数理解析研究所講究録 1404 巻 2004 年 1-11

Some Historical Remarks and Modern Questions Around the Ergodic Theorem

VITALY BERGELSON*
The Ohio State University
Columbus, OH 43210 USA

1 Some historical facts

In 1909-1910 P. Bohl, W. Sierpinski, and H. Weyl independently proved that for any irrational number ξ, the sequence $x_n = n\xi \pmod 1$, $n = 1, 2, \ldots$, is *uniformly distributed* (or, as it was called then, *uniformly dense*) in $[0,1]$, meaning that for any $0 \le a < b \le 1$, one has

$$\lim_{N\to\infty} \frac{\#\{1 \le n \le N : x_n \in (a,b)\}}{N} = b - a. \tag{1.1}$$

It was however the fundamental paper [We], published by Weyl in 1916, that gave rise to the theory of uniform distribution, which today has connections to numerous mathematical disciplines, including number theory, combinatorics, probability theory, harmonic analysis, and ergodic theory.

Weyl starts his paper by noting that a sequence $(x_n)_{n\in\mathbb{N}} \subset [0,1]$ satisfies (1.1) if and only if for any function f which is periodic with period 1 and Riemann integrable on $[0,1]$, one has

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(x_n) = \int_0^1 f\, dx. \tag{1.2}$$

While for Weyl the relation (1.2) expresses an analytic equivalent of the fact that the sequence $(x_n)_{n\in\mathbb{N}}$ is uniformly dense in $[0,1]$, it is the ergodic

*The author acknowledges support received from the National Science Foundation (USA) via grant DMS-0345350

character of (1.2) which we would like to emphasize here. Indeed, the left-hand side of (1.2) can be interpreted as a time average. (Think of $n = 1, 2, \dots$ as instances of time, x_n as the position occupied by a moving particle, and $f(x_n)$ as a result of the measurement of some parameter at time n.) The right-hand side of (1.2) is just the "space average" (which in a more general situation, when the interval $[0,1]$ is replaced by a finite measure space $(X, \mathcal{B}, \mu)$, would be written, for a function $f \in L^1(X, \mathcal{B}, \mu)$, as $\frac{1}{\mu(X)} \int_X f(x)\, d\mu(x)$).

Weyl also observes that since by the theory of Fourier series, any periodic function can be represented as a linear combination of special periodic functions of the form $e^{2\pi imx}$, $m \in \mathbb{Z}$, one has the following convenient criterion for the equidistribution of a sequence $(x_n)_{n\in\mathbb{N}} \subset [0,1]$:

$$\forall m \in \mathbb{Z},\ m \neq 0,\ \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} e^{2\pi imx} = 0. \tag{1.3}$$

Applying this criterion (and simultaneously extending the discussion to higher dimensions), Weyl obtains many by now classical results, among which the following is perhaps the most popular.

Theorem 1.1 *If a real polynomial $p(t) = \alpha_k t^k + \alpha_{k-1} t^{k-1} + \dots + \alpha_0$ has the property that at least one coefficient other than α_0 is irrational, then the sequence $(p(n) \bmod 1)_{n\in\mathbb{N}}$ is uniformly distributed.*

We will return to this result in Section 2 when discussing some modern ramifications of the ergodic theorem.

It took another 15 years for the ergodic idea expressed by relation (1.2) to take on the form of the ergodic theorem.

In 1931, B. Koopman published a short paper ([K]) which amounted to a very simple but significant observation: if T is an invertible measure preserving transformation of a measure space $(X, \mathcal{B}, \mu)$, then the operator U, defined on $L^2(X, \mathcal{B}, \mu)$ by $(Uf)(x) = f(Tx)$, is unitary. The following passage from an article by P. Halmos ([H], p. 91) gives a colorful description of the story of the inception of J. von Neumann's ergodic theorem.

> *Koopman's observation was simultaneously a challenge and a hint. If there is an intimate connection between measure preserving transformations and unitary operators, then the known analytic theory of such*

operators must surely give some information about the geometric behavior of the transformations. By October of 1931, von Neumann had the answer; the answer was the mean ergodic theorem.

Here is the modern formulation of von Neumann's ergodic theorem, obtained in [N1].

Theorem 1.2 *Let U be a unitary operator on a Hilbert space $\mathcal{H}$. Denote by P the orthogonal projection onto the subspace $\mathcal{H}_{inv} = \{f \in \mathcal{H} : Uf = f\}$. For any $f \in \mathcal{H}$, one has*

$$\lim_{N-M\to\infty} \left\| \frac{1}{N-M} \sum_{n=M}^{N-1} U^n f - Pf \right\|_{\mathcal{H}} = 0.$$

Corollary 1.3 *Assume that $(X, \mathcal{B}, \mu)$ is a finite measure space. Let $T : X \to X$ be an invertible measure preserving transformation which is* ergodic, *that is, for any $A \in \mathcal{B}$ with $0 < \mu(A) < \mu(X)$, one has $\mu(A \triangle TA) \neq 0$. Then for any $f \in L^1(X, \mathcal{B}, \mu)$, one has*

$$\lim_{N-M\to\infty} \frac{1}{N-M} \sum_{n=M}^{N-1} f(T^n x) = \frac{1}{\mu(X)} \int_X f(x)\, dx$$

in L^1-norm.

Remark 1.4
1. The proofs of Theorem 1.2 and Corollary 1.3 can be found in any standard text on ergodic theory. See, for example, [P] or [Wa].
2. In [N1], von Neumann deals with a unitary $\mathbb{R}$-action $\left(U_t\right)_{t\in\mathbb{R}}$ induced by a continuous family of measure preserving transformations. While his notation is cumbersome and the proof is complicated by the usage of overly sophisticated machinery (such as the spectral resolution, obtained by M. Stone in [S]), it is a truly outstanding paper.

In October 1931, von Neumann communicated his result to G.D. Birkhoff, who was able to prove, (see [Bi1], [Bi2]) by an original and rather classical argument, the almost everywhere statement which, in modern terms, can be formulated as follows.

Theorem 1.5 *Assume that $(X, \mathcal{B}, \mu)$ is a finite measure space and let $T : X \to X$ be an invertible measure preserving transformation. For any $f \in L^1(X, \mathcal{B}, \mu)$, there exists a function $\overline{f} \in L^1(X, \mathcal{B}, \mu)$ such that*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(T^n x) = \overline{f}(x) \ a.e.$$

If the transformation T is ergodic, then $\overline{f} = \frac{1}{\mu(X)} \int_X f(x)\, dx$ a.e.

Curiously enough, Birkhoff's theorem, published in two articles in PNAS (Proceedings of the National Academy of Sciences), appeared in the December 1931 issue, whereas von Neumann's paper appeared in the January 1932 issue of PNAS. It looks like Birkhoff was in a hurry: [Bi1] was submitted on November 27, 1931, and [Bi2] was submitted on December 1, 1931. Von Neumann's paper [N1] was submitted on December 10, 1931. This controversy is (partly) explained in an account by S. Ulam ([U], p. 98). In the following passage from [U], G.D. stands for Birkhoff and Johnny for von Neumann.

> *Von Neumann never quite forgave G.D. for having "scooped" him in the affair of the ergodic theorem: von Neumann had been first in proving what is now called the weak ergodic theorem. By a sheer virtuoso type of combinatorial thinking, Birkhoff managed to prove a stronger one, and – having more influence with the editors of the Proceedings of the National Academy of Sciences – he published his paper first. This was something Johnny could never forget. He sometimes complained about this to me, but always in a most indirect and oblique way.*

The tension between von Neumann and Birkhoff can also be detected in the way they wrote about the importance and physical adequacy of their results. The reader may find the following three quotations both instructive and amusing.

(i) From [N2], which was submitted on January 21, 1932. In the text below, (1) stands for von Neumann's norm convergence result of the uniform averages $\frac{1}{t-s} \int_s^t f(T^\tau x)\, d\tau$, and (2) stands for Birkhoff's result on the almost everywhere convergence of the averages $\frac{1}{t} \int_0^t f(T^\tau x)\, d\tau$.

It is of interest to decide which of the two formulations, (1) or (2), corresponds to the actual physical problem of the ergodic hypothesis. It turns out that the weaker form of statement (1) is sufficient, – that it, indeed, is the precise mathematical equivalent of the physical state of affairs. It is to be noted, further, that the knowledge of the spectral resolution $E(\lambda)$, which is fundamental in Koopman's method, enables one to dominate the physical situation here completely; in particular, it furnishes a numerical estimation of the degree of convergence of the limiting process connected with the ergodic hypothesis, whereas Birkhoff's existence proof for (2) is of a non-constructive character.

(ii) From [BiK], which was submitted on February 13, 1932. After agreeing that von Neumann's result is "sufficient for the needs of the kinetic theory," the authors still write on the subject of Birkhoff's theorem:

From the viewpoint of the detailed statistics along an individual path-curve, it is fundamentally more far-reaching: in it is proved for the first time that the relative time of sojourn along almost every individual path-curve exists, *a result often assumed implicitly in the writing of physicists, but never proved.*

(iii) In his American Mathematical Monthly article [Bi3], Birkhoff writes:

The integral of Lebesgue (1901), founded upon Borel measure, has been a dominating weapon in the striking advance of Analysis during the present century. Perhaps the Ergodic Theorem (1931) is destined to hold a central position in this development.

As if this were not enough, Birkhoff adds the following in a footnote:

Our discussion here deals only with the "Ergodic Theorem," and not at all with the "Mean Ergodic Theorem" of von Neumann, which stimulated me to reconsider some old ideas, and so led me to the discovery and proof of the Ergodic Theorem, embodying a strong, precise result which, so far as I know, had never been hoped for.

While the speculations offered in [N2] and [BiK] are interesting, they are perhaps not totally convincing, especially when they address measurement and numerical estimation. We shall offer in Section 2 one more take on the question of which of the two ergodic theorems is more useful/relevant in studying physical phenomena.

2 Some questions related to modern developments

Recall that a measure space $(X, \mathcal{B}, \mu)$ is called a Lebesgue space if it is measure-theoretically isomorphic to the unit interval, equipped with the standard Lebesgue measure. The following classical theorem shows that this assumption is in no way too restrictive.

Theorem 2.1 (Cf. [R], Ch. 15, Sec. 5, Thm. 16) *Assume that μ is the completion of a finite Borel measure on a complete separable metric space X. If μ is normalized (i.e. $\mu(X) = 1$) and has no atoms, then $(X, \mathcal{B}, \mu)$ (where $\mathcal{B}$ is the completion of the algebra of Borel sets in X) is measure-theoretically isomorphic to the unit interval with Lebesgue measure.*

Assume now that $(T_v)_{v\in\mathbb{R}}$ is an ergodic continuous measure preserving flow on a Lebesgue space $(X, \mathcal{B}, \mu)$. It is not hard to show that for all but countably many $s \in \mathbb{R}$, the element $T_s = S$ is ergodic. (See for example [CFS], Ch. 12, Sec. 1, Lemma 1, or [Bel], p. 122.) Note also that since the ergodicity of the flow $(T_v)_{v\in\mathbb{R}}$ is obviously equivalent, for any real $c \neq 0$, to the ergodicity of the flow $(T_{cv})_{v\in\mathbb{R}}$, it follows that for all but countably many $s \in \mathbb{R}$, the transformation $T_s = S$ is totally ergodic, meaning that S^n is ergodic for any nonzero $n \in \mathbb{Z}$.

Consider now the following situation. In order to study the continuous averages $\frac{1}{t}\int_0^t f(T_v x)\, dv$, a physicist picks some $s \in \mathbb{R}$ and considers the averages of the form $A_N(f) = \frac{1}{N}\sum_{n=0}^{N-1} f(S^n x)$, where $S = T_s$. (One may think of the averages $A_N(f)$ as corresponding to the average of measurements performed at times $t = 0, 1, \ldots, N-1$.) In reality, the measurements can be done only approximately at times $t = i$, $i \in \{0, 1, \ldots, N-1\}$, and so perhaps it is more natural to consider the "perturbed" averages $A_N^*(f) = \frac{1}{N}\sum_{n=0}^{N-1} f(S^{n+\delta_n} x)$, where $(\delta_n)_{n\in\mathbb{N}}$ is an independent random sequence in a small interval $[-\epsilon, \epsilon]$. Assuming that the flow element $S = T_s$ was chosen with some luck, i.e. that S is ergodic, our physicist would like to know whether he can expect that for large N, the averages $A_N^*(f)$ are close to $\int_X f(x)\, d\mu$. Note that the assumption of ergodicity of S is not unreasonable, since as was noted above, for all but countably many $s \in \mathbb{R}$, T_s is ergodic. Moreover, if the flow $(T_v)_{v\in\mathbb{R}}$ happens to be weakly mixing (which physically means that the system has no angular variables, cf. [KN]), then one can actually show that, for any nonzero $s \in \mathbb{R}$, T_s is ergodic and even weakly mixing.

The following considerations show that the answer to the question whether $A_N^*(f)$ is close to $\int_X f(x)\, d\mu(x)$ is quite satisfactory if one is concerned with norm convergence.

First, note that even if $A_N^*(f)$ does not converge in norm to $\int_X f(x)\, d\mu(x)$, the averages $A_N^*(f)$ will be close in norm to $A_N(f)$ if ϵ is small enough. (Just use the triangle inequality and the fact that, for any f, $\|T_v f - f\| \underset{v\to 0}{\longrightarrow} 0$.) Now, since S is ergodic, $\lim_{N\to\infty} A_N(f) = \int_X f(x)\, d\mu(x)$ in norm, which implies that for all large enough N (and small enough ϵ) the expression $A_N^*(f)$ is close in norm to $\int_X f(x)\, d\mu(x)$. Moreover, this reasoning equally applies to the expressions $A_{N,M}^*(f) = \frac{1}{N-M}\sum_{n=M}^{N-1} S^{n+\delta_n} f$ for large enough $N - M$.

The following theorem says that if one is interested in norm convergence, the expressions $A_N^*(f)$ are well behaved for a *typical* sequence $(\delta_n)_{n\in\mathbb{N}}$.

Theorem 2.2 *Let $I = [-\epsilon, \epsilon]$ and let Π be the countably infinite cartesian power of I, equipped with the normalized product measure m_∞ induced by the Lebesgue measure m on I. Let $S = T_s$ be an ergodic element of an ergodic measure preserving flow $(T_v)_{v\in\mathbb{R}}$ acting on a Lebesgue space $(X, \mathcal{B}, \mu)$. Then for any $f \in L^2(X, \mathcal{B}, \mu)$, the set $\mathcal{D}$ of sequences $(\delta_n)_{n\in\mathbb{N}} \in \Pi$ for which*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} f(S^{n+\delta_n} x) = \int_X f(x)\, d\mu(x)$$

in the L^2 norm satisfies $m_\infty(\mathcal{D}) = 1$.

Theorem 2.2 should be juxtaposed with the following negative result pertaining to almost everywhere convergence.

Theorem 2.3 *Under the assumptions and notational agreements of Theorem 2.2, there exists a set $C \subset \Pi$ with $m_\infty(C) = 1$ such that for any sequence $(\delta_n)_{n\in\mathbb{N}} \in C$, there exists a function $f \in L^\infty(X, \mathcal{B}, \mu)$ such that the averages $A_N^*(f)$ fail to converge almost everywhere.*

One can succinctly summarize the (mathematical) content of Theorems 2.2 and 2.3 as follows: for randomized sampling from an ergodic flow, von Neumann's ergodic theorem is more useful than that of Birkhoff. We leave it to the reader to (try to) interpret these two theorems from the point of view of their physical content.

Here are some comments of the proofs of Theorems 2.2 and 2.3.

The reason for Theorem 2.2 to hold is that one can utilize the spectral theorem in order to reduce the problem to a question on the equidistribution of a random sequence in Π. The result then follows from an application of a form of the law of large numbers.

As for Theorem 2.3, it immediately follows from the following fact proved in [BeBB].

Theorem 2.4 (See [BeBB], Thm 1.1.) *Let $\lambda = (\lambda_n)_{n\in\mathbb{N}}$ be a sequence of real numbers which is linearly independent over $\mathbb{Q}$. Then for any non-periodic flow $\left(X, \mathcal{B}, \mu, (T_v)_{v\in\mathbb{R}}\right)$, there exists a bounded function $f \in L^\infty(X, \mathcal{B}, \mu)$ for which the averages*

$$A_N(f) = \frac{1}{N}\sum_{n=1}^{N} f(T_{\lambda_n}x)$$

fail to converge almost everywhere.

To see that Theorem 2.3 follows from Theorem 2.4, it is enough to observe that there exists a set $E \subset \Pi$ with $m_\infty(E) = 1$ such that for any $(\delta_n)_{n\in\mathbb{N}} \in E$, the sequence $(n + \delta_n)_{n\in\mathbb{N}}$ is linearly independent over $\mathbb{Q}$.

Note that the conditions on the sequence $(\lambda_n)_{n\in\mathbb{N}}$ in Theorem 2.4 are met by some known "deterministic" sequences, such as $(\sqrt{p_n})_{n\in\mathbb{N}}$ where the p_n are distinct primes. One can show that there exists an exceptional countable set $P \subset \mathbb{R}$ such that the sequence $(n^\alpha)_{n\in\mathbb{N}}$ with $\alpha \in \mathbb{R}\setminus P$ consists of linearly independent numbers. One can also show that sequences of the form $(n^r)_{n\in\mathbb{N}}$, where $r \in \mathbb{Q}\backslash\mathbb{Z}$, $r > 0$, also satisfy the conclusion of Theorem 2.4. The reason behind this is that while, for example, numbers of the form $\lambda_n = \sqrt{n}$, $n \in \mathbb{N}$ are linearly dependent over $\mathbb{Q}$, one can extract a subsequence $\lambda_{n_k} = \sqrt{n_k}$ such that (i) $(\lambda_{n_k})_{k\in\mathbb{N}}$ are linearly independent over $\mathbb{Q}$, and (ii) the sequence $(n_k)_{k\in\mathbb{N}}$ has positive density. (See [BeBB], Section 2, for more details.)

On the other hand, the following result, also proved in [BeBB], shows that one has a strong positive result involving the norm convergence.

Theorem 2.5 *Let $\alpha_1, \alpha_2, \ldots, \alpha_k \in (0,1)$, $\alpha_i \neq \alpha_j$ for $i \neq j$. Let $(T_v)_{v\in\mathbb{R}}$ be an ergodic measure preserving continuous flow acting on a Lebesgue space $(X, \mathcal{B}, \mu)$. Then for any $f_1, \ldots, f_k \in L^\infty(X, \mathcal{B}, \mu)$, one has*

$$\lim_{N\to\infty}\left\|\frac{1}{N}\sum_{n=1}^{N} f_1(T_{n^{\alpha_1}}x)\ldots f_k(T_{n^{\alpha_k}}x)-\int_X f_1\,d\mu\ldots\int_X f_k\,d\mu\right\|_{L^2}=0.$$

We will discuss now some natural problems which are suggested by the following beautiful theorem due to J. Bourgain (see [Bo1], [Bo2]).

Theorem 2.6 *Let $(X,\mathcal{B},\mu,T)$ be a measure preserving system and let $p(t)$ be a polynomial with integer coefficients. Then for any $p>1$ and any $f\in L^p(X,\mathcal{B},\mu)$, the averages $\frac{1}{N}\sum_{n=1}^{N} f\big(T^{p(n)}x\big)$ converge almost everywhere.*

One can check that when T is totally ergodic, for any $f\in L^p(X,\mathcal{B},\mu)$, where $p>1$, one has

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f\big(T^{p(n)}x\big)=\int_X f(x)\,d\mu(x)\text{ a.e.}\tag{2.1}$$

Indeed, one needs only to check that, when T is totally ergodic, the averages $\frac{1}{N}\sum_{n=1}^{N} f\big(T^{p(n)}x\big)$ converge to a constant in L^2-norm. (See for example [F], Ch. 3, Sec. 4, or [Be2].)

While from the L^2 result the convergence in any L^p space, where $1\le p<\infty$, follows almost immediately, the question whether (2.1) holds in $L^1(X,\mathcal{B},\mu)$ is open (and is perhaps one of the most interesting problems in the area of almost everywhere convergence).

Here is another interesting set of problems, related to the physical interpretation of Bourgain's theorem. Assume that the transformation T in Bourgain's theorem is totally ergodic. (As we explained above, if one has an ergodic flow $(T_v)_{v\in\mathbb{R}}$, then for all but countably many s, the transformation T_s will also be totally ergodic.) For, say, a bounded and measurable function f, one will have then, for almost every x, the equality of the space average $\int f\,d\mu$ and the time average $\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f\big(T^{p(n)}x\big)$, taken along the polynomial sequence $p(n)$, $n=1,2,\ldots$. What is the physical meaning of this? Why does Nature (in the case of totally ergodic transformations) work so well along the polynomials? It seems, intuitively, that the higher the degree of $p(n)$, the slower the rate of convergence should be. Can one make a mathematical theorem out of this sentiment? Note that the last question is not so simple since it is well known that "there is no speed of convergence" in Birkhoff's ergodic theorem (see, for example, [P], p. 99, Ex. 3). But can one estimate the speed of convergence (as a function of the degree of the polynomial $p(n)$) for smooth enough T and f?

Acknowledgement. It is my pleasure to thank Professors Yoichiro Takahashi and Michiko Yuri for their excellent organization of the RIMS Symposium on Dynamics of Complex Systems, which took place in Kyoto in January of 2004. I would also like to thank Y. Pesin for many useful communications.

References

[Be1] V. Bergelson, Independence properties of continuous flows, *Almost everywhere convergence (Columbus, OH, 1988)*, 121–130, Academic Press, Boston, MA, 1989.

[Be2] V. Bergelson, Weakly mixing PET, *Erg. Th. and Dynam. Sys.* **7** (1987), 337–349.

[BeBB] V. Bergelson, M. Boshernitzan, and J. Bourgain, Some results on nonlinear recurrence, *J. Anal. Math.* **62** (1994), 29–46.

[Bi1] G. Birkhoff, Proof of a recurrence theorem for strongly transitive systems, *Proc. Nat. Acad. Sci. USA* **17** (1931), 650–655.

[Bi2] G. Birkhoff, Proof of the ergodic theorem, *Proc. Nat. Acad. Sci, USA* **17** (1931), 656-660.

[Bi3] G. Birkhoff, What is the ergodic theorem? *Amer. Math. Monthly* **49** (1942), 222–226.

[BiK] G.D. Birkhoff and B.O. Koopman, Recent contributions to the ergodic theory, *Proc. Nat. Acad. Sci. USA* **18** (1932), 279–282.

[Bo1] J. Bourgain, On the maximal ergodic theorem for certain subsets of the integers, *Israel J. Math.* **61** (1988), 39–72.

[Bo2] J. Bourgain, Pointwise ergodic theorems for arithmetic sets (appendix: The return time theorem), *Publ. Math. I.H.E.S.* **69** (1989), 5–45.

[CFS] I. Cornfeld, S. Fomin, Y. Sinai, *Ergodic Theory*, Springer, 1982.

[F] H. Furstenberg, *Recurrence in Ergodic Theory and Combinatorial Number Theory*, Princeton University Press, 1981.

[H] P. Halmos, Von Neumann on measure and ergodic theory, *Bull. AMS* **64** (1958), 86–94.

[K] B.O. Koopman, Hamiltonian systems and transformations in Hilbert space, *Proc. Nat. Acad. Sci. USA* **17** (1931), 315–318.

[KN] B.O. Koopman and J. von Neumann, Dynamical systems of continuous spectra, *Proc. Nat. Acad. Sci. USA* **18** (1932), 255–263.

[N1] J. von Neumann, Proof of the quasi-ergodic hypothesis, *Proc. Nat. Acad. Sci. USA* **18** (1932), 70–82.

[N2] J. von Neumann, John, Physical applications of the ergodic hypothesis, *Proc. Nat. Acad. Sci. USA* **18** (1932), 263–266.

[P] K. Petersen, *Ergodic Theory*, Cambridge Univ. Press, Cambridge, 1981.

[R] H.L. Royden, *Real Analysis*, Macmillan, New York, 1968.

[S] M.H. Stone, Linear transformations in Hilbert space. III. Operational methods and group theory, *Proc. Nat. Acad. Sci. USA* **16** (1930), 172–175.

[U] S.M. Ulam, *Adventures of a Mathematician*, Charles Scribner's Sons, New York, 1983.

[Wa] P. Walters, *An Introduction to Ergodic Theory*, Springer-Verlag, New York, 1982.

[We] H. Weyl, Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77** (1916), 313–352.