PART I

SURVEY ARTICLES

POINTWISE ERGODIC THEOREMS VIA HARMONIC ANALYSIS

JOSEPH M. ROSENBLATT AND MÁTÉ WIERDL

CONTENTS

INTRODUCTION

Historical remarks

It has been eighty-five years since Bohl [1909], Sierpiński [1910] and Weyl [1910] proved the now famous equidistribution theorem: if α is an irrational number then the sequence $\alpha, 2\alpha, 3\alpha \dots$ is uniformly distributed mod 1. This means that for each subinterval I of the unit interval $[0,1)$ we have

$$\lim_{N\to\infty} \frac{\#\{n \mid n \le N, \langle n\alpha \rangle \in I\}}{N} = |I|, \tag{1}$$

where $\langle x \rangle$ denotes the fractional part of x, that is $\langle x \rangle = x - [x]$, and $|I|$ is the length of the interval I. In fact, Weyl went on to prove, in [Weyl, 1916], that the sequence $\alpha, 2^2\alpha, 3^2\alpha \dots$ is uniformly distributed mod 1. A bit less than twenty years later Vinogradov [cf. Ellison & Ellison, 1985] proved, as a byproduct of his solution of the 'odd' Goldbach conjecture[1], that the sequence $(p_n\alpha)$, where p_n denotes the n-th prime number, is uniformly distributed mod 1. On the other hand, it is easy to see that for some irrational α the sequence $(2^n\alpha)$ is not uniformly distributed mod 1.

Now the question arises what happens if we replace the interval I in (1) by an arbitrary Lebesgue measurable subset of $[0,1)$. What kind of extensions do the results of Weyl and Vinogradov have in this direction? We cannot expect a word-for-word generalization of their results since the Lebesgue measure of any fixed sequence is 0, and I may even be disjoint from the sequence! In the beginning of the 30's Birkhoff [1931] and Khintchin [1933] proved the appropriate generalization of (1): for any fixed Lebesgue measurable $I \subset [0,1)$, for almost every x we have

$$\lim_{N\to\infty} \frac{\#\{n \mid n \le N, \langle x + n\alpha \rangle \in I\}}{N} = |I|, \tag{2}$$

where now $|I|$ denotes the Lebesgue measure of I. (It is not at all clear at first sight, but the result in (2) *does* imply the one in (1).) This is an instance of the individual (or pointwise) ergodic theorem. But then it took more than fifty years to obtain similar generalizations of the other result of Weyl and the result of Vinogradov! Bourgain [1988, 1988a, 1989] developed a very powerful method with which he proved in 1987: for any fixed Lebesgue measurable $I \subset [0,1)$, for almost every x we have

$$\lim_{N\to\infty} \frac{\#\{n \mid n \le N, \langle x + n^2\alpha \rangle \in I\}}{N} = |I|, \tag{3}$$

and

$$\lim_{N\to\infty} \frac{\#\{n \mid n \le N, \langle x + p_n\alpha \rangle \in I\}}{N} = |I| \tag{4}$$

[1] That every large enough odd number is a sum of three prime numbers.

(recall that p_n is the n-th prime). Bourgain's method is a wonderful blend of (analytic) number theory, Fourier analysis and ergodic theory: he uses estimates on the Fourier transforms (or exponential sums)

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i n^2 t}$$

and

$$\frac{1}{N}\sum_{n=1}^{N} e^{2\pi i p_n t}$$

respectively, as they were obtained even by Weyl and Vinogradov.

In this paper we will examine the following general question: for what sequences of integers (a_n) do we have that for any fixed Lebesgue measurable $I \subset [0,1)$, for almost every x

$$\lim_{N\to\infty} \frac{\#\{n \mid n \le N, \langle x + a_n\alpha\rangle \in I\}}{N} = |I|? \tag{5}$$

While our primary goal is to introduce the reader to Bourgain's method that led to the results in (3) and (4), we will also discuss other methods and results of related interest, in particular the "early" results of Krengel, Bellow and others. In fact, the problem formulated in (5) is only the starting point of our investigations. To give an idea of some of the questions we shall examine, let us give the analytic reformulation of the results above.

Let f denote the characteristic function of the interval $I \subset [0,1)$. Then (1) can be rewritten as

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\langle n\alpha\rangle) = \int_0^1 f(x)dx, \tag{1'}$$

and an approximation argument shows that in the above we can take f to be any Riemann integrable function defined on $[0,1)$.

Similarly, if f denotes the indicator function of the Lebesgue measurable set $I \subset [0,1)$, then (2) can be written as

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\langle x + n\alpha\rangle) = \int_0^1 f(x)dx. \tag{2'}$$

Now Khintchin showed that in the above we can take f to be any Lebesgue integrable funcion. How about a similar generalization of (3) and (4)? That is, writing

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\langle x + n^2\alpha\rangle) = \int_0^1 f(x)dx, \tag{3'}$$

and

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(\langle x+p_n\alpha\rangle)=\int_0^1 f(x)dx, \tag{4'}$$

can we take f to be any Lebesgue integrable function? The answer is not known! All we know is that (3') and (4') hold for every L^p for $p>1$. So we arrive at the following question: let (a_n) be a sequence of integers (or real numbers). For what values of p, $1\le p\le\infty$, do we have that for each $f\in L^p$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(\langle x+a_n\alpha\rangle)=\int_0^1 f(x)dx \tag{6}$$

for almost every x? If for some p for every $f\in L^p$ we have (6) a.e., then we call the sequence (a_n) a *universally good averaging sequence* (for L^p), because there is no restriction on the irrational α.

We have already mentioned that for some irrational α the sequence $(2^n\alpha)$ is not uniformly distributed mod 1. But we also have a result of Weyl which says that for almost every α the sequence $(2^n\alpha)$ *is* uniformly distributed mod 1. So at least there is something good to be said here. The picture changes dramatically in the ergodic-theoretical setting. It is a result of Bellow that for *every* irrational α there is a characteristic function f for which

$$\frac{1}{N}\sum_{n=1}^{N} f(\langle x+2^n\alpha\rangle)$$

diverges for almost every x. We can say that in a sense the sequence (2^n) is a *universally bad averaging sequence.*

The methods — mostly developed by Bourgain — to solve the above "subsequence" problems helped to settle other almost everywhere convergence problems. We will mention a number of these problems, but we will develop the method in the special context of these subsequence problems, and usually we refer to other type of applications in the notes after the chapters.

Prerequisites

We do not assume that the reader is familiar with any deeper theories, but this does not mean that she/he has an easy task. The difficulty is that we use tools and results from five branches of mathematics.

Thorough knowledge of the Lebesgue integral is certainly assumed, as well as the elements of functional analysis, such as Baire's category theorem and the uniform boundedness principle. The books [Wheeden, Zygmund, 1977] or [Royden, 1988] contain all the material that is needed from measure theory, and Chapters 7 and 9 of [Royden, 1988] have all the functional analysis we need.

We will use basic facts about harmonic analysis on the classical groups $\mathbb{T}$, $\mathbb{Z}$ and $\mathbb{R}$. We will also use facts about the Hardy-Littlewood maximal function (both on the integers and on the real line), although we will give a proof on $\mathbb{Z}$, since we will use it to prove the ergodic maximal inequality. At certain points — not crucial — we will use the M. Riesz interpolation theorem. Herglotz's theorem on the representation of positive definite sequences is used. All the material that is needed from harmonic analysis can be found in the first two chapters of [Helson, 1991]. Straightforward treatments of the Hardy-Littlewood maximal function can be found in section 4.6 of [Helson, 1991] or in Chapter 9 of [Wheeden, Zygmund, 1977].

From number theory we use the basic properties of congruences. Some of the examples and exercises will refer to deep results of analytic number theory, such as the prime number theorem for arithmetic progressions, but skipping these will not seriously affect the reader's understanding of the other parts of the material. All the material needed from elementary number theory can be found in the first eleven chapters of [Weil, 1979].

The material we assume to be known from probability theory is, in addition to elementary concepts, the moment estimate of Marcinkiewicz and Zygmund. This inequality is a generalization of Khintchin's inequality for the Rademacher functions. The facts we use from probability theory can be found in the first chapter of [Durrett, 1991]. The moment estimate of Marcinkiewicz and Zygmund is in [Garsia, 1970].

Although we will give proofs of both the mean and pointwise ergodic theorems, it is desirable that the reader have some idea of the significance of these results. The reader should know what an aperiodic and an ergodic transformation is, and should know about Rokhlin's tower construction. Strictly speaking, all the material we need from ergodic theory is the first 7 sections (the seventh is "Consequences of ergodicity") and the section "Uniform topology" from [Halmos, 1956]

Summing up: the material is accesible for a third–year (US) graduate student, but she/he may want to skip some of the examples at first reading.

How to use these notes

Our presentation is fairly concise. Although we will try to give careful explanations of the underlying ideas of each proof, we will leave the routine computations to the reader.

There are exercises throughout the text, not just at the ends of the sections. Some of the exercises are difficult, but we provide hints for most of them. The exercises form an integral part of the text, and they often contain interesting developments of the preceding results.

We hope that these notes will motivate the reader to think about some of the unsolved problems of this field (many of which are mentioned in the sequel).

Finally, it is important to note the following. Even with the assumed prerequisites we could not give a full account of Bourgain's method; we give his fundamental inequality without proof, and we refer the interested reader either to Bourgain's original paper [1989] or to Thouvenot's [1990]

Acknowledgements

We want to thank all the mathematicians whose work is discussed in the text; we have freely quoted from the work of Professors M. Akcoglu, A. Bellow, M. Boshernitzan, J. Bourgain, Y. Huang, A. del Junco, R. L. Jones, K. Reinhold-Larsson and H. White. We have benefitted from discussions with Professor M. Lacey.

We also want to thank Professors K. Petersen and I. Salama for inviting us to prepare this article as a more detailed exposition of the survey lecture given at the conference "*Ergodic Theory and Its Connections with Harmonic Analysis*" held in Alexandria, Egypt, in May of 1993.

List of Symbols

We give the page number where the symbol first occurs. We explain the meaning of the symbol or notation only if a brief explanation is possible.

$\mathbb{Z}$: the set of integers	8
$\mathbb{T}$: the torus; $\mathbb{R}/\mathbb{Z}$	8
$\mathbb{R}$: the set of real numbers	8
$\mathbb{C}$: the set of complex numbers	14
$\mathbb{Z}_+$: the set of positive integers	16
$\mathbb{R}_+$: the set of positive real numbers	21
$\mathbb{N}$: the natural numbers	107
$\log x$: natural logarithm of x	27
$[x]$: largest integer not exceeding x	5
$\langle x\rangle$: fractional part of x; $\langle x\rangle = x - [x]$	5
$\#A$: the cardinality of the set A	5
(X, Σ, m, τ): dynamical system	14
$A(t)$: $A(t) = \#\{a \mid a \in A, 1 \le a \le t\}$ for $A \subset \mathbb{Z}_+$	14
$A(t, q, h)$: $= \#\{a \mid a \in A, 1 \le a \le t, a \equiv h \mod q\}$	15
(a_n): $\subset \mathbb{Z}_+$, strictly increasing (unless we say otherwise)	6
τ: a point tranformation	14
$M_t(A, f)$: $= \frac{1}{A(t)} \sum_{\substack{1\le a\le t\\ a\in A}} f(\tau^a x) = \frac{1}{t}\sum_{n\le t} f(\tau^{a_n} x)$	14
T_τ: the operator induced by τ	32
$e(y)$: $= e^{2\pi i y}$	18
$\widehat{M_t}(\beta)$: $= \frac{1}{A(t)} \sum_{\substack{1\le a\le t\\ a\in A}} e(a\beta)$	31
δ_a: the Dirac mass at the point a	15
1_B: indicator function of the set B	54
X_q: $= \{0, 1, \dots, q-1\}$	15
τ_q: $\tau_q x = x + 1 \mod q$	15
m_q: measure on X_q; $m_q(x) = 1/q$ for $x \in X_q$	15
$\|f\|_p = \|f\|_{L^p(X)}$: $= \left(\int_X \|f\|^p\right)^{1/p}$	17
$\|f\|_p = \|f\|_{\ell^p}$: $= \left(\sum_{j\in\mathbb{Z}} \|f(j)\|^p\right)^{1/p}$	49
I_ρ: (first meaning) $I_\rho = \{\rho^k\}_{k\in\mathbb{Z}_+}$ for $\rho > 1$	17
I_ρ: (second meaning)	22
$\widehat{X} = \widehat{X}_q$: $= \{0, 1/q, \dots, (q-1)/q\}$	17
$\widehat{m} = \widehat{m}_q$: measure on $\widehat{X}_q$; $\widehat{m}_q(x) = 1$	17
$\widehat{f} = \mathcal{F}f$: (on a periodic system) Fourier-transform;	17
$\widehat{f} = \mathcal{F}f$: (for $f : \mathbb{Z} \to \mathbb{C}$) $\widehat{f}(\beta) = \sum_{j\mathbb{Z}} f(j) e(-j\beta)$	31
$\check{f} = \mathcal{F}^{-1}f$: (on a periodic system) inverse F-transform;	18
$\check{f} = \mathcal{F}^{-1}f$: (for $f : \mathbb{T} \to \mathbb{Z}$) $\mathcal{F}^{-1}f(j) = \int_{\mathbb{T}} f(\beta) e(j\beta) d\beta$	82
$\overline{d}(A)$: $= \limsup_{t\to\infty} \frac{A(t)}{t}$	21
$\underline{d}(A)$: $= \liminf_{t\to\infty} \frac{A(t)}{t}$	21