is ergodic then $A_n(f) = \frac{1}{n}\sum_{k=1}^n f(T^k x)$ is a distributive sequence.

Now the first important result in [26] is the following.

Theorem 2. *Let (A_n) be a distributive sequence of linear operators on $L^p(\Omega)$, where each A_n is continuous in measure and maps $L^p(\Omega)$ to $L(\Omega)$. If $1 \le p \le 2$, and $A^* f(x) = \sup_n |A_n f(x)|$ is finite μ almost everywhere for all*

$f \in L^p(\Omega)$ then there exists a constant C such that

$$\mu\{x : A^*f(x) \geq \lambda\} \leq \frac{C}{\lambda^p} \int_\Omega |f(x)|^p d\mu(x) \tag{2}$$

for all $f \in L^p(\Omega)$ and $\lambda > 0$.

If (2) holds then A^* is of weak type (p, p).

Corollary 3. *Let (A_n) be a distributive sequence of continuous linear transformations on $L^p(\Omega)$, where $1 \leq p \leq 2$. Then, either A^* is of weak type (p, p), or $A^*f(x) = \infty$, μ almost everywhere for all $f \in L^p(\Omega)$ with the possible exception of a set which is of the first Baire category in $L^p(\Omega)$.*

Corollary 4. *Let A_n be as above with the additional assumption that $\lim_{n\to\infty} A_nf(x)$ exists μ almost everywhere for all f in a dense subset of $L^p(\Omega)$. Then A^* is of weak type (p, p) if and only if $\lim_{n\to\infty} A_nf(x)$ exists μ almost everywhere* **for all** *$f \in L^p(\Omega)$.*

This corollary can be used the following way: we know convergence for a nice set of functions, like bounded, or continuous functions plus we know that the maximal operator A^* is of weak type (p, p). Then we have convergence almost everywhere for all functions.

In the earlier theorem and corollary the assumption $1 \leq p \leq 2$ is needed by a counterexample of Stein. However, if the sequence A_n is *nonnegative*, that is, $f(x) \geq 0$ implies that $A_nf(x) \geq 0$ almost everywhere then the following theorem holds.

Theorem 5. *Let (A_n) be a distributive sequence of continuous in measure nonnegative linear transformations of $L^p(\Omega)$ into $L(\Omega)$, and $1 \leq p \leq \infty$. Assume that $A^*f(x)$ is finite μ almost everywhere for all $f \in L^p(\Omega)$. Then for $p < \infty$, A^* is of weak type (p, p) and there exists C such that $||A^*f||_\infty \leq C \cdot ||f||_\infty$ holds for all $f \in L^\infty(\Omega)$.*

Ergodic averages, our standard example A_n, are nonnegative so Theorem 5 is applicable for them.

The third important classical tool is the Conze principle [16]. A rough statement of this says that if a maximal inequality holds for one ergodic dynamical system then it does for every dynamical system.

For a more precise statement of this important result let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system and ν a probability measure on the integers, $\mathbb{Z}$. Given a measurable function f, and an invertible measure preserving transformation T we put $\nu^T f(x) = \sum_{k=-\infty}^{\infty} \nu(k) f(T^k x)$.

In case we consider the usual ergodic averages we can consider the sequence of measures for which $\nu_N(k) = \frac{1}{N}$ if $1 \leq k \leq N$, $\nu_N(k) = 0$ otherwise. Then $\nu_N^T(x) = \frac{1}{N} \sum_{k=1}^{N} f(T^k x)$.

Theorem 6. *Let (ν_N) be a sequence of probability measures on $\mathbb{Z}$. Suppose that for some ergodic T there is a universal constant C such that*

$$\mu\{x : \sup_{N\geq 1} |\nu_N^T f(x)| > \lambda\} \leq \frac{C}{\lambda^p} ||f||_p^p \tag{3}$$

for all $f \in L^p(\Omega, \mu)$. Then the same holds for any measure preserving transformation S in place of T.

One of the cornerstones of the pointwise convergence in ergodic theory is given by the following weak type $(1,1)$ inequality

Theorem 7. *Let T be a measure preserving transformation on $(\Omega, \mathcal{B}, \mu)$ and denote by $A^*(f)(x) = \sup_{N\geq 1} \frac{1}{N} \sum_{n=1}^{N} |f(T^n x)|$ the maximal function. Then for all $\lambda > 0$ and all $f \in L^1(\mu)$ we have*

$$\mu\{x : A^*(f)(x) > \lambda\} \leq \frac{C}{\lambda} \int_\Omega |f| d\mu(x). \tag{4}$$

The importance of L^1 for this type of maximal operators is the fact that most of the time they map L^∞ functions to L^∞ functions. This is the case of the maximal function for the ergodic averages in Theorem 7. We clearly have $\|A^*(f)\|_\infty \leq \|f\|_\infty$. The importance of (4) is reflected in the possibility of recapturing all intermediate weak type inequalities for A^* in L^p for $1 \leq p \leq \infty$ by interpolation methods.

3 Bourgain's paper

The starting point of the paper containing the recent results discussed in Section 4 is a paper by Bourgain [12]. Here we discuss the minimum amount of results which are needed later.

We assume in this section that $(\Omega, \mathcal{B}, \mu, T)$ is a dynamical system on a probability space, and T is an automorphism, that is T and its inverse are both measure preserving.

Consider an increasing sequence of integers n_k and the averages $A_N f(x) = \frac{1}{N}\sum_{k=1}^{N} f(T^{n_k}x)$. We will talk then of the convergence of the ergodic averages along the subsequence n_k.

The first nontrivial result that started the study of the averages along subsequences is the paper by A. Brunel and M. Keane [13].

Later Krengel [19] showed the surprising result that there are subsequences of integers that are universally bad for the pointwise ergodic theorem. This means that in any ergodic dynamical system one can find a function f for which the averages along this subsequence do not converge almost everywhere. Actually the bad function is the characteristic function of a set of positive measure.

Then the natural question became the precise knowledge of the universally good subsequences and the bad ones. One of the first nonlinear subsequence that one could naturally think of would be the squares and the averages $A_N f(x) = \frac{1}{N}\sum_{n=1}^{N} f(T^{n^2}x)$, or for a more general case, n^2 is replaced by a polynomial with integer coefficients. Observe that this operator is again nonnegative, continuous in measure and, whenever T is ergodic, it is distributive.

The main question discussed in [12] is the almost everywhere convergence of $A_N f$. Almost everywhere convergence was first established for $f \in L^2$, then later for $f \in L^p$ when $p > 1$. The case $p = 1$ was left open and this is a situation where the change is coming when we have $p = 1$. Originally the second and third author of this paper had an attempt to settle this problem but they did not succeed. Fortunately, after the first author had learned about this attempt suggested to try to use its method on his counting problem and we managed to solve that.

The second and the third listed authors also strongly hope that in 2004 in a forthcoming paper they will be able to show that there are L^1 functions for which we do not have almost everywhere convergence of $A_N f$ when the squares are considered.

One of the key steps in Bourgain's paper is again a maximal inequality.

Theorem 8. *Let $(\Omega, \mathcal{B}, \mu, T)$ be a dynamical system, $p(x)$ a polynomial with integer coefficients and $A_N f(x) = \frac{1}{N}\sum_{n=1}^{N} f(T^{p(n)}x)$. Then there is a maximal inequality*

$$||\sup_N |A_N f|\,||_r \leq C \cdot ||f||_r$$

for $f \in L^r(\Omega, \mu)$, $r > 1$. The constant C depends only on $r > 1$ and the polynomial $p(x)$. Moreover, $A_N f$ converges almost surely as $N \to \infty$.

Methods for $p(n) = n^2$ are not quite identical to the ones used for the conventional ergodic averages. There is an additional difficulty. The maximal inequality helps to prove that the set of convergence is dense in L^r. But finding a dense set where the pointwise convergence holds is not obvious for the averages along the squares. In the case of the conventional ergodic averages the dense set is given by coboundaries i.e. by functions of the form $f - f \circ T$. The cancellation of the averages for such functions reduces the problem of pointwise convergence to the simple problem of the convergence of the sequence $\frac{f(T^n x)}{n}$ to zero. This later convergence is obtained in a simple way by using the Borel Cantelli lemma. Note that for each $t > 0$ we have $\sum_{n=1}^{\infty} \mu\{x : \frac{|f(T^n x)|}{n} > t\} = \sum_{n=1}^{\infty} \mu\{x : |f(x)| > tn\} < \infty$, as $f \in L^1(\mu)$.

The convergence along the squares is obtained by establishing a stronger maximal inequality, sometimes called variational inequality, that shows basically that the sequence $A_n(f)$ is Cauchy pointwise. Observe that if one can find a universal constant C such that for any increasing sequence n_k going to infinity and each positive integer N we have

$$\sum_{k=1}^{N} \| \sup_{n_k \le n \le n_{k+1}} |A_n(f) - A_{n_k}(f)| \|_2^2 \le C \cdot \log(N+1) \|f\|_2^2,$$

then we must have the almost everywhere convergence of the sequence $A_n(f)$. If not, we could find a positive integer δ and a sequence n_k such that

$$\| \sup_{n_k \le n \le n_{k+1}} |A_n(f) - A_{n_k}(f)| \|_2^2 \ge \delta.$$

Thus the previous inequality would give us

$$N\delta = \sum_{k=1}^{N} \delta \le \sum_{k=1}^{N} \| \sup_{n_k \le n \le n_{k+1}} |A_n(f) - A_{n_k}(f)| \|_2^2 \le C \log(N+1) \|f\|_2^2.$$

From this one can derive easily a contradiction. This is essentially the path followed in [12].

There are several variants of the above theorem. One interesting is when Λ denotes the sequence of prime numbers, Λ_N denotes the prime numbers less or equal than N, and $A_N f(x) = \frac{1}{\#\Lambda_N} \sum_{n \in \Lambda_N} f(T^n x)$. The main result says again that $A_N f$ converges almost surely for $f \in L^r$, $r > 1$.

Bourgain's paper contains a famous appendix by Bourgain, Fürstenberg, Katznelson and Ornstein about Return Times of Dynamical Systems.

We assume that $(X, \mathcal{B}, \mu, T)$ is and ergodic system, A is measurable and of positive measure. For an $x \in X$, the return time sequence to A is given by $\Lambda_x = \{n \in \mathbb{Z}_+ : T^n x \in A\}$.

Poincaré's recurrence theorem implies that almost every x returns to A infinitely often, so Λ_x is indeed an infinite sequence for almost every x. Birkhoff's ergodic theorem implies that Λ_x has positive density for μ almost every $x \in X$.

The first theorem in the Appendix of [12] is the following.

Theorem 9. *Λ_x satisfies for μ almost every x the pointwise ergodic theorem, that is,*

$$A_N g = \frac{1}{N} \sum_{1 \le n \le N,\, n \in \Lambda_x} S^n g \tag{5}$$

converges almost surely for any measure preserving system $(Y, \mathcal{B}, \nu, S)$ and $g \in L^1(Y)$.

In this theorem the operator A_N is not "normed properly" but this is no problem since for almost every x the sequence Λ_x is of positive density.

Of course, one can rephrase (5) as

$$A_N g(y) = \frac{1}{N} \sum_{1 \le n \le N} \chi_A(T^n x) g(S^n y).$$

Now it is quite natural to generalize it by replacing the characteristic function by a function coming from a more general class.

Theorem 10. *Given any measure preserving transformation of a probability space $(X, \mathcal{B}, \mu, T)$ and $f \in L^\infty$, there exists $X_0 \subset X$ of full measure such that for all $x_0 \in X_0$ and second dynamical system $(Y, \mathcal{C}, \nu, S)$ and $g \in L^1(\nu)$ there exists Y_0 of full measure, depending on the two systems and x_0, such that for all $y_0 \in Y_0$*

$$B_N g(y_0) = \frac{1}{N} \sum_{1 \le n \le N} f(T^n x_0) g(S^n y_0)$$

converges.

This can be generalized further to a version where $1 \le p \le \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, $f \in L^p(X, \mu)$ and $g \in L^q(Y, \nu)$.

It is important to observe that X_0 is universal for all dynamical systems! Otherwise, fixing f and g, by using Birkhoff's theorem for the product system the projection of the convergence set onto the first factor obtained by Birkhoff's theorem depends on both functions. Hence the Return Time theorem is not saying anything about the case of $f \in L^1$ and $g \in L^1$ while Birkhoff's theorem, on the other hand, guarantees convergence for $f \otimes g \in L^1 \times L^1$, $\mu \otimes \nu$-a.e.

This leads to the (L^1, L^1) problem of I. Assani see [1], [2], later also mentioned by other authors in [18], [25]. This question is about the validity of Theorem 10 when $f \in L^1(\mu)$, $g \in L^1(\nu)$? The main result of our joint paper [7] yields a negative answer to this question. One can point out that if one considers instead the norm convergence in L^1, I. Assani [3] has shown that given a function $f \in L^1(\mu)$ there exists a set X_f of full measure such that for each $x \in X_f$ and all measure preserving system $(Y, \mathcal{G}, \nu, S)$ and all $g \in L^1(\mu)$ the averages $\frac{1}{N}\sum_{n=1}^{N} f(T^n x) g \circ S^n$ converge in $L^1(\nu)$ norm. Thus we had the universality for pairs of L^1 functions with respect to the norm convergence. The (L^1, L^1) problem was the universality for the pointwise convergence for pairs of (L^1, L^1) functions.

Before discussing some of the key steps in our solution to this (L^1, L^1) problem we briefly talk about past and current work on the return times. This list is not exhaustive. A more exhaustive one can be found in [5]. The results on the return times theorem before Bourgain's paper [12] were all respecting the duality. The result of [13] by A. Brunel and M. Keane mentioned above was the first return times type theorem. Michael Keane at the recent ergodic workshop held at Chapel Hill in February 2004, indicated to us that the main idea was already in Antoine Brunel's thesis. More can be found also in U. Krengel's book [19]. After Krengel's book the next published reference we know of appeared in [9] by A. Bellow and V. Losert. There they did show that with an additional spectral assumption for T (T has countable Lebesgue spectrum) the Return Times Theorem holds. The class of transformations with this spectral property is known to be strictly bigger than the class of K automorphims (Kolmogorov automorphisms). On page 335 of their paper they wrote that the Return Times Theorem had been obtained for K automorphims by H. Fürstenberg, M. Keane, J.P. Thouvenot and B. Weiss. An extension to amenable groups was given by D. Ornstein and B. Weiss in [22]. They extended the Return Times Theorem to the setting of group actions. These authors assume that G (a countable group) is the union of a Fϕlner sequence of sets (A_n) (i.e., $|gA_n \triangle A_n|/|A_n| \to 0$ for all $g \in G$) for which $|A_n^{-1}A_n|/|A_n|$ is bounded. Then for any ergodic action of G on $(X, \mathcal{B}, \mu)$ and any $b \in L^\infty(\mu)$ for almost every $x \in X$ the sequence $b(gx)$ gives good weights for pointwise convergence almost everywhere (with respect to averages over the A_n).

In 1991 I. Assani asked if a multiple version of Bourgain's theorem could be obtained. At the same time he raised the (L^1, L^1) problem.

A partial answer was given in [6] by Assani, Lesigne and Rudolph. A multiple L^1 return times theorem for i.i.d. random variables was established in [1]. A multiple term return times theorem for L^∞ stationary processes was later proved by D. Rudolph [25].

The first result breaking the duality was obtained by I. Assani in [1] and then in [2].

Theorem 11. *Let X_n be a sequence of i.i.d. symmetric random variables on the probability space $(X, \mathcal{G}, P)$ with finite p moment for some $1 < p \leq \infty$ then there exists a set of full measure $\overline{X}$ such that for each $\omega \in \overline{X}$ for all $1 < r \leq \infty$ for all measure preserving systems $(Y, \mathcal{G}, \nu, S)$ and $g \in L^r(\nu)$ the averages*

$$\frac{1}{N} \sum_{n=1}^{N} X_n(\omega) g(S^n y) \tag{6}$$

converge ν almost everywhere.

The symmetric assumption was later removed in [4]. Using in part the method in [3] related results have been obtained for weighted averages (i.e averages of the form $\frac{1}{N}\sum_{n=1}^{N} a_n X_n$ where $\sup_N \frac{1}{N}\sum_{n=1}^{N} |a_n|^q < \infty$ for some $1 < q \leq \infty$) by J. Baxter, R. Jones, M. Lin and J. Olsen (see [8]). The case corresponding to $a_n = f(T^n x)$ for some functions $f \in L^q(\mu)$ for some $1 < q \leq \infty$ was obtained in [3] (Theorem 2).

4 An L^1 Counting Problem in Ergodic Theory

In this section the new results of [7] are discussed. The difficult proof of Theorems 12 and 13 below are in this paper. Here to illustrate how these theorems work we present the details of how to deduce Theorems 14 and 15, which are mentioned without proof as corollaries in [7].

Again $(X, \mathcal{B}, \mu)$ denotes a probability measure space, T is an invertible measure preserving transformation and $f \in L^1_+(\mu)$, that is, f is non-negative and belongs to L^1. A consequence of Birkhoff's Ergodic Theorem that $\frac{f(T^n x)}{n} \to 0$ μ-almost everywhere, hence the following function, $\mathbf{N}_n(f)(x) = \#\left\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\right\}$ is finite almost everywhere.

We consider the following

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

In Assani [1] and [2] the maximal operator $\sup_n \frac{\mathbf{N}_n(f)(x)}{n}$ is used to study the pointwise convergence of $\frac{\mathbf{N}_n(f)(x)}{n}$.

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 almost everywhere to $\int f d\mu$.

For non-ergodic T the limit is the conditional expectation of the function f with respect to the σ field of the invariant sets for T.

Hence, the limit is the same as the limit of the ergodic averages $A_N(f)(x) = \frac{1}{N}\sum_{n=1}^N f(T^n x)$.

By Birkhoff's Ergodic Theorem, $A_N(f)(x)$ converges almost everywhere for any function $f \in L^1(\mu)$.

It is natural to ask whether $\frac{\mathbf{N}_n(f)(x)}{n}$ also converges almost everywhere when $f \in L^1(\mu)$.

On the other hand, for independent identically distributed random variables $X_n \in L^1$ it was shown by Assani in [1] that

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

converges almost everywhere to the common expected value, $E(X_1)$.

Later in [18] by R. Jones, J. Rosenblatt and M. Wierdl the counting problem was discussed further.

To understand the way this counting function works next we discuss a simple example of a non-ergodic transformation and a bounded f for which we have pointwise convergence of $\frac{\mathbf{N}_n(f)(x)}{n}$ and we can see some heuristic why can one expect a maximal inequality.

Let $X = \mathbb{T} = [0,1)$ and $\mu =$Lebesgue measure. Assume a natural number $M > 4$ is fixed and $T(x) = x + \frac{1}{M} \mod 1$. Put $f(x) = 1$ if $x \in [0, \frac{1}{M})$, and $f(x) = 0$ otherwise.

Then $f(T^k x)/k > 1/n$ if $T^k x \in [0, 1/M)$ and $k < n$, therefore

$$\left[\frac{n}{M}\right] - 1 \leq \mathbf{N}_n(f)(x) \leq \left[\frac{n}{M}\right] + 1$$

and

$$\frac{\frac{n}{M} - 2}{n} \leq \frac{\mathbf{N}_n(f)(x)}{n} \leq \frac{\frac{n}{M} + 1}{n}$$

hence, $\frac{\mathbf{N}_n(f)(x)}{n} \to \frac{1}{M} = \int f$.

Assume $2 < m < M$, $\lambda = 1/m$. If $x \in [1 - \frac{j}{M}, 1 - \frac{j-1}{M})$ with a $j \in \{1, ..., M-1\}$ and $j + lM \leq n - 1 < j + (l+1)M$ with $l \in \{0, 1, ...\}$ then $\mathbf{N}_n(f)(x) = 1 + l$ and hence

$$\frac{\mathbf{N}_n(f)(x)}{n} \leq \frac{1 + l}{(j+1) + lM} \leq \frac{1}{j+1}.$$

Therefore $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} = \frac{1}{j+1}$. From $\frac{1}{j+1} > \frac{1}{m}$ it follows that $j \leq m - 2$.

$$\mu(\{x : \sup_n \frac{\mathbf{N}_n(f)(x)}{n} > \lambda = \frac{1}{m}\}) = \mu([1 - \frac{m-2}{M}, 1) =$$

$$= \frac{m-2}{M} < \frac{\frac{1}{M}}{\frac{1}{m}} = \frac{\int_X f}{\lambda}.$$

Which shows that in this specific case we have a weak-(1,1) maximal inequality as well. In fact, this observation can be extended to any characteristic function of a measure preserving system. Indeed, we have

$$\frac{\#\{k : \frac{\mathbf{1}_A(T^k x)}{k} > \frac{1}{n}\}}{n} = \frac{1}{n} \sum_{k=1}^{n-1} \mathbf{1}_A(T^k x).$$

Thus the weak type (1,1) maximal inequality for the ergodic averages gives the following inequality

$$\mu\{x : \sup_n \frac{\#\{k : \frac{\mathbf{1}_A(T^k x)}{k} > \frac{1}{n}\}}{n} > \lambda\} \leq \frac{1}{\lambda}\mu(A).$$

So we have the weak type (1,1) inequality for characteristic functions of measurable sets. Such a property is called restricted weak type inequality and was introduced in [28]. The L^1 problem for the counting was the search of an extension of this restricted weak type (1,1) inequality to a weak type (1, 1) inequality or the non existence of such extension.

By using a generalized version of the Stein-Sawyer result from Assani [1] one can state the following equivalent problem to the counting problem.

Counting Problem II.

Does there exist a finite positive constant C such that for all measure preserving systems and all $\lambda > 0$

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

The main result of [7] is to show that this equivalent problem has a negative answer:

Theorem 12.

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

With a little extra work this can be rephrased in a stronger form

Theorem 13. *For any nonatomic ergodic system one can find an integrable f for which $\sup_n \frac{\mathbf{N}_n(f)(x)}{n} = \infty$ almost everywhere.*

The first consequence of Theorems 12 and 13 is linked to the study of the maximal function $\mathbf{N}^*(f)(x) = \sup_n \frac{\mathbf{N}_n(f)(x)}{n}$. This result is called the Return Times for the Tail (of the Cesaro averages).

Definition 2. Let $(X, \mathcal{B}, \mu, T)$ be a measure preserving dynamical system. The Return Times for the Tail Property holds in $L^r(\mu)$, for $1 \leq r \leq \infty$ 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 $\frac{f(T^n x)\cdot g(S^n y)}{n} \to 0$ for ν almost every y.

From Theorem 12 it follows that

Theorem 14. *The Return Times for the Tail Property does not hold for $r = 1$.*

Proof. Theorem 14 follows from Theorem 8 in [1]. It was shown there that for a sequence of nonnegative numbers c_n such that $\lim_n c_n/n = 0$ the following two statements are equivalent

1. $$\sup_n \frac{\#\{k : \frac{c_k}{k} > \frac{1}{n}\}}{n} < \infty;$$

   and

2. for all measure preserving systems $(Y, \mathcal{G}, \nu, S)$ and all $g \in L^1(\nu)$, the sequence $c_n \cdot g(S^n y)/n$ converges to zero ν a.e.

Taking the sequence $c_n = f(T^n x)$ for an ergodic transformation T shows that if the validity of the Return Time for the Tail Property in L^1 were to hold, then we should have for all $f \in L^1_+(\mu)$ for a.e. x,

$$\sup_n \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n} < \infty. \tag{7}$$

Condition (7) for all $f \in L^1_+(\mu)$ is equivalent to saying that

$$\sup_{\alpha > 0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} < \infty$$

for all $f \in L^1_+(\mu)$ for a.e. x. Consider an enumeration of the positive rational numbers r_k and define for each k the function $\mathbf{T}_k(f)(x) = \frac{\mathbf{N}_{r_k}(f)(x)}{r_k}$. We have

$$\sup_{\alpha > 0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} = \sup_k \mathbf{T}_k(f)(x)$$

When T is ergodic it commutes with the family of powers of T. By the ergodic theorem this family is mixing. Indeed, we have

$$\lim_N \frac{1}{N} \sum_{n=1}^{N} \mu(A \cap T^n(B)) = \mu(A)\mu(B)$$

so for each $\rho \geq 1$ there exists an n such that $\mu(A \cap T^n(B)) \leq \rho\mu(A)\mu(B)$. For each $\gamma \geq 1$ we have

$$\sup_k \mathbf{T}_k(\gamma f)(x) = \gamma \sup_k \mathbf{T}_k(f)(x).$$

Thus we can apply Theorem 4 of [1] to conclude that there exists a finite positive constant C such that for all $f \in L^1_+$,

$$\mu\{x : \sup_k \mathbf{T}_k(f)(x) > 1\} \leq C \int f d\mu.$$

This means that

$$\mu\left\{x : \sup_{\alpha>0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha} > 1\right\} \leq C \int f d\mu.$$

Replacing the function f by f/λ provides a maximal inequality for the maximal function

$$\sup_{\alpha>0} \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{\alpha}\}}{\alpha}.$$

From this we obtain easily a maximal inequality with the same constant C for

$$\sup_n \frac{\#\{k : \frac{f(T^k x)}{k} > \frac{1}{n}\}}{n}.$$

Having this constant for one ergodic transformation provides the same constant for all ergodic transformations. The ergodic decomposition would then show that

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

This would contradict Theorem 12. □

Again if we have stronger assumptions than L^1 about f then Assani in [1] and [2] showed that the Return Times for the Tail Property holds in L^r for $1 < r \leq \infty$ and even in $L \log L$.

As it was mentioned in the end of the previous section from Theorem 14 it follows that

Theorem 15. *Bourgain's Return Time Theorem does not hold for pairs of* (L^1, L^1) *functions.*

Proof. We can argue also by contradiction. If we had the validity of the Return Times for Pairs property for (L^1, L^1) spaces then we would have the convergence in the universal sense of the averages

$$\frac{1}{N}\sum_{n=1}^{N} f(T^n x)\cdot g(S^n y) = \frac{\sigma_N}{N}$$

for $g \in L^1(\nu)$. This would imply the convergence to zero of

$$\frac{\sigma_N}{N} - \frac{\sigma_{N-1}}{N-1} = \frac{f(T^N x) g(S^N y)}{N} + \frac{\sigma_{N-1}}{N-1}\cdot\frac{N-1}{N} - \frac{\sigma_{N-1}}{N-1}.$$

This in turn would give the validity of the Return Time for the Tail property in L^1, but this was disproved in Theorem 14. □

5 Some consequences

We indicate here two applications of Theorem 14. The second is mentioned in [7].

5.1 Random ergodic theorems with universally representative sequences

In [20] some random ergodic theorems were studied. Assume $(\Omega, \mathcal{B}, P)$ is a probability measure space, $\frac{1}{2} \leq \sigma \leq 1$ and $Y_1, Y_2, ...$ is a sequence of i.i.d. random variables of values in $\{-1, 1\}$ with $P(Y_n = 1) = \sigma$ and $P(Y_n = -1) = 1 - \sigma$. Set $a_n(\omega) = \sum_{k=1}^{n} Y_k(\omega)$. By the strong law of large numbers we know that $\lim_n \frac{a_n(\omega)}{n} = E(Y_1) = 1$. In particular if $\sigma > \frac{1}{2}$ then for μ almost every ω we have $\lim_n a_n(\omega) = \infty$. Fix such an ω and consider for $f \in L^p(\mu)$, $1 \leq p \leq \infty$ the averages

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{a_n(\omega)} x).$$

It is proved in [20] that if $p > 1$ and $\sigma > \frac{1}{2}$ then the averages

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{a_n(\omega)}x),$$

converge almost everywhere. In the same paper a question was raised about the case $p = 1$. Based on the reduction made in [20], Theorem 13 immediately shows the following

Theorem 16. *Consider a sequence of i.i.d. random variables* $Y_1, Y_2, \ldots$ *defined on the probability measure space* $(\Omega, \mathcal{B}, P)$ *of values in* $\{-1, 1\}$*. Assume that* $P(Y_n = 1) = \sigma$ *and* $P(Y_n = -1) = 1 - \sigma$ *with* $\sigma > \frac{1}{2}$*. Set* $a_n(\omega) = \sum_{k=1}^{n} Y_k(\omega)$*, and fix* ω *in the set where* $\lim_n a_n(\omega) = \infty$*. In any aperiodic dynamical system there exists a function* $f \in L^1(\mu)$ *such that the averages*

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{a_n(\omega)}x),$$

do not converge almost everywhere.

5.2 Continuous analog of the counting function

Most graduate textbooks in harmonic analysis give a proof of the weak type $(1, 1)$ inequality for the Hardy–Littlewood maximal function

$$H(f)(x) = \sup_t \int_0^t f(x - u)du.$$

It is also shown that this weak type $(1, 1)$ inequality is equivalent to the weak type $(1, 1)$ inequality for the maximal operator for the ergodic averages. It seems then a natural question to ask if the counting function translates also for some maximal operator on $\mathbb{T}$ and what happens in this case. The continuous analog of the counting function is given by

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

and was introduced by the first author. Here again by considering a simple characteristic function of a measurable set B one can see that A satisfies a restricted weak type $(1, 1)$ inequality. However, another consequence of Theorem 14 is the following result;

Theorem 17. *There exists a function $f \in L^1(\mathbb{T})$ such that $A(f)(x)$ is not finite almost everywhere.*

6 Some Open Problems

We present here some open problems

Problem A *What is the correct Orlicz class of functions for which the averages $\frac{\mathbf{N}_n(f)(x)}{n}$ converge a.e.?*

Problem B *What are the precise values for p and q for which the return times theorem holds? Does it hold only when the duality is respected?*

References

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

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

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

[4] I. Assani, Wiener Wintner dynamical systems, *Erg. Th. and Dyn. Syst.*, 23, 1637-1654, 2003.

[5] I. Assani, *Wiener Wintner Ergodic theorems*: 228 pages, World Scientific Pub Co, (2003), ISBN: 9810244398.

[6] I. Assani, E. Lesigne and D. Rudolph, Wiener-Wintner return-times ergodic theorem, *Israel J. Math.*, 92, (1995), no. 1-3, 375–395.

[7] I. Assani, Z. Buczolich and D. Mauldin An L^1 Counting Problem in Ergodic Theory, *submitted.*

[8] J. Baxter, R. Jones, M. Lin and J. Olsen, SLLN for weighted independent identically distributed random variables, *to appear in Jour. of Theoretical Probability*, (2004).

[9] A. Bellow and V. Losert, The weighted pointwise ergodic theorem and the individual ergodic theorem along subsequences, *Trans. Amer. Math. Soc.,* 288, 1, (1988), 307-345.

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

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

[12] J. Bourgain, Pointwise Ergodic Theorems for Arithmetic Sets, with an Appendix by J. Bourgain, H. Fürstenberg, Y. Katznelson, and D. S. Ornstein, *Publ. Mat. IHES* **69** (1989), 5-45.

[13] A. Brunel, M. Keane, Ergodic theorems for operator sequences, *Zeitschr. Wahrsc. verw. Gebiete*, **12** (1969), 231-240.

[14] Z. Buczolich, Arithmetic averages of rotations of measurable functions, *Ergod. Th. and Dynam. Sys.* **16** (1996), 1185-1196.

[15] Z. Buczolich, Ergodic averages and free $\mathbb{Z}^2$ actions, *Fund. Math.* **160** (1999), 247-254.

[16] J. P. Conze, Convergence des moyennes ergodiques pour des sous suites, *Bull. Soc. Math. France* **35** (1973), 7-15.

[17] A. Garsia, *Topics in Almost Everywhere Convergence*, Chicago, Markham Publ. Co. 1970.

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

[19] U. Krengel, *Ergodic Theorems*, De Gruyter Studies in Math 6, (1985).

[20] M. Lacey, K. Petersen, D. Rudolph and M. Wierdl Random ergodic theorems with universally representative. *Ann. Inst. H. Poincaré Probab. Statist.*, 30 (1994), no. 3, 353–395.

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

[22] D. Ornstein and B. Weiss, Subsequence ergodic theorems for amenable groups, *Israel J. Math.*, 79, (1992), no. 1, 113–127.

[23] J. Rosenblatt and M. Wierdl Pointwise ergodic theorems via harmonic analysis. *Ergodic theory and its connections with harmonic analysis (Alexandria, 1993)*, 3–151, London Math. Soc. Lecture Note Ser., 205, Cambridge Univ. Press, Cambridge, 1995.

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

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

[26] S. Sawyer, Maximal inequalities of weak type, *Ann. of Math.* **84** (1966) no. 4, 157-174.

[27] E. M. Stein, On limits of sequences of operators, *Ann. of Math.* **74** (1961) no. 4, 140-170.

[28] E. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Mathematical Series, No. 32. Princeton University Press, Princeton, N.J., 1971.

[29] R. E. Svetic, A function with locally uncountable rotation set, *Acta Math. Hungar.* **81** (1998) no. 4, 305-314.