de Gruyter Studies in Mathematics 6

Editors: Heinz Bauer · Peter Gabriel

Ulrich Krengel

Ergodic Theorems

With a Supplement by Antoine Brunel

Walter de Gruyter
Berlin · New York 1985

Author
Dr. Ulrich Krengel
Professor of Mathematics
Universität Göttingen

Library of Congress Cataloging in Publication Data

Krengel, Ulrich, 1937–
Ergodic theorems.
(De Gruyter studies in mathematics ; 6)
Bibliography: p.
Includes index.
1. Ergodic theory. I. Brunel, Antoine. II. Title.
III. Series.
QA313.K74 1985 515.4'2 85-4457
ISBN 0-89925-024-6 (U.S.)

CIP-Kurztitelaufnahme der Deutschen Bibliothek

Krengel, Ulrich:
Ergodic theorems / Ulrich Krengel. With a suppl. by Antoine Brunel.
– Berlin ; New York : de Gruyter 1985.
(De Gruyter studies in mathematics ; 6)
ISBN 3-11-008478-3

NE: GT

Cover design: Rudolf Hübler, Berlin. Typesetting and Printing: Tutte Druckerei GmbH, Salzweg-Passau. Binding: Lüderitz & Bauer, Berlin.

Lesen heißt borgen,
daraus erfinden abtragen.

G. C. Lichtenberg,
Sudelbuch F

Preface

The study of ergodic theorems is the oldest branch of ergodic theory. It was started in 1931 by von Neumann and Birkhoff, having its origins in statistical mechanics. While new applications to mathematical physics continued to come in, the theory soon earned its own rights as an important chapter in functional analysis and probability.

So far, a comprehensive treatment has been neglected, and this book tries to provide it. Most of its material has not appeared in any other book. This applies even to older results, but the main body of the results is less than twenty years old and several interesting topics have just been added in the last decade.

Roughly speaking we ask: When do averages of quantities generated in a stationary manner converge? In the classical situation the stationarity is described by a measure preserving transformation τ, and one considers averages taken along a sequence $f, f \circ \tau, f \circ \tau^2, \ldots$ for integrable f. This corresponds to the probabilistic concept of stationarity. More generally, τ can be replaced by an operator T in a function space and $f \circ \tau^i$ by $T^i f$. As T^i is the result of the iterated action of the *same* operator we again have some kind of stationarity. Generalizing further, we study semigroups $\{T_g, g \in G\}$ of operators and limits of averages of $T_g f$ over subsets $I_n \subset G$. The term "ergodic theorem" has been used by some authors for quite distinct limit theorems, but we reserve it for convergence theorems dealing with such averages and for their close relatives. This meaning seems most widely accepted. Among the relatives we count subadditive ergodic theorems, local ergodic theorems (generalizing the differentiation of integrals), ratio ergodic theorems and ergodic theorems for information.

The modes of convergence under consideration mostly are norm convergence for "mean" ergodic theorems, and convergence almost everywhere for "individual" (or "pointwise") ergodic theorems. Recently, weak convergence has gained importance for nonlinear ergodic theorems and almost uniform convergence for ergodic theorems in von Neumann algebras. Convergence in distribution will not be considered. Typically, it applies to renormed averages rather than to averages, and it requires different tools.

I have tried to make the various parts of the book independently readable. The reader should start with any section which is of interest to him. He will then notice which previous results enter and find that often just a few will suffice. Of course, this implies some redundancy. On the other hand, I hope that this way large portions of the book can serve as a textbook, and that this approach will render this monograph useful and readily accessible for non-specialists.

I have not always given the shortest proof. Sometimes a longer proof seemed more transparent. Another aspect has been the wish to introduce a variety of methods. In some "additive" ergodic theorems the proof of convergence could have been simplified by the use of subadditive theorems. However, the longer additive arguments give access to an evaluation of the limit.

I presuppose knowledge of basic measure theory, and, for many sections, some functional analysis. But I tried to help the non-experts with references even for standard theorems.

Surely this book is biased towards my personal interests and even more so since I have included a number of new results and proofs. But I also tried hard not to miss any important result and to give it fair coverage. If a good presentation existed I sometimes may have just quoted it. I apologize in advance to anyone whose contributions were overlooked.

Concerning convergence almost everywhere the book of Stout [1974] covers many of the themes in the complement of this book.

Most sections end with Notes containing additional information. But I did include credits in the main text when it seemed possible without much delay.

Theorems, lemmas, definitions etc. are numbered consecutively in each section. A quotation "theorem 2.3.4" refers to chapter 2, section 3, theorem 3.4. A quotation "theorem 3.4" refers to the current chapter.

My indebtedness extends to many. First, I would like to thank my teacher, Konrad Jacobs, for generating my interest in ergodic theory, for giving me a sound introduction, and for suggesting a fertile area of research. I also owe much to Louis Sucheston and his infectious enthusiasm. I am very grateful to M. Lin, A. del Junco, Y. Derriennic, G. Keller, M. Akcoglu, R. Nagel, M. Mathieu, H.-J. Borchers, A. Bellow, R. Jajte, W. Takahashi, J. Fritz, M. Keane, M. Denker and many others for useful hints and encouragement. Martina Hochhaus helped with the bibliography. Marrie Powell contributed her unusual skill at mathematical typing. Special thanks go to Heinz Bauer for inviting the book into this series.

The idea for this book arose in 1976 at the end of a pleasant and interesting sabbatical I spent at the University of Paris VI. Antoine Brunel and I agreed that a book covering the whole spectrum of ergodic theorems was badly needed and planned to write it jointly. Unfortunately, grave personal reasons prevented this. I am the more grateful to Antoine Brunel for writing the supplement on Harris processes, a topic to which he contributed so much.

I devote this book to my wife Beate, and to Jeannette Brunel who died of cancer in 1981. They provided the environment for us in which devotion to mathematical work was possible.

Göttingen, April 1985 Ulrich Krengel

Contents

Chapter 1: Measure preserving and null preserving point mappings

We begin with the classical ergodic theorems for measure preserving transformations and with their role in the theory of stationary processes. Then the subadditive ergodic theorem is proved. We use it to derive the multiplicative ergodic theorem of Oseledeč, a powerful tool in the study of dynamical systems. Recurrence is discussed for the wider class of null preserving transformations. The dominated ergodic theorem provides important estimates of the supremum of averages. Some topics on measure preserving transformations like weak mixing, multiparameter semigroups, vector valued ergodic theorems, and the ergodic theorem for information are postponed although they could be read right away.

§ 1.1 Von Neumann's mean ergodic theorem, ergodicity

1. Definitions and examples. A *measurable space* $(\Omega, \mathscr{A})$ consists of a non empty set Ω and a σ-algebra $\mathscr{A}$, i.e., a non empty class of subsets of Ω, closed under the formation of complements and countable unions. A *measure* μ on $(\Omega, \mathscr{A})$ is a non negative set function $\mu\colon \mathscr{A} \to \mathbb{R}^+ \cup \{\infty\}$ with $\mu(\emptyset) = 0$ which is σ-additive. The triple $(\Omega, \mathscr{A}, \mu)$ is called a *measure space*. In the case $\mu(\Omega) = 1$, μ is called a *probability measure*, and $(\Omega, \mathscr{A}, \mu)$ a *probability space*.

Let $(\Omega, \mathscr{A})$ and $(\Omega', \mathscr{A}')$ be measurable spaces. A mapping $\tau\colon \Omega \to \Omega'$ is called *measurable* (or $\mathscr{A} - \mathscr{A}'$-measurable) if $\tau^{-1}\mathscr{A}' = \{\tau^{-1}A'\colon A' \in \mathscr{A}'\} \subset \mathscr{A}$. τ is called a *homomorphism* of $(\Omega, \mathscr{A}, \mu)$ into $(\Omega', \mathscr{A}', \mu')$ if τ is measurable and the measure $\mu \circ \tau^{-1}$ defined on $\mathscr{A}'$ by $(\mu \circ \tau^{-1})(A') = \mu(\tau^{-1}A')$ agrees with μ'.

$\tau\colon \Omega \to \Omega$ is called *measure preserving* if it is measurable and satisfies $\mu \circ \tau^{-1} = \mu$. In this case μ is called *invariant* or τ-invariant. A measure preserving transformation will also be called *endomorphism* (of $(\Omega, \mathscr{A}, \mu)$). If τ is an invertible endomorphism of Ω onto Ω for which τ^{-1} is an endomorphism, then τ is called an *automorphism*.

Examples of endomorphisms of measure spaces turn up in many branches of mathematics. Perhaps the simplest examples are translations in $\mathbb{R}^n$ and rotations $x \to x + \alpha \pmod 1$ in $[0, 1[$ with Lebesgue measure and, more generally, translations in locally compact groups Ω with left Haar measure.

Another class of measure preserving transformations in the same space is given by the continuous grouptheoretic automorphisms of Ω.

A simple number theoretic example can be defined via the expansion of $\omega \in \Omega = [0,1] \setminus \mathbb{Q}$ as a continued fraction: Identify ω with $(\omega_1, \omega_2, \dots) \in \mathbb{N}^{\mathbb{N}}$, where

$$\omega = \frac{1}{\vert \omega_1} + \frac{1}{\vert \omega_2} + \frac{1}{\vert \omega_3} + \dots .$$

Now $\omega \to \tau\omega = (\omega_2, \omega_3, \dots)$ defines an endomorphism in $(\Omega, \mathscr{A}, \mu)$ when μ is the measure with density $(1+x)^{-1}$ with respect to Lebesgue measure; see e.g. Billingsley [1965].

If Ω is a compact Hausdorff space and $\tau: \Omega \to \Omega$ continuous, there always exists an invariant measure on the σ-algebra $\mathscr{A}$ of Baire sets. As we want to return to this example later and a proof is simple if we make use of Banach limits, we take this liberty.

A *Banach limit* L is a linear functional defined on the space ℓ_∞ of bounded sequences $x = (x_0, x_1, \dots)$ of real numbers such that

(i) $L(x) \geqq 0$ holds for all x with $x_i \geqq 0 \quad (i = 0, 1, \dots)$,

(ii) $L((x_1, x_2, x_3, \dots)) = L((x_0, x_1, x_2, \dots)) \quad (x \in \ell_\infty)$ and

(iii) $L((1, 1, 1, \dots)) = 1$.

Banach limits exist; see theorem 3.4.1. Using a fixed Banach limit L and a fixed $\omega \in \Omega$, we can define a positive linear functional μ_ω on the space $C(\Omega)$ of continuous functions on Ω by $\mu_\omega(f) = L((f(\tau^n\omega))_{n=0}^\infty)$.

By the Riesz representation theorem (see Bauer [1981]) this linear functional is of the form $\mu_\omega(f) = \int f(\eta)\mu_\omega(d\eta)$ for some measure μ_ω on $\mathscr{A}$. The properties of L imply that μ_ω is an invariant probability measure.

The classical examples of endomorphisms which have originally motivated the search for ergodic theorems arise in statistical mechanics. A theorem of Liouville asserts the invariance of the $6r$-dimensional Lebesgue measure under the Hamiltonian flow in phase space, see Khintchine [1949].

Still other examples can be found in § 1.4. A rich collection of examples is given in the book of Cornfeld, Fomin, Sinai [1982].

If $(\Omega, \mathscr{A}, \mu)$ is a measure space, $\mathscr{L}_p = \mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ denotes the space of real or complex valued measurable functions f with $\|f\|_p := (\int |f|^p d\mu)^{1/p} < \infty$, $(1 \leqq p < \infty)$. $\mathscr{L}_\infty$ denotes the space of measurable functions for which $\|f\|_\infty := \inf\{\alpha > 0: \mu(\{|f| > \alpha\}) = 0\}$ is finite. We shall also write $\mathscr{L}_p(\mu)$ or $\mathscr{L}_p(\mathscr{A})$ if we want to mention the underlying measure or σ-algebra. $\{|f| > \alpha\}$ is a shorthand for $\{\omega \in \Omega: |f(\omega)| > \alpha\}$. We shall use such a shorthand notation also for other sets defined by properties of functions. Frequently we abreviate even further and write $\mu(|f| > \alpha)$ for $\mu(\{|f| > \alpha\})$.

Recall that $f = g \pmod{\mu}$ means $\mu(f \neq g) = 0$, and that equality mod μ is an equivalence relation in the space of measurable functions and in each space $\mathscr{L}_p$. The space $L_p = L_p(\Omega, \mathscr{A}, \mu)$ of equivalence classes in $\mathscr{L}_p$ is a Banach space with norm $\|\cdot\|_p$.

Most of the time we shall not distinguish between elements $f \in L_p$ and their representatives. In all statements involving only a sequence of elements of $\mathscr{L}_p$ or L_p and holding only mod μ, the difference is irrelevant.

The function 1_A which is 1 on $A \subset \Omega$ and 0 on the complement A^c is called *indicator function* of A. A and B are equal mod μ if the measure of their *symmetric difference* $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$ is zero. Again, sets which are equal mod μ will usually not be distinguished. $\mathbb{1}$ denotes the function $\equiv 1$.

A measure v on $\mathscr{A}$ is called *μ-continuous* if $v(A) = 0$ holds for all $A \in \mathscr{A}$ with $\mu(A) = 0$. We then write $v \ll \mu$. Call v *equivalent* to $\mu (v \sim \mu)$ if $v \ll \mu \ll v$.

If $(\Omega, \mathscr{A})$ is a measurable space and $\tau: \Omega \to \Omega$ a measurable mapping, the operator $f \to Tf := f \circ \tau$ is called *composition with* τ. It is a linear operator in the space of measurable functions on Ω.
A measurable map $\tau: \Omega \to \Omega$ with $\mu \circ \tau^{-1} \ll \mu$ is called *null preserving*. An invertible null preserving map τ of Ω onto Ω for which also τ^{-1} is null preserving is called *nonsingular*. For null preserving τ the composition operator is well defined in the spaces of equivalence classes because then $f_1 = f_2$ mod μ implies $f_1 \circ \tau = f_2 \circ \tau$ mod μ.

We shall make frequent use of the following notions from functional analysis:

Definition 1.1. If $\mathfrak{X}$, $\mathfrak{Y}$ are normed vector spaces and T a linear operator mapping $\mathfrak{X}$ into $\mathfrak{Y}$, the *norm* $\|T\|$ of T is given by

$$\|T\| = \sup_{\|f\| \leqq 1} \|Tf\|.$$

T is called *bounded* if $\|T\|$ is finite, and T is called a *contraction* if $\|T\| \leqq 1$. T is called an *isometry* if $\|Tf\| = \|f\|$ holds for all f.

If a partial order is defined in $\mathfrak{X}$, the *positive cone* $\{f \in \mathfrak{X}: f \geqq 0\}$ of $\mathfrak{X}$ is denoted by $\mathfrak{X}^+$. T is called *positive* if $T\mathfrak{X}^+ \subset \mathfrak{Y}^+$. In the case $\mathfrak{X} = \mathfrak{Y}$ we speak of a linear operator (contraction, ...) in $\mathfrak{X}$. $\mathscr{L}(\mathfrak{X})$ denotes the set of bounded linear operators in $\mathfrak{X}$. I denotes the identity operator.

If τ is an endomorphism of $(\Omega, \mathscr{A}, \mu)$ the composition with τ is a positive isometry in each L_p. If τ is only null preserving or nonsingular, it still is a positive contraction in L_∞, but it need not map L_p into L_p for $1 \leqq p < \infty$.

2. Contractions in Hilbert space. We denote the scalar product of two elements f, h of a Hilbert space $\mathfrak{H}$ by $\langle f, h \rangle$. The dual T^* of a bounded linear operator T in $\mathfrak{H}$ is characterized by $\langle Tf, h \rangle = \langle f, T^* h \rangle$, valid for all f and h.

Lemma 1.2. *If $T: \mathfrak{H} \to \mathfrak{H}$ is a contraction in a real or complex Hilbert space and $g \in \mathfrak{H}$, then $g = Tg$ holds if and only if $g = T^* g$.*

Proof. If, for some g, $\langle g, Tg \rangle = \|g\|^2$, then $\langle g, Tg \rangle$ is real and $\langle g, Tg \rangle = \langle Tg, g \rangle$. We then get

$$\begin{aligned} \| Tg - g \|^2 &= \langle Tg - g, Tg - g \rangle = \| Tg \|^2 + \| g \|^2 - 2\langle g, Tg \rangle \\ &\leqq 2\|g\|^2 - 2\|g\|^2 = 0. \end{aligned}$$

Thus, $g = Tg$ is equivalent to $\|g\|^2 = \langle g, Tg \rangle = \langle T^* g, g \rangle$. Applying this equivalence to T^* the identity $g = T^* g$ follows. □

Lemma 1.3. *Let $\mathscr{T}$ be a family of contractions T in a Hilbert space $\mathfrak{H}$. Then the orthogonal complement $F^\perp$ of $F = \{g \in \mathfrak{H}: Tg = g \forall T \in \mathscr{T}\}$ is the closure of the subspace N spanned by $\{h - Th: h \in \mathfrak{H}, T \in \mathscr{T}\}$.*

Proof. Write $g \perp \mathfrak{H}_0$ if g is orthogonal to a subspace $\mathfrak{H}_0$. Now $g \perp N \Leftrightarrow \langle g, (T - I)h \rangle = 0 \forall h \in \mathfrak{H}, T \in \mathscr{T} \Leftrightarrow \langle T^* g - g, h \rangle = 0 \forall h, T \Leftrightarrow T^* g = g \forall T \Leftrightarrow Tg = g \forall T \Leftrightarrow g \in F$.

Thus N, and hence also its closure $cl\, N$, is orthogonal to the closed subspace F. As any vector orthogonal to $cl\, N$ belongs to F we have $F^\perp = cl\, N$. □

The following notation will be used frequently for linear operators T:

$$S_n f := S_n(T) f := \sum_{i=0}^{n-1} T^i f, \quad A_n f := A_n(T) f := n^{-1} S_n(T) f.$$

If the operator under consideration is S we may also write $A_n f$ for $A_n(S) f$. If T is the composition with τ we sometimes write $A_n(\tau)$ for $A_n(T)$.

Theorem 1.4 (Mean ergodic theorem of von Neumann). *If T is a contraction in a Hilbert space $\mathfrak{H}$, and P the projection on $F = \{g \in \mathfrak{H}: Tg = g\}$, then $A_n f$ converges in norm to Pf for $f \in \mathfrak{H}$, $(n \to \infty)$.*

Proof. If $f = (T - I)h$ for some h then

$$\begin{aligned} \|A_n f\| &= n^{-1} \| Th - h + T^2 h - Th + \ldots + T^n h - T^{n-1} h \| \\ &= n^{-1} \| T^n h - h \| \leqq 2n^{-1} \|h\| \to 0. \end{aligned}$$

By approximation this yields $\|A_n f\| \to 0$ for all f in the closure of $(T - I)\mathfrak{H}$, and now the assertion follows from Lemma 1.3. □

If τ is an endomorphism in $(\Omega, \mathscr{A}, \mu)$ and $p = 2$, it follows that for any $f \in L_p$ there is an $\bar{f} \in L_p$ with $\bar{f} = \bar{f} \circ \tau$ and $\|\bar{f}\|_p \leqq \|f\|_p$ such that $\|A_n(\tau) f - \bar{f}\|_p \to 0$ $(n \to \infty)$. Approximation arguments (or theorem 2.1.1) show that this remains true for $1 < p < \infty$. If $\mu(\Omega) = \infty$ the analogous statement does not always hold for $p = 1$; e.g., take $\Omega = \mathbb{R}^1$, $\mu =$ Lebesgue measure, $\tau\omega = \omega + 1$, and $f = 1_{[0,1[}$. For $f \in L_\infty$, $\|A_n(\tau) f - \bar{f}\|_\infty \to 0$ need not even hold in the case $\mu(\Omega) < \infty$. One can take $\Omega = [0, 1[$, $\mu =$ Lebesgue measure, $\tau\omega = \omega + \alpha \pmod 1$ with an irrational α, and a suitable (highly discontinuous) $f \in L_\infty$. We leave this as an exercise to the reader.

3. Absorbing and invariant sets. In many cases the property $\bar{f} = \bar{f} \circ \tau$ of the limit function can be used to find the explicit form of $\bar{f}$. For future use we discuss the relevant notions in some more generality than needed here.

Definition 1.5. Let τ be null preserving in $(\Omega, \mathscr{A}, \mu)$. A set $A \in \mathscr{A}$ is called τ-*absorbing* if $A \subset \tau^{-1} A$, τ-absorbing mod μ if $\mu(A \setminus \tau^{-1} A) = 0$, τ-*invariant* if $\tau^{-1} A = A$, and τ-*invariant* mod μ if $\mu(A \Delta \tau^{-1} A) = 0$. An $\mathscr{A}$-measurable function f is called τ-*invariant* if $f = f \circ \tau$ and τ-invariant mod μ if $\mu(f \neq f \circ \tau) = 0$.

Thus A is τ-absorbing if no orbit $\omega, \tau\omega, \tau^2\omega, \ldots$ starting in A leaves A. If A is τ-absorbing mod μ the set $A^1 = A \setminus \bigcup_{k=0}^{\infty} \tau^{-k}(A \setminus \tau^{-1} A)$ differs from A only by a set of measure 0 and is τ-absorbing. Similarly, if A is τ-invariant mod μ the set $A^2 = \bigcup_{k=0}^{\infty} \tau^{-k} A^1$ is τ-invariant and equal to A mod μ. Thus, in all considerations where sets of measure 0 do not matter, we need not distinguish the notions τ-invariant and τ-invariant mod μ. A real valued f is τ-invariant if and only if $\{f > \alpha\}$ is τ-invariant for all $\alpha \in \mathbb{R}$. A complex valued f is τ-invariant if both the real and the complex part are τ-invariant. These observations can be used to show that for functions f which are τ-invariant mod μ there exists a τ-invariant f' with $\mu(f \neq f') = 0$, and we need not distinguish the notions τ-invariant and τ-invariant mod μ for functions either.

For any null preserving τ the family $\mathscr{I}$ of invariant sets clearly is a σ-algebra. By the above remarks a function f is τ-invariant if and only if it is $\mathscr{I}$-measurable. If τ is an endomorphism and μ finite, the σ-algebra $\mathscr{I}$ can be used to to express the limit $\bar{f}$ of $A_n(\tau) f$ as a *conditional expectation*:

Let $\mathscr{F}$ be a sub-σ-algebra of $\mathscr{A}$, such that the restriction of μ to $\mathscr{F}$ is σ-finite. Recall that for any $f \in L_1(\Omega, \mathscr{A}, \mu)$ there exists (by the Radon-Nikodym theorem) an $f_0 \in L_1$ which is $\mathscr{F}$-measurable and satisfies

$$\int_A f_0 \, d\mu = \int_A f \, d\mu \quad \forall A \in \mathscr{F},$$

and that f_0 is uniquely determined mod μ. f_0 is called the conditional expectation of f with respect to $\mathscr{F}$, and denoted by $E(f \mid \mathscr{F})$ or by $E_\mu(f \mid \mathscr{F})$ if we emphasize the fact that the basic measure is μ. It follows from the Jensen inequality that the map $f \to E(f \mid \mathscr{F})$ is a contraction $E_{\mathscr{F}}$ in each space $L_p(\Omega, \mathscr{A}, \mu)$.

Proposition 1.6. *If τ is an endomorphism in $(\Omega, \mathscr{A}, \mu)$, and μ σ-finite on the σ-algebra $\mathscr{I}$ of τ-invariant sets, then, for any $f \in L_2(\Omega, \mathscr{A}, \mu)$, the norm-limit $\bar{f}$ of $A_n(\tau) f$ is given by $\bar{f} = E(f \mid \mathscr{I})$.*

Proof. We may assume $\mu(\Omega) < \infty$. $\bar{f}$ is $\mathscr{I}$-measurable. For any $A \in \mathscr{I}$ we have

$\int_A f \circ \tau^k d\mu = \int_A f d\mu$ because A is τ-invariant and τ an endomorphism. Now $\int_A f d\mu = \langle A_n(\tau)f, 1_A \rangle \to \langle \bar{f}, 1_A \rangle = \int_A \bar{f} d\mu$ because strong convergence implies weak convergence. □

Definition 1.7. A null preserving transformation τ in $(\Omega, \mathscr{A}, \mu)$ is called *ergodic* if all τ-invariant sets A have the property that $\mu(A) = 0$ or $\mu(A^c) = 0$.

Thus, τ is ergodic if the space cannot be decomposed into two non trivial τ-invariant subsets. If an endomorphism is ergodic and $0 < \mu(\Omega) < \infty$, the limit $\bar{f}$ is simply given by $\bar{f} = \mu(\Omega)^{-1} \int f d\mu$. This means that for large n the space average $\mu(\Omega)^{-1} \int f d\mu$ is very close to the time average $A_n(\tau)f$. The important *ergodic hypothesis* in statistical mechanics is the assumption that these two averages are asymptotically equal for the endomorphisms arising in the Hamiltonian flow in phase space. Von Neumann's theorem made it clear that the limit of the time averages does exist in the sense of L_2-convergence and that it is equal to the space average in the ergodic case. The question of ergodicity of the transformations constituting the Hamiltonian flow was left open. For many endomorphisms the proof of their ergodicity is fairly simple, but for some, including the endomorphisms coming from the Hamiltonian flow, the question whether they are ergodic is very deep; see the Notes.

4. Criteria of ergodicity. We end this section by describing some general necessary and sufficient conditions for ergodicity. They do not go very far beyond reformulations of the definition. The proof of the ergodicity for specific examples usually requires arguments which exploit the specific nature of the examples.

(a) For general null preserving τ we can just say that τ is ergodic iff each measurable τ-invariant function f is constant μ-a.e. This follows from our above remarks on τ-invariant sets and functions. (As usual "iff" means "if and only if").

(b) If τ is nonsingular, τ is ergodic iff $\mu(A) > 0$ implies that the complement of $A^* = \bigcup_{k=-\infty}^{+\infty} \tau^{-k} A$ is a nullset. A^* is the smallest τ-invariant set containing A. An obvious equivalent condition is that $\mu(A) > 0$, $\mu(B) > 0$ imply the existence of an integer k with $\mu(\tau^k A \cap B) > 0$.

Proposition 1.8. *For an endomorphism τ of a finite measure space $(\Omega, \mathscr{A}, \mu)$ each of the following conditions is equivalent to the ergodicity of τ:*

(e 1) $\mu(A) > 0 \Rightarrow \mu((A^-)^c) = 0$, *where* $A^- = \bigcup_{k=0}^{\infty} \tau^{-k} A$;

(e 2) $\mu(A) > 0$, $\mu(B) > 0 \Rightarrow \exists k \geqq 0$ *with* $\mu(\tau^{-k} A \cap B) > 0$;

(e 3) *For all* $A, B \in \mathscr{A}$, $\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} \mu(\tau^{-k} A \cap B) = \mu(A)\mu(B)\mu(\Omega)^{-1}$;

(e 4) There exists a family $\mathcal{M} \subset \mathcal{A}$ such that the linear combinations of the functions 1_E ($E \in \mathcal{M}$) lie dense in L_2 and which has the property that $A_n(\tau)1_E$ converges weakly in L_2 to a constant;

(e 5) For all $f, g \in L_2$, $\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \langle f \circ \tau^k, g\rangle = \langle f, \mathbb{1}\rangle \langle \mathbb{1}, g\rangle \mu(\Omega)^{-1}$.

Proof. The equivalence of (e 1) and (e 2) to ergodicity follows because $\tau^{-1}A^- \subset A^-$ and $\mu(\tau^{-1}A^-) = \mu(A^-)$ imply $\mu(A^- \setminus \tau^{-1}A^-) = 0$. Von Neumann's theorem yields the existence of the limit in (e 5) (and hence in (e 3) and (e 4)) and the limit is $\langle \bar{f}, g\rangle$. If τ is ergodic $\bar{f}$ is constant so that the identity $\langle \bar{f}, \mathbb{1}\rangle = \langle f, \mathbb{1}\rangle$ implies $\bar{f} = \mu(\Omega)^{-1}\langle f, \mathbb{1}\rangle$. But $\langle \bar{f}, g\rangle = \langle f, \mathbb{1}\rangle\langle \mathbb{1}, g\rangle \mu(\Omega)^{-1}$. Taking $f = 1_A$, $g = 1_B$ we get (e 3). If τ is not ergodic, (e 3) and hence (e 5) cannot hold because (e 2) doesn't. The equivalence of (e 4) is obtained by an approximation argument. □

Notes

Koopman [1931] observed that, using Liouville's theorem and the measure preserving property, the Hamiltonian flow could be studied via the induced group of unitary operators in Hilbert space. This idea led von Neumann [1931] to a proof of his ergodic theorem via spectral theory.

The measure theoretic concept of ergodicity seems to go back to Birkhoff and Smith [1928], who used the term *metrically transitive*, still favoured by some authors.

Sinai [1963] announced a theorem which – roughly speaking – says that a system of n balls of equal diameter following the laws of elastic reflection in a cubic box is ergodic. No published proof of this result seems to exist. Sinai [1970] considered the movement of only *one* ball in a 2-dimensional domain with smooth strictly positive curvature (dispersing billiards). This was simplified and generalized by Bunimovich-Sinai [1973]. Also Kubo [1976], Kubo-Murata [1981], Gallavotti [1975], and Keller [1977] contributed to this subject.

Assuming, in addition, "finite horizon", Gallavotti and Ornstein [1974] proved that dispersing billiards are isomorphic to Bernoulli shifts.

We refer to Bunimovich [1982] for a survey on recent developments in this area.

§ 1.2 Birkhoff's ergodic theorem

1. Discrete time. Our next aim is to give a proof of George D. Birkhoff's famous pointwise ergodic theorem. For later use we formulate some of the arguments in the more general operator theoretic setting. We write

$$(2.1) \qquad M_n^S f = \operatorname{Max}(S_1 f, \ldots, S_n f), \quad M_n f = \operatorname{Max}(A_1 f, \ldots, A_n f),$$

and

$$M_\infty f = \sup_{n \geq 1} M_n f.$$

When this notation is applied to an endomophism τ, T will be the composition with τ.

The key step in the proof of the pointwise ergodic theorem is the following *maximal ergodic theorem*, which was discovered for endomorphisms by Yosida-Kakutani [1939] and for general positive contractions in L_1 by Hopf [1954]:

Theorem 2.1. *Let T be a positive contraction in $L_1(\Omega, \mathscr{A}, \mu)$. For real valued $f \in L_1$ put $E_n = \{M_n f \geqq 0\}$ $(= \{M_n^S f \geqq 0\})$ and $E_\infty = \bigcup_{n=1}^{\infty} E_n$. Then*

$$\int_{E_n} f d\mu \geqq 0 \quad \text{and} \quad \int_{E_\infty} f d\mu \geqq 0 .$$

Proof. We follow the elegant argument of A.M. Garsia [1965]: For $k = 1, \ldots, n$, $(M_n^S f)^+ \geqq S_k f$ and hence $f + T(M_n^S f)^+ \geqq f + TS_k f = S_{k+1} f$. Thus $f \geqq S_k f - T(M_n^S f)^+$ holds for $k = 1, \ldots, n$ because it is trivial for $k = 1$. Passing to maxima one obtains $f \geqq M_n^S f - T(M_n^S f)^+$. Now integrate over E_n:

$$\begin{aligned}
\int_{E_n} f d\mu &\geqq \int_{E_n} (M_n^S f - T(M_n^S f)^+) d\mu \\
&= \int_{E_n} ((M_n^S f)^+ - T(M_n^S f)^+) d\mu \\
&= \int_{\Omega} (M_n^S f)^+ d\mu - \int_{E_n} T(M_n^S f)^+ d\mu \\
&\geqq \int_{\Omega} (M_n^S f)^+ d\mu - \int_{\Omega} T(M_n^S f)^+ d\mu \geqq 0
\end{aligned}$$

because $\int h d\mu \geqq \int T h d\mu$ holds for $h \in L_1^+$. A passage to the limit yields the second inequality. □

This proof is a bit miraculous. In section 3.2 a discussion of the filling scheme will provide a longer but more intuitive proof.

The following *maximal ergodic inequality* was already known to Wiener [1939]. It admits also a simple direct proof (see theorem 5.2), and suffices for the proof of Birkhoff's theorem given in theorem 7.3.

Corollary 2.2. *If τ is an endomorphism in a measure space $(\Omega, \mathscr{A}, \mu)$, then*

$$\mu(M_n f \geqq \alpha) \leqq \alpha^{-1} \|f\|_1$$

holds for any real valued $f \in L_1$ and $\alpha > 0$.

Proof. It is enough to prove $\|f\|_1 \geqq \alpha \mu(A)$ for arbitrary sets A of finite measure contained in $\{M_n f \geqq \alpha\}$. Put $E_{n,A} = \{M_n(f - \alpha 1_A) \geqq 0\}$. By theorem 2.1,

$$\int_{E_{n,A}} (f - \alpha 1_A)\, d\mu \geqq 0 .$$

For any $\omega \in A \subset \{M_n f \geqq \alpha\}$ there exists a $k \leqq n$ with $A_k f(\omega) \geqq \alpha$. This implies $S_k(f - \alpha \mathbb{1})(\omega) \geqq 0$ and $S_k(f - \alpha 1_A)(\omega) \geqq 0$. Thus A is contained in $E_{n,A}$, and

$$\|f\|_1 \geqq \int_{E_{n,A}} f d\mu \geqq \alpha \int_{E_{n,A}} 1_A d\mu = \alpha\mu(A). \quad \square$$

Theorem 2.3 (Birkhoff's ergodic theorem). *If τ is an endomorphism in a measure space $(\Omega, \mathscr{A}, \mu)$ and $f \in L_1$ (real or complex), then the averages $A_n f$ converge μ — a.e. to some τ-invariant $\bar{f}$ with $\|\bar{f}\|_1 \leqq \|f\|_1$. For each τ-invariant $A \in \mathscr{A}$ with $\mu(A) < \infty$*

$$\int_A \bar{f} d\mu = \int_A f d\mu . \tag{2.1}$$

Proof. First consider a real valued f. Because of

$$A_{n+1} f = (n+1)^{-1} S_{n+1} f = (n+1)^{-1} f + \frac{n}{n+1}\left(\frac{1}{n} S_n f\right) \circ \tau$$

the functions $f^u = \limsup_{n\to\infty} A_n f$ and $f^l = \liminf_{n\to\infty} A_n f$ are τ-invariant. We show that f^u and f^l assume the values $\pm\infty$ only on a set of measure 0:

For any $\beta > 0$ the τ-invariant set $D_\beta = \{f^u > \beta\}$ is contained in the union of the increasing sequence $\{M_n f \geqq \beta\}$. By the maximal inequality $\mu(M_n f \geqq \beta) \leqq \beta^{-1}\|f\|_1$. Hence $\mu(D_\beta) \leqq \beta^{-1}\|f\|_1$. Passing with β to infinity $f^u < \infty$ a.e. follows. By symmetry $f^l > -\infty$ a.e., and $\mu(f^l < \alpha) = \mu(-\limsup A_n(-f) < \alpha) \leqq |\alpha|^{-1}\|f\|_1$ for $\alpha < 0$.

If $A_n f$ does not converge a.e. there exist rational numbers $\alpha < \beta$ such that the set $B = \{f^l < \alpha < \beta < f^u\}$ has positive measure. $\mu(B)$ cannot be infinite because $\alpha < 0$ or $\beta > 0$. By the τ-invariance of B the function $f' = (f - \beta) 1_B$ has the property that $f' \circ \tau^k$ vanishes outside B for all $k \geqq 0$ and $B = \{\omega : \exists n \geqq 1 \text{ with } S_n f'(\omega) > 0\}$.

The maximal ergodic theorem implies $\beta\mu(B) \leqq \int_B f d\mu$.

A symmetric argument with $f'' = (\alpha - f) 1_B$ gives us $\int_B f d\mu \leqq \alpha\mu(B)$. Together this contradicts $\alpha < \beta$.

We have shown that $f^u = f^l$ a.e., and that these functions μ — a.e. assume only finite values. By the decomposition $f = f^+ - f^-$ we may assume $f \in L_1^+$ for the proof of $\|\bar{f}\|_1 \leqq \|f\|_1$. The lemma of Fatou then yields

$$\int \bar{f} d\mu = \int \liminf A_n f d\mu \leqq \liminf \int A_n f d\mu = \int f d\mu$$

where the last equality follows from $\int f \circ \tau^k d\mu = \int f d\mu$.

To prove the last statement in the theorem we may assume $\Omega = A$, $\mu(\Omega) < \infty$, and $f \geqq 0$. For any $\varepsilon > 0$ there exists a $K_\varepsilon \geqq 0$ such that $g_\varepsilon = f - (f \wedge K_\varepsilon)$ has norm $\|g_\varepsilon\|_1 < \varepsilon$. Now

$$\int (A_n f - K_\varepsilon)^+ \, d\mu \leqq \int A_n g_\varepsilon \, d\mu < \varepsilon$$

shows that the sequence $A_n f$ is uniformly integrable. As any uniformy.integrable sequence converging μ-a.e. converges in L_1-norm, the assertion $\int \bar{f} d\mu = \int f d\mu$ follows from $\int A_n f = \int f d\mu$.

For complex valued f one can use the decomposition into the real and the imaginary parts to prove the existence of the limit, and observe $|\bar{f}| \leqq \lim A_n |f|$. □

The condition (2.1) determines $\bar{f}$ uniquely: We can assume that μ is σ-finite. There exists a set $\Omega_0 \in \mathscr{I}$ such that every $A \in \mathscr{I}$ with $\mu(A) < \infty$ is contained mod μ in Ω_0, and μ is σ-finite on $\Omega_0 \cap \mathscr{I}$. $\bar{f}$ vanishes in Ω_0^c and is given by $E(f \mid \mathscr{I})$ in Ω_0. In particular, if τ is an ergodic endomorphism in a σ-finite infinite measure space, then $\bar{f}$ must vanish for all f.

2. Continuous time. We have stated the ergodic theorem for a single endomorphism τ. Sometimes one is interested in a continuous time motion of the points ω, and in a corresponding continuous time theorem. There is no difficulty to derive such a result from theorem 2.3. By a *flow* $\{\tau_t, t \in \mathbb{R}\}$ we mean a group of measurable transformations $\tau_t\colon \Omega \to \Omega$ with $\tau_0 =$ identity, $\tau_{t+s} = \tau_t \circ \tau_s$, $(t, s \in \mathbb{R})$. The flow will be called measure preserving if the τ_t are measure preserving. The flow is called *measurable*, if the map $(\omega, t) \to \tau_t \omega$ from $\Omega \times \mathbb{R}^1$ into Ω is $\tilde{\mathscr{A}} - \mathscr{A}$-measurable, where $\tilde{\mathscr{A}}$ is the completion of the product-σ-algebra $\mathscr{A} \otimes \mathscr{B}$ of $\mathscr{A}$ with the Borel sets, and the completion is taken with respect to the product $\tilde{\mu}$ of μ on $\mathscr{A}$ and the Lebesgue measure λ on $\mathscr{B}$. The completely analoguous definitions can be given for *semiflows* $\{\tau_t, t \geqq 0\}$, (in which the τ_t need not be invertible). If $\{\tau_t, t \geqq 0\}$ is a measurable measure preserving semiflow in a σ-finite measure space, and if $f\colon \Omega \to \mathbb{R}$ is integrable, the function $\tilde{f}$ defined by $\tilde{f}(\omega, t) = f(\tau_t \omega)$ is, for all $T > 0$, integrable in $\Omega \times [0, T]$ (Fubini), and, hence, for μ-a.e. $\omega \in \Omega$ the integrals $\int_0^T f(\tau_t \omega)\, dt$ are well-defined. Note that $\int_0^n f(\tau_t \omega)\, dt = \sum_{i=0}^{n-1} F(\tau_1^i \omega)$ with $F(\omega) = \int_0^1 f(\tau_t \omega)\, dt$ and that F is integrable. Therefore the ergodic theorem implies that $n^{-1} \int_0^n f(\tau_t \omega)\, dt$ converges a.e. when the integers n tend to infinity. The ergodic theorem also implies $n^{-1} F_0 \circ \tau_1^{n-1} \to 0$ a.e., where $F_0 = \int_0^1 |f \circ \tau_t|\, dt$. For $n \leqq T < n+1$, $\left|\int_0^T f \circ \tau_t\, dt - \int_0^n f \circ \tau_t\, dt\right| \leqq F_0 \circ \tau_1^{n-1}$. Thus the convergence a.e. of $T^{-1} \int_0^T f \circ \tau_t\, dt$ when T tends to infinity along the reals is a rather trivial consequence of the discrete time theorem. Many of the discrete parameter theorems in this book will have such continuous parameter analogues and we shall usually not care to state the latter.

There is, however, another class of continuous parameter theorems which will be of more interest to us. These are the *local ergodic theorems*. They assert the convergence of continuous time averges over time intervals $[0, \varepsilon]$ when $\varepsilon \to 0 + 0$. They were introduced by N. Wiener [1939] for measurable measure preserving flows. In this case they are a consequence of a form of the fundamental theorem of calculus, which says that for Lebesgue integrable f on $[0, \infty[$ one has

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_s^{s+\varepsilon} f(t)\, dt = f(s)$$

for λ-almost all s; see Royden [1968: Ch. 5].

Theorem 2.4 (Wiener's local ergodic theorem). *If $(\Omega, \mathscr{A}, \mu)$ is a σ-finite measure space and $\{\tau_t, t \geqq 0\}$ a measure preserving measurable semiflow, $f \in \mathscr{L}_1(\mu)$, then* $\lim_{\varepsilon \to 0+0} \varepsilon^{-1} \int_0^{\varepsilon} f(T_\alpha \omega)\, d\alpha = f(\omega)$ *holds* μ-a.e..

Proof. Let $\tilde{N}$ denote the complement of the set of points (ω, t) in $\Omega \times [0, \infty[$ with $\lim_{\varepsilon \to 0+0} \varepsilon^{-1} \int_0^{\varepsilon} f(\tau_{t+\alpha}\omega)\, d\alpha = f(\tau_t \omega)$ and let $N_\omega = \{t \in [0, \infty[: (\omega, t) \in \tilde{N}\}$ and $N^t = \{\omega \in \Omega : (\omega, t) \in \tilde{N}\}$. For μ-almost all ω, $\int_0^t f(\tau_\alpha \omega)\, d\alpha$ is well defined for all t and $\tilde{f}(\omega, \alpha)$ is integrable in each $[0, T]$. The fundamental theorem of calculus implies $\lambda(N_\omega) = 0$ for these ω.

$\tilde{\mu}(\tilde{N}) = 0$ follows, and, by Fubini, λ-almost all t have the property that $\mu(N^t) = 0$. But $N^t = \tau_t^{-1} N^0$. As τ_t is measure preserving $\mu(N^0) = 0$ follows. □

The integrability of $\tilde{f}(\omega, \cdot)$ with respect to λ is also clear for bounded f, when the τ_t are not necessarily measure preserving. If $\mu(N^t) = 0$ for almost all $t > 0$ and the τ_t are null preserving $\mu(N^t) = 0$ follows for all $t > 0$. Therefore the above argument also proves:

Theorem 2.5. *If $(\Omega, \mathscr{A}, \mu)$ is a σ-finite measure space and $\{\tau_t, t \geqq 0\}$ a null preserving measurable semiflow with the property that*

$$\mu(\tau_t^{-1} A) = 0 \quad \textit{for all } t > 0 \quad \textit{implies } \mu(A) = 0$$

then

$$\lim_{\varepsilon \to 0+0} \varepsilon^{-1} \int_0^{\varepsilon} f(\tau_t \omega)\, dt = f(\omega) \quad \mu\text{-a.e.}$$

holds for all bounded measurable f.

3. Uniform convergence. Simple examples like the Bernoulli shifts (§ 1.4) show that $A_n f$ need not converge *everywhere* even when Ω is a compact topological space, and τ and f are continuous. However, for some interesting examples one obtains even uniform convergence by the following special case of the mean ergodic theorem in Banach spaces:

Theorem 2.6. *Let Ω be a compact metric space with metric ϱ. If $\tau\colon \Omega \to \Omega$ is continuous, and the functions $A_n f$, $(n \geqq 1)$, are equicontinuous, then $A_n f$ converges uniformly in Ω.*

(Recall that the functions f_n are called *equicontinuous* if for any $\varepsilon > 0$ there exists $\delta > 0$ such that $\varrho(\omega, \omega') < \delta$ implies $|f_n(\omega) - f_n(\omega')| < \varepsilon$ for all n).

Proof. Let $\bar{\mathcal{O}}(\omega)$ denote the closure of the orbit $\{\omega, \tau\omega, \ldots\}$ of a point ω. If some $g \in C(\Omega)$ is 0 on $\bar{\mathcal{O}}(\omega)$, then $\mu_\omega(g) = 0$, where μ_ω is the linear functional constructed in subsection 1 of § 1.1. Thus $\mu_\omega(\bar{\mathcal{O}}(\omega)) = 1$.

By the Birkhoff theorem there exists an $\omega^* \in \bar{\mathcal{O}}(\omega)$ for which $A_n f(\omega^*)$ converges to some finite $\bar{f}(\omega^*)$. Given $\varepsilon > 0$ find $\delta > 0$ by the equicontinuity of the $A_n f$ and then find m with $\varrho(\tau^m \omega, \omega^*) < \delta$. $|A_n f(\omega) - A_n f(\tau^m \omega)|$ tends to 0 for $n \to \infty$, since all but $2m$ terms in the sums cancel and f is bounded. $|A_n f(\tau^m \omega) - A_n f(\omega^*)| < \varepsilon$ implies $|A_n f(\omega) - \bar{f}(\omega^*)| < 2\varepsilon$ for large n. As ε was arbitrary $\lim A_n f(\omega)$ exists.

By the compactness of Ω there exist $\omega_1, \ldots, \omega_K \in \Omega$ so that the open $\delta-$balls B_k with these centers cover Ω. If N is large enough, then $|A_n f(\omega_k) - \bar{f}(\omega_k)| < \varepsilon$, for $n \geqq N$ and $k = 1, \ldots, K$. If ω is arbitrary there exists k with $\varrho(\omega, \omega_k) < \delta$. Now $|A_n f(\omega) - A_n f(\omega_k)| < \varepsilon$ shows that the convergence is uniform. □

Example. Consider the d-dimensional torus $\Omega = [0, 1[^d$, which is a compact group with coordinatewise addition mod 1. Take some $\alpha = (\alpha_1, \ldots, \alpha_d) \in \Omega$, for which $\alpha_1, \alpha_2, \ldots, \alpha_d, 1$ are *integrally independent*, i.e., $m_1 = m_2 = \ldots = m_d = 0$ shall be the only integers for which $\sum_{k=1}^{d} \alpha_k m_k$ is an integer.

The translation $\tau\colon \omega \to \omega + \alpha$ clearly preserves the d-dimensional Lebesgue measure $\mu = \lambda^d$. τ is ergodic. This can be proved using Fourier analysis; see e.g. Petersen [1983: p. 51]. We sketch a proof using only measure theory. First show by induction on d that the sequence $\ldots, -2\alpha, -\alpha, 0, \alpha, 2\alpha, 3\alpha, \ldots$ is dense in Ω: As all these points are different there is an accumulation point, and, hence, for any $\varepsilon > 0$, a $k \in \mathbb{Z}$ with $0 < |\alpha_i'| < \varepsilon/2d$, where $\alpha_i' = k\alpha_i \bmod 1$. Because of the induction hypothesis the line $\{x\alpha'\colon x \in \mathbb{R}\}$ through $\alpha' = (\alpha_1', \ldots, \alpha_d')$ and $(0, 0, \ldots, 0)$ lies densely in Ω. (It intersects $\{0\} \times [0,1[^{d-1}$ in a dense subset.) For any point z on this line there is some $m\alpha'$ $(m \in \mathbb{Z})$ in an $\varepsilon/2$-neighbourhood of z. Thus, for any $y \in \Omega$, there is some $mk\alpha$ in the ε-neighbourhood of y. (A similar argument shows the density of the "forward orbit" $\alpha, 2\alpha, \ldots$, which was already known to Kronecker [1884]).

Now let A be a τ-invariant set of positive measure, and let $0 < \eta < 1$ be given. For $B, C \in \mathscr{A}$ we say that B fills out $(1-\eta)$ of C when $\mu(B \cap C) \geqq (1-\eta)\,\mu(C)$. If $\xi > 0$ is sufficiently small one can fill out $(1-\eta)$ of Ω with say $k(\xi)$ disjoint cubes of side-length ξ which lie at a strictly positive distance of each other. There are arbibrarily small ξ such that A fills out $(1-\eta)$ of *some* cube C_ξ of side length ξ. By the density statement above, we now can find $k(\xi)$ disjoint translates $\tau^i C_\xi$ of C_ξ. As $\tau^i A = A$, A fills out $(1-\eta)$ of $\tau^i C_\xi$, and, hence, $(1-\eta)^2$ of Ω. As $\eta > 0$ was arbitrarily small $\mu(A) = 1$, and the ergodicity of τ follows.

An application of theorem 2.6 to this example now proves the theorem of Weyl [1916] on uniform distribution mod 1:

Theorem 2.7. *If τ in $\Omega = [0, 1[^d$ is the translation* mod 1 *by $\alpha = (\alpha_1, \ldots, \alpha_d)$ and the numbers $\alpha_1, \ldots, \alpha_d, 1$ are integrally independent, then, for each continuous f, $A_n f$ converges uniformly to $\int f d\mu$.*

Proof. It is simple to check that the $A_n f$ are equicontinuous. By the ergodicity the limit must be the constant $\int f d\mu$. □

Sometimes Weyl's theorem is spelled out for Riemann integrable functions, but this is really an equivalent formulation which can be obtained by a simple approximation. (To prove convergence a.e. for $f \in \mathscr{L}_1$ one needs a maximal inequality even in this special case).

Notes

Birkhoff [1931] has based his proof on the following (weaker) maximal inequality: If F_α is the τ-invariant set $\{\limsup A_n f \geqq \alpha\}$, then $\int_{F_\alpha} f d\mu \geqq \alpha\mu(F_\alpha)$.

He actually has formulated his a result only for indicator functions $f = 1_B$ (in the setting of a closed analytic manifold having a finite invariant measure). Khintchine [1933] then showed that Birkhoffs result remained true for integrable f on an abstract finite measure space. (Therefore theorem 2.3 is called Birkhoff-Khintchine-Theorem in a few countries. However, Khintchine himself emphasized that the idea of his proof was precisely that of Birkhoff).

Birkhoff's theorem (for finite μ) easily implies norm convergence of $A_n f$ in L_p for $f \in L_p$ $(1 \leq p < \infty)$, and therefore contains the special case of von Neumann's theorem which motivated his work. As Birkhoff's paper appeared earlier, it is of interest that von Neumann's theorem was proved first and was known to Birkhoff.

The original proof of Yosida-Kakutani's maximal ergodic theorem, which used ideas of Kolmogorov [1937], remains of interest; see Petersen [1983]. Simplified and alternative arguments (sometimes for continuous time or special cases) were given by Riesz [1931], [1932], [1942], [1945], Pitt [1942], Hopf [1947], Dowker [1950] and others. Kamae [1982] gave a proof using nonstandard analysis. Using some of his ideas, Katznelson and Weiss [1982] gave a short proof without explicit use of a maximal ergodic theorem.

E. Bishop [1966], [1967], [1968] gave a proof of Birkhoff's theorem using *upcrossing inequalities* similar to those in martingale theory; see § 4.1. These inequalities are constructively valid, but Birkhoff's theorem is not. Nuber [1972] has sought conditions which constructively imply the conclusion of Birkhoff's theorem.

The first *ratio ergodic theorem* was proved by Stepanov [1936] (in the ergodic case) and Hopf [1937]: If τ is an endomorphism of a σ-finite $(\Omega, \mathscr{A}, \mu)$, $f \in \mathscr{L}_1$ and $g \in \mathscr{L}_1^+$, then $S_n f / S_n g$ converges a.e. on each $\{S_k g > 0\}$. This is now a special case of the Chacon-Ornstein theorem.

It is easy to see that Birkhoff's theorem implies the convergence a.e. of $B_n^\lambda f := (2n+1)^{-1} \sum_{k=-n}^{+n} e^{2\pi i k \lambda} f \circ \tau^k$ for fixed $\lambda \in \mathbb{R}$. This was strengthened by Wiener and Wintner [1941] who proved:

Theorem 2.8. *If τ is an automorphism of a finite measure space, then there exists, for each $f \in \mathscr{L}_1$, a nullset N_f such that, for $\omega \in N_f^c$, $B_n^\lambda f(\omega)$ converges for all $\lambda \in \mathbb{R}$.*

Wiener and Wintner [1941a] also investigated when $F(t) = f(\tau_t \omega)$ is almost periodic and studied $\lim (2t)^{-1} \int_{-t}^{+t} F(s+u) \bar{F}(s)\, ds$, where $\bar{F}$ is the complex conjugate function.

Speed of convergence. When a convergence statement has been proved, one of the questions of interest is whether one can assert something about the speed of convergence. The famous law of the iterated logarithm is an example where a positive result is possible. In the form proved by Hartman and Wintner [1941] it can be stated in our language as follows: If $\mu(\Omega) = 1$, τ is an endomorphism, $Tf = f \circ \tau$, f has integral $\int f d\mu = 0$ and L_2-norm $\|f\|_2 = 1$, and if $f, f \circ \tau, f \circ \tau^2, \ldots$ are independent, then

$$\limsup_{n \to \infty} A_n f / \sqrt{2 \log \log n / n} = 1 \quad \mu\text{-a.e.}.$$

By symmetry $\liminf = -1$, so that $A_n f = 0\,((n^{-1} \log \log n)^{1/2})$ a.e. Independence is crucial! No general positive ergodic theoretic result of this type is possible even for slower speeds. Indeed, Krengel [1978a] has shown: If τ is an ergodic endomorphism of the torus $[0, 1[$ with Lebesgue measure, and (α_n) any null sequence of positive numbers, there exists a *continuous* f with integral 0 and

$$\limsup \alpha_n^{-1} |A_n f| = \infty \quad \text{a.e.}. \tag{2.2}$$

On the other hand, Halász [1976] proved: For any non decreasing sequence (c_n) of positive numbers with $c_1 \geqq 2$ and tending to ∞, and for any ergodic automorphism of $[0, 1[$ there exists A with $\lambda(A) = \frac{1}{2}$ and $|S_n 1_A - n/2| \leqq c_n$ for all n. Thus, the convergence *can* be arbitrarily fast.

The following deep theorem of O'Brien [1983] contains a limit version of Halász' result and (2.2) for measurable f:

Theorem 2.9. *If (b_n) is a sequence of positive numbers tending to ∞ and satisfying* $\liminf n^{-1} b_n = 0$, *there exists τ and a $\{+1, -1\}$-valued f with* $\limsup b_n^{-1} S_n f = 1$, *and this f can be constructed in such a way that the sequences $f, f \circ \tau, f \circ \tau^2, \ldots$* and $-f, -f \circ \tau, -f \circ \tau^2, \ldots$ *have the same joint distribution.*

Kakutani and Petersen [1981] proved another strengthening of (2.2) for measurable f: They construct, for any sequence (b_n) of positive numbers with divergent sum, a bounded measurable f with integral 0 and $\sup_k |\sum_{i=1}^{k} b_i A_i f| = \infty$ a.e. Dowker and Erdös [1959] showed the existence of a bounded f with $\int f d\lambda = 0$ for which $\sum_{i=1}^{k} b_i f \circ \tau^i$ fails to converge in measure on any subset of $[0,1[$ having positive measure, and with $\sup|\sum_{i=1}^{k} b_i f \circ \tau^i| = \infty$ a.e., see also Halmos [1948]. Baum and Katz [1965] showed: $\sum \mu(|A_n f| > \varepsilon)/n$ converges for f with $\int f d\mu = 0$ if $f, f \circ \tau \ldots$ are independent.

The uniform boundedness principle implies that the existence of a sequence α_n tending to ∞, with $\limsup \alpha_n \|A_n f - Pf\| < \infty$ for all f, is equivalent to $\|A_n - P\| \to 0$. It is therefore easy to see that there is no speed of convergence in von Neumann's mean ergodic theorem.

There are a few positive results on speed of convergence for specific transformations and functions; see e.g. Kuipers and Niederreiter [1974], or Kowada [1973].

Non-integrable functions. If τ is an ergodic endomorphism in a probability space, and a non negative f has a infinite integral, then Birkhoff's theorem implies $\lim_{n\to\infty} A_n f = \infty$, because

$$\liminf_{n\to\infty} A_n f \geqq \lim_{n\to\infty} A_n(f \wedge k) = \int (f \wedge k)\, d\mu \quad \text{for all } k.$$

J. Aaronson [1977] has proved that also other norming factors of $S_n f$ than n^{-1} cannot produce convergence a.e. to a non zero finite function:

Theorem 2.10. *If τ is an ergodic endomorphism of a probability space $(\Omega, \mathscr{A}, \mu)$, $f \geqq 0$, and $\int f d\mu = \infty$, then for any sequence (b_n) of positive reals, one has either*

$$\limsup_{n\to\infty} b_n^{-1} S_n f = \infty \quad \text{a.e.} \quad \textit{or} \tag{2.3}$$

$$\liminf_{n\to\infty} b_n^{-1} S_n f = 0 \quad \text{a.e.}. \tag{2.4}$$

There is a similar theorem for σ-finite measure spaces; see also Aaronson [1979].

Aaronson [1981] also considered this problem for functions f which need not be non negative: If b_n/n is non decreasing and tends to ∞, then (2.3) or (2.4) hold with $S_n f$ replaced by $|S_n f|$. In the case where the $f \circ \tau^i$ are independent Feller [1946] proved this assertion for arbitrary non decreasing sequences. But in the general ergodic case the condition $b_n/n \to \infty$ cannot be deleted: Aaronson [1977] gave an example of an f with $\int |f| d\mu = \infty$ and $S_n f/n \to 1$ a.e. A related result is that of Kesten [1975], who showed that $\liminf A_n f > 0$ a.e. if $S_n f \to \infty$ a.e..

Let (a_n) be a non decreasing sequence of positive numbers. Tanny [1974] showed that the condition $\liminf a_{k\cdot n}/a_n > 1$ for some $k > 1$ implies that for any ergodic τ in a probability space and all f, $\limsup f \circ \tau^n/a_n = \infty$ a.e. or $\limsup f \circ \tau^n/a_n = 0$ a.e. O'Brien [1982] proved that the condition is also necessary.

Dowker and Erdös [1959] have given several more examples: E.g. they showed that $S_n(\tau)g/S_n(\tau^{-1})g$ may diverge for ergodic automorphisms of a σ-finite measure space.

Del Junco and Steele [1977] have shown that for any ergodic endomorphism τ in $[0,1]$ (which Lebesgue measure) and for any increasing sequence $0 < b_1 \leqq b_2 \leqq$ of integers b_n with $n^{-1} b_n \to 0$ there exists an indicator function $f = 1_A$ such that $\limsup b_n^{-1} \sum_{i=n}^{n+b_n} f \circ \tau^i = 1$ μ-a.e. and the corresponding lim inf is 0 μ-a.e. (Wiener and Wintner [1941a] have made a similar observation; see also Pfaffelhuber [1975]).

By an example of Burkholder [1962] the averages considered in Birkhoff's theorem may diverge a.e. after an application of a single conditional expectation operator. Isaac [1973] has studied similar questions for ratios.

We refer to Kuipers and Niederreiter [1974] for the theory of uniform distribution mod 1.

For abstract ergodic theorems in a Boolean algebra and in a logic, see Bunjakov [1973], Dvurecenskij and Riecan [1980], and Pulmannová [1982].

§ 1.3 Recurrence

1. The conservative and dissipative part. We call a null preserving τ *recurrent* if, for all $A \in \mathscr{A}$, μ-almost all $\omega \in A$ belong to the set $A_{ret} = A \cap \bigcup_{k=1}^{\infty} \tau^{-k} A$ of points returning at least once, and *infinitely recurrent* if, for all $A \in \mathscr{A}$, μ-almost all $\omega \in A$ belong to the set

$$A_{inf} = \{\omega \in A: \tau^k \omega \in A \text{ for infinitely many } k \geqq 1\} = A \cap \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \tau^{-k} A.$$

A set $W \in \mathscr{A}$ is called *wandering* if the sets $\tau^{-k} W$ $(k \geq 0)$ are disjoint, or, equivalently, if no point in W returns to W. τ is called *conservative* if there exists no wandering set of positive measure. Finally, τ is called *incompressible*, if there exists no $A \in \mathscr{A}$ with $A \subset \tau^{-1} A$ and $\mu(\tau^{-1} A \setminus A) > 0$. Passing to complements one sees that τ is incompressible if and only if $\tau^{-1} B \subset B$ implies $\mu(B \setminus \tau^{-1} B) = 0$.

Theorem 3.1 (Recurrence theorem). *Let τ be null preserving in a measure space $(\Omega, \mathscr{A}, \mu)$. The following conditions are equivalent:*

(i) τ is conservative,

(ii) τ is recurrent,

(iii) τ is infinitely recurrent,

(iv) τ is incompressible.

Proof. (i) $\Rightarrow$ (ii): For any A the set $A_0 = A \setminus A_{ret}$ of points in A which never return is wandering, and, hence, a nullset.

(ii) $\Rightarrow$ (iii): For any $A \in \mathscr{A}$, $\mu(A_0) = 0$. The set $\tau^{-k} A_0$ is the set of points which visit A at time k for the last time. If $\omega \in A$ does not return to A infinitely often there must be some $k \geqq 0$ with $\omega \in \tau^{-k} A_0$. Hence $A_{inf} = A \setminus \bigcup_{k=0}^{\infty} \tau^{-k} A_0$ differs from A by a nullset.

(iii) $\Rightarrow$ (ii) is trivial.

(ii) $\Rightarrow$ (iv): If $\tau^{-1} B \subset B$, then the sequence $\tau^{-k} B$, $k = 1, 2, \ldots$ is decreasing, and $B \setminus \tau^{-1} B = B \setminus \bigcup_{k=1}^{\infty} \tau^{-k} B$ is the set of points in B which never return. As τ is recurrent $\mu(B \setminus \tau^{-1} B) = 0$.

(iv) $\Rightarrow$ (i): If W is wandering put $B = \bigcup_{k=0}^{\infty} \tau^{-k} W$. Clearly $\tau^{-1} B \subset B$. By the disjointness of the sets $\tau^{-k} W$ we have $W = B \setminus \tau^{-1} B$. Hence $\mu(W) = 0$. □

Clearly, if τ is an endomophism of a finite measure space, there cannot exist a wandering set of positive measure, and, therefore, τ must be infinitely recurrent. This is the recurrence theorem of Poincaré [1899], perhaps the oldest result in ergodic theory. Actually, Poincaré was most interested in a topological type of recurrence, which can be deduced from the result stated above: If Ω is a topological space with a countable basis $\mathscr{B} \subset \mathscr{A}$, and τ is infinitely recurrent, then almost every $\omega \in \Omega$ returns infinitely often to any neighborhood of itself. To see this, observe that the union E of all sets $A \setminus A_{inf}$ with $A \in \mathscr{B}$ has measure 0. If U is a neighborhood of some $\omega \in E^c$, there is an $A \in \mathscr{B}$ with $\omega \in A \subset U$. As $\omega \notin A \setminus A_{inf}$, ω returns infinitely often to A, and, hence, to U.

In general, if μ is σ-finite one can find a maximal subset of Ω on which τ is conservative.

Theorem 3.2 (Hopf decomposition). *If τ is null preserving in the σ-finite measure space $(\Omega, \mathscr{A}, \mu)$, there exists a decomposition of Ω into two disjoint measurable sets C and D, the conservative and the dissipative part, such that*

(i) C is τ-absorbing,

(ii) the restriction of τ to C is conservative, and

(iii) $D = \Omega \setminus C$ is an at most countable union of wandering sets.

If τ is even nonsingular, C is τ-invariant and there exists a wandering W_0 with $D = \bigcup_{k=-\infty}^{+\infty} \tau^k W_0$. ($\tau$ is called *dissipative* if $\Omega = D$.)

The proof is based on an *exhaustion argument*. As this type of argument is used frequently, let us explain this simple technique: Let **P** be a certain property of measurable sets in the σ-finite measure space $(\Omega, \mathscr{A}, \mu)$ which is such that any subset of a set with this property again has property **P**. As μ is σ-finite, there exists a finite measure ν equivalent to μ. Put $\alpha_1 = \sup\{\nu(A): A \text{ has property } \mathbf{P}\}$. Pick some $A_1 \in \mathscr{A}$ with property **P** and $\nu(A_1) \geqq 2^{-1}\alpha_1$. If $A_1, \ldots, A_n$ have been determined, put $\alpha_{n+1} = \sup\{\nu(A): A \text{ has property } \mathbf{P} \text{ and } A \subset \Omega \setminus \bigcup_{k=1}^{n} A_k\}$. Then find $A_{n+1} \subset \Omega \setminus \bigcup_{k=1}^{n} A_k$ with property **P** such that $\nu(A_{n+1}) \geqq 2^{-1}\alpha_{n+1}$. Let $\Omega_1 = \bigcup_{i=1}^{\infty} A_i$. The complement Ω_2 of Ω_1 is such that no subset of Ω_2 with positive measure has property **P**. Moreover, if Ω is the disjoint union of Ω_1' and Ω_2' such that Ω_1' is a countable union of sets with property **P**, and no subset of Ω_2' has positive measure and property **P**, then Ω_1 coincides up to nullsets with Ω_1'.

Proof of theorem 3.2. Take the property of being wandering and put $\Omega_1 = D$, Ω_2

$= C$. As $\tau^{-1} W$ is wandering if W is wandering we see that $\tau^{-1} D \subset D$ up to nullsets. Hence C is τ-absorbing mod μ. Changing C on a nullset we can assume that it is τ-absorbing.

If τ is nonsingular, also τW is wandering if W is, and we can infer that D and C are τ-invariant. W_0 is constructed as follows: Take the A_i from the construction of D, put $A_i^* = \bigcup_{k=-\infty}^{+\infty} \tau^k A_i$, and

$$W_0 = \bigcup_{i=1}^{\infty} (A_i \setminus \bigcup_{j=1}^{i-1} A_j^*). \quad \square$$

Also for general null preserving τ the description of D can be made somewhat more precise: Put $W_1 = A_1$, and for $n \geqq 1$ put

$$W_{n+1} = A_{n+1} \cup (W_n \cap (\bigcup_{i=0}^{\infty} \tau^{-i} A_{n+1})^c).$$

Then the sequence $D_n = \bigcup_{i=0}^{\infty} \tau^{-i} W_n$ is increasing with union D, and each W_n is wandering.

The following complement to the Hopf decomposition has been observed by J. Feldman [1962] in a more general situation. We leave it as an exercise.

Proposition 3.3. *Also in the case where τ is an endomorphism, C is τ-invariant.*

Halmos [1947] proved that the powers of a conservative τ are conservative. In fact we have:

Theorem 3.4. *If τ is null preserving in a σ-finite measure space $(\Omega, \mathscr{A}, \mu)$, the dissipative parts D^n of τ^n are the same for all $n \geqq 1$.*

Proof. If W is wandering for τ it is wandering for τ^n, so that $D^1 \subset D^n$. Now let $W_1, W_2, \ldots$ be wandering sets for τ^n with union D^n. Put $h = \sum_{k=1}^{\infty} 2^{-k} 1_{W_k}$. Then h is strictly positive on D^n and

$$\sum_{i=0}^{\infty} h \circ \tau^{in} = \sum_{k=1}^{\infty} 2^{-k} \sum_{i=0}^{\infty} 1_{W_k} \circ \tau^{in} \leqq \sum_{k=1}^{\infty} 2^{-k} \leqq 1 \text{ on } \Omega.$$

If D^n is not contained in $D^1 = D$, there exists an $\varepsilon > 0$ and $A \subset C$ with $\mu(A) > 0$ and $h \geqq \varepsilon 1_A$. Now

$$\sum_{i=0}^{\infty} 1_A \circ \tau^i \leqq \varepsilon^{-1} \sum_{i=0}^{\infty} h \circ \tau^i = \varepsilon^{-1} \sum_{j=0}^{n-1} (\sum_{i=0}^{\infty} h \circ \tau^{in}) \circ \tau^j \leqq \varepsilon^{-1} n \text{ on } \Omega.$$

But then no point in A can return to A more than $\varepsilon^{-1} n$ times, a contradiction to $A \subset C$. $\square$

The transformation $k \to k+1$ on $\Omega = \mathbb{Z}$ with the counting measure is an example of an ergodic measure preserving invertible transformation which is dissipative. This is essentially the only invertible example because of the last assertion in theorem 3.2. In particular a nonsingular ergodic τ in a non atomic measure space must be conservative. There exist endomorphisms in non atomic σ-finite measure spaces which are dissipative and yet ergodic; see § 3.1 (Notes).

Proposition 3.5. *If a null preserving transformation τ is conservative and A τ-absorbing* mod μ, *then A is τ-invariant* mod μ.

Proof. We may assume $\tau^{-1}A \supset A$ by modifying A on a nullset. Almost every point of $E = \tau^{-1}A \setminus A$ returns to E. But $\tau^{-1}A \supset A$ implies that, for all $\omega \in E$, the orbit $\tau\omega, \tau^2\omega, \ldots$ is contained in $A \subset E^c$. Thus E must have measure 0. □

It follows that a conservative null preserving transformation is ergodic if and only if $\mu(A) > 0$ implies $\mu((\bigcup_{k=0}^{\infty} \tau^{-k}A)^c) = 0$.

2. Induced transformations. If τ is recurrent, the *return time*

$$r_A(\omega) := \inf\{k \geqq 1 : \tau^k \omega \in A\}$$

to $A \in \mathscr{A}$ is finite a.e. in A. In the measure preserving case we can evaluate the integral of r_A:

Theorem 3.6 (Recurrence theorem of Kac [1947]). *If τ is a conservative endomorphism of $(\Omega, \mathscr{A}, \mu)$ and $A \in \mathscr{A}$, then*

$$\int_A r_A(\omega)\,\mu(d\omega) = \mu\Big(\bigcup_{n=0}^{\infty} \tau^{-n}A\Big).$$

In particular, if τ is ergodic and $\mu(A) > 0$, then $\int_A r_A\,d\mu = \mu(\Omega)$.

Proof. We may assume $0 < \mu(A) < \infty$. Put $A_0 = A$ and

$$A_k = \tau^{-1}A_{k-1} \cap A^c, \quad R_k = \tau^{-1}A_{k-1} \cap A \quad (k \geqq 1).$$

R_k is the set of points in A with $r_A(\omega) = k$. For $k \geqq 1$, A_k is the set of points in A^c which visit A for the first time at time k.

We have $\mu(A_k) = \mu(A_{k+1}) + \mu(R_{k+1})$, $(k \geqq 0)$, and, hence,

$$\mu(A_n) = \sum_{k=n+1}^{\infty} \mu(R_k) + \lim_{k\to\infty} \mu(A_k).$$

By the recurrence of τ, $\mu(A_0) = \sum_{k=1}^{\infty} \mu(R_k)$, which implies $\lim_{k\to\infty} \mu(A_k) = 0$. Now the assertion of the theorem follows from

$$\mu\Big(\bigcup_{n=0}^{\infty} \tau^{-n} A\Big) = \sum_{n=0}^{\infty} \mu(A_n) = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} \mu(R_k)$$
$$= \sum_{k=1}^{\infty} k\,\mu(R_k) = \int_A r_A \, d\mu. \quad \square$$

Higher moments of r_A have been studied by Blum and Rosenblatt [1967] and by Wolfowitz [1967].

For conservative null preserving τ, we can define a map τ_A of A_{ret} into A by $\omega \to \tau^{r_A(\omega)}\omega$. ω is mapped to the place where it first returns to A. Because of $\tau_A^{-1} B = \bigcup_{k=1}^{\infty} R_k \cap \tau^{-k} B$, this map is measurable and null preserving. It maps A_{inf} into A_{inf} and one has $\tau_{A_{inf}} = \tau_A$ on A_{inf}. Neglecting a nullset, we assume $A = A_{inf}$. τ_A is called the transformation *induced* by τ on A. Obviously, τ_A is conservative.

Theorem 3.7 (Kakutani [1943]). *If τ is a conservative endomorphism of $(\Omega, \mathscr{A}, \mu)$, the induced transformation is an endomorphism of $(A, A \cap \mathscr{A}, \mu_A)$, where μ_A is the restriction of μ to A.*

Proof. Take some measurable $B \subset A$, and put $B_0 = B$ and

$$B_{k+1} = \tau^{-1} B_k \cap A^c, \quad B'_{k+1} = \tau^{-1} B_k \cap A \quad (k \geqq 0).$$

We have

$$\mu(B) = \mu(B'_1) + \mu(B_1) = \mu(B'_1) + \mu(B'_2) + \mu(B_2) = \dots$$
$$= \sum_{k=1}^{\infty} \mu(B'_k) + \lim_{k\to\infty} \mu(B_k) \geqq \sum_{k=1}^{\infty} \mu(B'_k).$$

But B'_k is the set $R_k \cap \tau^{-k} B = R_k \cap \tau_A^{-1} B$. Thus $\mu(B) \geqq \mu(\tau_A^{-1} B)$. The theorem now follows by an application of the following lemma to τ_A. $\square$

Lemma 3.8. *If τ is conservative in $(\Omega, \mathscr{A}, \mu)$ and $\mu(\tau^{-1}A) \leqq \mu(A)$ holds for all $A \in \mathscr{A}$, then μ is τ-invariant.*

Proof. Otherwise there exists an A with $\mu(A) > \mu(\tau^{-1}A)$. Construct $A_0, A_1, \dots$ $R_1, R_2, \dots$ as in the proof of the Kac recurrence theorem. Then $\mu(A) > \mu(\tau^{-1}A) = \mu(A_1) + \mu(R_1) \geqq \mu(R_1) + \mu(R_2) + \mu(A_2) \ \dots \ \geqq \sum_{k=1}^{\infty} \mu(R_k) = \mu(A)$ yields a contradiction. $\square$

Proposition 3.9. *If a null preserving transformation τ in $(\Omega, \mathscr{A}, \mu)$ is conservative, and $\mathscr{I}$ is the σ-algebra of τ-invariant sets, then $\mathscr{I} \cap A$ is the σ-algebra of τ_A-invariant sets. In particular the ergodicity of τ implies that of τ_A.*

Proof. $I \in \mathscr{I}$ means that $\omega \in I$ is equivalent to $\tau\omega \in I$, and, hence, to $\tau^2\omega \in I$, etc. Therefore $\omega \in A \cap I$ is equivalent to $\tau_A\omega \in A \cap I$, and $A \cap I$ is τ_A-invariant. If, for some I', $A \cap I'$ is not τ_A-invariant, there exist an $\omega \in A$ such that $\omega \in I'$ and $\tau_A\omega \notin I'$, or $\omega \notin I'$ and $\tau_A\omega \in I'$. So the equivalence of the assertions $\tau^n\omega \in I'$ must fail for some n and I' is not τ-invariant. □

Sometimes it is useful to reverse the process of inducing a transformation on a subset: Let τ be an endomorphism of $(\Omega, \mathscr{A}, \mu)$ and let $R_1, R_2, \ldots$ be a *partition* of Ω, i.e., the R_i form a sequence of disjoint measurable sets with union Ω. Let $\Omega_0 = \Omega$, $\Omega_i = \bigcup_{k=i+1}^{\infty} R_k$, and $\Omega^* = \{(k, \omega) : k \geqq 0, \omega \in \Omega_k\}$, and for measurable $A \subset \Omega_k$ put $\mu^*(\{k\} \times A) = \mu(A)$. We have defined a measure space consisting of disjoint copies of the Ω_i. Define an endomorphism τ^* in Ω^* by

$$\tau^*(k, \omega) = \begin{cases} (k+1, \omega) & \text{if} \quad \omega \in \Omega_{k+1} \\ (0, \tau\omega) & \text{if} \quad \omega \notin \Omega_{k+1}. \end{cases}$$

If we identify Ω with $\{(0, \omega) : \omega \in \Omega_0\}$, then $\tau^*_\Omega = \tau$. This is Kakutani's *skyscraper* construction. Clearly one can do the same thing with a null preserving τ and come out with a null preserving τ^*. If τ is an automorphism of $(\Omega, \mathscr{A}, \mu)$, τ^* is again an automorphism. If τ is nonsingular, τ^* is. Clearly τ^* is ergodic if and only if τ is ergodic. Therefore the skyscraper construction is a very simple way of constructing ergodic automorphisms in non atomic infinite measure spaces. Simply start with an ergodic automorphism, say, in the unit interval, and choose the R_k with $\sum_{k=1}^{\infty} k \cdot \mu(R_k) = \infty$.

Notes

In his famous book, Hopf [1937] described the decomposition $\Omega = C \cup D$ for invertible transformations, and, in 1954, for the much more general situation of a positive contraction in L_1; see section 3.1. But concepts like wandering set or induced transformation have no equally simple and intuitive extensions to the operator case. Sucheston [1957], Helmberg [1965], [1965a], [1966], Simons [1965], [1971], Wright [1961], Roos [1964], Tsurumi [1958], Choksi [1961], Helmberg and Simons [1969], and others have therefore studied the case of non invertible null preserving transformations. One can split D further into the set of points in D whose orbit enter C and the rest, see Kopf [1982], [1978].

Call two measure spaces *isomorphic* mod 0 if – after deleting nullsets N and N' from Ω and Ω'-there exists a bijective map φ: $\Omega \setminus N \to \Omega' \setminus N'$ such that φ and φ^{-1} are measurable and $\mu \circ \varphi^{-1} = \mu'$. Endomorphisms τ and τ' in Ω and Ω' are called isomorphic mod 0 if such a φ can be found with $\tau' \circ \varphi = \varphi \circ \tau$; the sets N, N' shall be such that $\Omega \setminus N$ and $\Omega' \setminus N'$ are invariant under τ and τ' respectively. Two dissipative automorphisms τ and τ' in σ-finite spaces $(\Omega, \mathscr{A}, \mu)$ and $(\Omega', \mathscr{A}', \mu')$ are isomorphic mod 0 if and only if the measure spaces $(W_0, W_0 \cap \mathscr{A}, \mu_{W_0})$ and $(W_0', W_0' \cap \mathscr{A}', \mu'_{W_0'})$ are isomorphic mod 0, where W_0, W_0' are the sets with $\Omega = \bigcup_{k=-\infty}^{+\infty} \tau^k W_0$, $\Omega' = \bigcup_{k=-\infty}^{+\infty} \tau'^k W_0'$ constructed in theorem 3.2.

In particular the set W_0 is determined uniquely up to measure theoretic isomorphism mod 0; see Krengel [1968/69: part I], where also dissipative flows are classified by showing that they are isomorphic to a flow τ_s: $(\omega_0, t) \to (\omega_0, t+s)$ in a space $W_0 \times \mathbb{R}$, with a W_0 which is uniquely determined mod 0. The recurrence properties of nonsingular flows and null preserving semiflows were discussed in part II of this paper. Helmberg [1969] has derived a number of results on the mean recurrence time of measure preserving flows and semiflows, thereby giving a subtle continuous time version of the theorem of Kac.

If τ is an automorphism of a probability space $(\Omega, \mathscr{A}, \mu)$, and we put, for some $A \in \mathscr{A}$, $v_A(\omega) = r_A(\omega)$ $(\omega \in A)$, $v_A(\omega) = 0$ $(\omega \in A^c)$, then τ^{v_A} is again an automorphism. Neveu [1969] has studied the question, when, for a random variable v: $\Omega \to \mathbb{Z}^+$, τ^v is again an automorphism. It turns out that v must be such that τ^v is obtained by iterated compositions of induced transformations on a decreasing sequence of subsets of Ω.

Related results for more general groups of automorphisms were given by Geman and Horowitz [1975].

Jacobs [1967] has shown that the assertion of the Poincaré recurrence theorem remains valid if one replaces the stationarity assumption $\mu = \mu \circ \tau^{-1}$ by a recurrence property for the measure: Let μ be a probability measure on a complete, separable, metric space and let τ: $\Omega \to \Omega$ be continuous. If the orbit $\mu \circ \tau^{-1}, \mu \circ \tau^{-2}, \dots$ of μ returns into each neighborhood of μ (in the weak topology given by $\mu_n \to \mu$ if $\int f d\mu_n \to \int f d\mu$ for all bounded continuous f), then the orbit $\tau\omega, \tau^2\omega, \dots$ of μ-a.e. ω returns into each neighborhood of ω infinitely often.

Kurth [1975] considered homeomorphisms τ of a topological space Ω with countable basis and with a finite invariant measure, and gave a decomposition related to Poincarés theorem into departing points, asymptotic points and recurrent points.

Barone and Bhaskara Rao [1981] studied the recurrence theorem for *finitely* additive measures on a σ-algebra. Then a.e. point is k times recurrent for fixed k, but not necessarily infinitely recurrent.

The following result of Khintchine [1934] is proved in many books (and never applied): If τ is an automorphism of a probability space, then for $\varepsilon > 0$ and $B \in \mathscr{A}$ there exists L such that any interval of length L contains at least one k with $\mu(\tau^k B \cap B) \geqq \mu(B)^2 - \varepsilon$.

D. Maharam [1964] has "extended" nonsingular transformations τ of Ω to endomorphisms $\tilde{\tau}$ of $\Omega \times \mathbb{R}^+$ mapping the fiber of ω onto the fiber of $\tau\omega$. $\tilde{\tau}$ is conservative iff τ is conservative.

§ 1.4 Shift transformations and stationary processes

1. Canonical representation of processes. We now discuss a class of examples which is of great importance in probability theory.

Let $(E, \mathscr{F})$ be a measurable space and $J \neq \emptyset$ an arbitrary index set. The product space E^J is the space of all functions ω: $J \to E$. If ω_j is the value the function ω takes in $j \in J$, the map X_j: $\omega \to \omega_j$ is called the *j-th coordinate map* and $X_j(\omega) = \omega_j$ the j-th coordinate of ω. In the case $J = \mathbb{Z}^+$ we can identify ω with the *unilateral* sequence $(\omega_0, \omega_1, \omega_2, \dots)$, and in the case $J = \mathbb{Z}$ with the *bilateral* sequence $(\dots, \omega_{-1}, \omega_0, \omega_1, \omega_2, \dots)$.

An E-valued random variable Z defined on an abstract probability space $(\Omega', \mathscr{A}', P)$ is nothing but a measurable map Z: $\Omega' \to E$. A family $Y = \{Y_j, j \in J\}$

of E-valued random variables is called an E-valued *stochastic process* with *parameter space* J. We may write $Y_j = X_j \circ Y$, when Y is considered as a map $\Omega' \to E^J$. E is called *state space* of Y. The *product σ-algebra* $\mathscr{F}^J$ is the smallest σ-algebra in E^J containing all σ-algebras $X_j^{-1}\mathscr{F}$ $(j \in J)$. Y can be considered as a random variable $(\Omega', \mathscr{A}') \to (E^J, \mathscr{F}^J)$.

The *distribution* of Z is the measure $P \circ Z^{-1}$ on $\mathscr{F}$, and the distribution of Y is the probability measure $P \circ Y^{-1}$ on $\mathscr{F}^J$. For distinct $j_1, j_2, \ldots, j_n \in J$ and $J_0 = \{j_1, \ldots, j_n\}$, E^{J_0} can be identified with E^n. The map Y_{J_0}: $\omega \to (Y_{j_1}(\omega), \ldots, Y_{j_n}(\omega))$ is an E^n-valued random variable (for the σ-algebra $\mathscr{F}^n = \mathscr{F}^{J_0}$). The distribution of Y_{J_0} is called the *n-dimensional marginal distribution* corresponding to $\{j_1, \ldots, j_n\}$.

The distribution of Y is uniquely determined by the finitedimensional marginal distributions: $P \circ Y^{-1}$ is the unique measure on $\mathscr{F}^J$ such that

$$(P \circ Y^{-1})\Big(\bigcap_{j \in J_0} X_j^{-1} A_j\Big) = P\Big(\bigcap_{j \in J_0} Y_j^{-1} A_j\Big)$$

holds for all finite $J_0 \subset J$ and all choices of $A_j \in \mathscr{F}$.

If $\{\mu_j, j \in J\}$ is a family of probability measures in $(E, \mathscr{F})$ the *product measure* $\mu = \prod_{j \in J} \mu_j$ is the unique probability measure in E^J such that

$$\mu\Big(\bigcap_{j \in J_0} X_j^{-1} A_j\Big) = \prod_{j \in J_0} \mu_j(A_j)$$

holds for all J_0 and all choices of $A_j \in \mathscr{F}$. The random variables Y_j $(j \in J)$ are called *independent* if

$$P\Big(\bigcap_{j \in J_0} Y_j^{-1} A_j\Big) = \prod_{j \in J_0} P(Y_j^{-1} A_j)$$

holds for all J_0, and all A_j. Thus, the random variables Y_j $(j \in J)$ are independent iff $P \circ Y^{-1}$ is the product of the distributions $\mu_j = P \circ Y_j^{-1}$.

Now assume $J = \mathbb{Z}$ or $J = \mathbb{Z}^+$. The shift θ: $E^J \to E^J$ is the transformation defined by $X_k(\theta\omega) = X_{k+1}(\omega)$. The shift in $\Omega := E^{\mathbb{Z}}$ is often called *bilateral shift*. It is bijective and both θ and θ^{-1} are measurable with respect to $\mathscr{A} = \mathscr{F}^{\mathbb{Z}}$. The shift in $\Omega^+ := E^{\mathbb{Z}^+}$ is measurable with respect to $\mathscr{A} = \mathscr{F}^{\mathbb{Z}^+}$ and surjective, but not invertible. It is called *unilateral shift*. (It will be convenient to use some notation like θ, $\mathscr{A}$, X_k both in Ω and Ω^+. It should always be clear from the context, which interpretation applies).

If E is a topological space, the unilateral shift is continuous in the product topology, and the bilateral shift is even a homeomorphism.

The simplest way to define a θ-invariant measure in Ω is to take a product measure. It is easy to see that $\mu = \prod_{j \in \mathbb{Z}} \mu_j$ is θ-invariant iff all μ_j are identical. The automorphism θ in $(\Omega, \mathscr{A}, \mu)$ is then called a (bilateral) *Bernoulli shift*. Unilateral Bernoulli shifts are defined in the same way in Ω^+ with $\mu = \prod_{j \in \mathbb{Z}^+} \mu_j$ and identical μ_j.

Random variables $Y_j\,(j \in J)$ are called *identically distributed* if their distributions μ_j agree. Consequently, if $Y = (Y_0, Y_1, \dots)$ is a sequence of independent identically distributed random variables, the distribution $\mu = P \circ Y^{-1}$ of Y is the the shift invariant measure used for the definition of a Bernoulli shift.

Kolmogorov's *strong law of large numbers* asserts that $n^{-1} \sum_{k=0}^{n-1} Y_k$ converges *almost surely* (i.e., P-a.e.) to $E(Y_0) = \int Y_0\, dP$, when the Y_i are real valued, independent, identically distributed, integrable random variables. Let us see how this follows from Birkhoff's theorem. We need the ergodicity of Bernoulli shifts and show a bit more:

An endomorphism τ of a finite measure space $(\Sigma, \mathscr{B}, \nu)$ is called *mixing* if

$$\lim_{n \to \infty} \nu(A \cap \tau^{-n} B) = \nu(A)\, \nu(B)/\nu(\Sigma) \tag{4.1}$$

holds for all $A, B \in \mathscr{B}$. Mixing implies ergodicity because (4.1) cannot hold for $A = B$ when A is a τ-invariant set with $0 < \nu(A) < \nu(\Sigma)$. Let $\mathscr{A}(m, n)$ denote the σ-algebra which is generated by the σ-algebras $X_k^{-1} \mathscr{F}$ $(m \leqq k \leqq n, k \neq \pm\infty)$. If θ is a Bernoulli shift, $\mu(A \cap \theta^{-n} B) = \mu(A)\, \mu(B)$ holds for all $A \in \mathscr{A}(k_1, k_2)$, $B \in \mathscr{A}(j_1, j_2)$ with $k_1 \leqq k_2 < \infty$ and $j_1 \leqq j_2 < \infty$ as soon as $j_1 + n > k_2$, because μ is a product measure and $\theta^{-n} B \in \mathscr{A}(j_1 + n, j_2 + n)$. Thus

$$\lim_{n \to \infty} \mu(A \cap \theta^{-n} B) = \mu(A)\, \mu(B) \tag{4.2}$$

holds for all A, B in the union of the σ-algebras $\mathscr{A}(k, l)$ with $k, l \neq \pm\infty$. This union is dense in $\mathscr{A}$ in the metric $d(A, A') = \mu(A \Delta A')$. Therefore (4.2) holds for all $A, B \in \mathscr{A}$ and θ is mixing.

Now put $K_x = \{\omega \in \Omega^+ \colon \lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} X_0 \circ \theta^k(\omega) = E(Y_0)\}$ and $K_y = \{\omega' \in \Omega' \colon \lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} Y_k(\omega') = E(Y_0)\}$. Birkhoff's theorem implies $\mu(K_x) = 1$. $X_k = X_0 \circ \theta^k$ and $Y_k = X_k \circ Y$ yield $K_y = Y^{-1}(K_x)$, so that $P(K_y) = (P \circ Y^{-1})(K_x) = \mu(K_x) = 1$. But this is the assertion of Kolmogorov's strong law.

Here we have applied the ergodic theorem only to a function on Ω^+ depending on only one coordinate. But we could take any integrable function $f \colon \Omega^+ \to \mathbb{R}$, and then the sequence $f \circ \theta^k$ would in general not be a sequence of independent random variables. Thus, even in the special case of a Bernoulli shift the ergodic theorem is strictly more informative than the strong law. The main usefulness of the ergodic theorem, however, is due to the fact that there are many processes $Y = (Y_0, Y_1, \dots)$ for which $P \circ Y^{-1}$ is θ-invariant although the $Y_0, Y_1, \dots$ are not independent.

If $Y = (Y_0, Y_1, \dots)$ is an E-valued process, the distribution $\mu = P \circ Y^{-1}$ is θ-invariant iff $\mu = P \circ Y^{-1} \circ \theta^{-1} = P \circ (\theta \circ Y)^{-1}$. Now $\theta \circ Y$ is the process $(Y_1, Y_2, \dots)$ for which the time scale is shifted by one. It is therefore natural to introduce the following definition: An E-valued stochastic process $Y = (Y_j, j \in J)$ with $J = \mathbb{Z}$ or $J = \mathbb{Z}^+$ is called *stationary* if Y and $\theta \circ Y$ have the same distribution. As the distribution is determined by the probabilities of events of the type $\{Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n\}$, we have the equivalent definition: Y is called stationary if

$$(4.3) \qquad P(Y_{t_1} \in A_1, \ldots, Y_{t_n} \in A_n) = P(Y_{t_1+s} \in A_1, \ldots, Y_{t_n+s} \in A_n)$$

holds for all n, all choices of $A_1, \ldots, A_n \in \mathscr{F}$, all $t_1, \ldots, t_n \in \mathbb{Z}^+$ (or $\mathbb{Z}$), and for $s = 1$, (and hence for all s). Intuitively, a process is stationary if the random mechanism generating the process is invariant under a translation of the time scale.

For a process $Y = (Y_t, t \in J)$ with $J = \mathbb{R}$ or $J = \mathbb{R}^+$, observed continuously, one can proceed very similarly: For an additive semigroup J we define a family of shifts θ_t in the space E^J by $X_s(\theta_t \omega) = X_{t+s}(\omega)$, where X_s is the s-th coordinate. One can then call Y stationary if the distribution is θ_s-invariant for all $s \in J$. This is the case when (4.3) holds for all $t_1, \ldots, t_n \in J$ and all $s \in J$. (This time it is not sufficient to ask for the validity of (4.3) for a single s, because $\mathbb{R}$ is not a group generated by a single element).

Let us return to the case of discrete time. By the definition of stationarity a process is stationary if it generates an endomorphism. Conversely, an endomorphism generates many stationary processes:

Proposition 4.1. *If τ is an endomorphism of a probability space $(\Omega, \mathscr{A}, \mu)$, $(\tilde{E}, \tilde{\mathscr{F}})$ a measurable space, and $f: \Omega \to \tilde{E}$ measurable, then the sequence $(Z_i = f \circ \tau^i$, $(i \geqq 0))$ is a stationary process. Similarly, if τ is an automorphism, $(Z_i, (i \in \mathbb{Z}))$ is stationary.*

Proof. If $n \in \mathbb{N}$, $A_1, \ldots, A_n \in \tilde{\mathscr{F}}$ and $t_1, \ldots, t_n \in \mathbb{Z}^+$ (or $\mathbb{Z}$) are given, we have

$$\begin{aligned}
\mu(Z_{t_1} \in A_1, \ldots, Z_{t_n} \in A_n) &= \mu(f \circ \tau^{t_1} \in A_1, \ldots, f \circ \tau^{t_n} \in A_n) \\
&= \mu(\tau^{-1}\{f \circ \tau^{t_1} \in A_1, \ldots, f \circ \tau^{t_n} \in A_n\}) \\
&= \mu(f \circ \tau^{t_1+1} \in A_1, \ldots, f \circ \tau^{t_n+1} \in A_n) \\
&= \mu(Z_{t_1+1} \in A_1, \ldots, Z_{t_n+1} \in A_n). \quad \square
\end{aligned}$$

Corollary 4.2. *If $Y = (Y_i, i \in J)$ is an E-valued stationary process on $(\Omega', \mathscr{A}', P)$,* ($J = \mathbb{Z}$ or $\mathbb{Z}^+$), *and $f: E^J \to \tilde{E}$ measurable, then $\tilde{Y}_i = f \circ \theta^i \circ Y$ $(i \in J)$ defines an $\tilde{E}$-*valued stationary process. *(For $J = \mathbb{Z}^+$ this means that $\tilde{Y}_i = f(Y_i, Y_{i+1}, Y_{i+2}, \ldots)$ is stationary).*

Proof. Assume $J = \mathbb{Z}^+$. Apply proposition 4.1 with Ω^+ instead of Ω, and take $\tau = \theta$ and $\mu = P \circ Y^{-1}$. The stationarity of $Z = (Z_0, Z_1, \ldots)$ implies that $\mu \circ Z^{-1}$ is shift-invariant in $\tilde{\Omega}^+ = \tilde{E}^{\mathbb{Z}^+}$. We have $\mu \circ Z^{-1} = P \circ Y^{-1} \circ Z^{-1} = P \circ (Z \circ Y)^{-1}$. But $Z \circ Y$ is the process $\tilde{Y} = (\tilde{Y}_0, \tilde{Y}_1, \ldots)$. The same proof works for $J = \mathbb{Z}$. $\square$

For example, if the process $Y_0, Y_1, \ldots$ is real valued and stationary, the process

$\tilde{Y}_i$ defined by the *moving averages* $\tilde{Y}_i = M^{-1} \sum_{k=i}^{i+M-1} Y_k$, $M > 1$ fixed, is stationary. Clearly, the $\tilde{Y}_i$ will in general be dependent even if the Y_i are independent. For the probabilist, the object of primary interest is the stationary process, and not the measure preserving transformation. Therefore, it is of interest to express the ergodicity of θ in terms of the process:

If $Y = (Y_j, j \in \mathbb{Z}^+)$ is defined on $(\Omega', \mathscr{A}', P)$, call $B \in \mathscr{A}'$ *invariant* if there exists some $A \in \mathscr{A}$ (in Ω^+) such that

$$B = \{(Y_n, Y_{n+1}, Y_{n+2}, \dots) \in A\} \tag{4.4}$$

is true for all $n \geqq 0$. This is equivalent to the existence of an $A^* \in \mathscr{A}$ with $B = Y^{-1} A^*$ and $\theta^{-1} A^* = A^*$. One direction of this equivalence ist obvious: If such an A^* exists, we have $\theta^{-n} A^* = A^*$ and $B = Y^{-1} \theta^{-n} A^* = \{(Y_n, Y_{n+1}, \dots) \in A^*\}$ for all $n \geqq 0$. If A with (4.4) exists A must be θ-invariant mod μ, because $Y^{-1}(A \Delta \theta^{-1} A) = \{(Y_0, Y_1, \dots) \in A$ and $(Y_1, Y_2, \dots) \notin A\} \cup \{(Y_0, Y_1, \dots) \notin A$ and $(Y_1, Y_2, \dots) \notin A\}$ is empty. ($A = \theta^{-1} A$ need not hold!) In section 1.1 we have sketched how to find for A a θ-invariant A^2 with $\mu(A \Delta A^2) = 0$. It is an exercise to show that (4.4) implies $B = Y^{-1} A^2$, so that we can use A^2 for A^*. The invariant sets in Ω' for the process Y form a σ-algebra $\mathscr{I}'$, and we have $\mathscr{I}' = Y^{-1} \mathscr{I}$, where $\mathscr{I}$ is the σ-algebra of θ-invariant sets in Ω^+. We could also call a set $B' \in \mathscr{A}'$ invariant mod P if there exists an $A \in \mathscr{A}$ for which $P(B' \Delta \{(Y_n, Y_{n+1}, \dots) \in A\}) = 0$ is true for all n; but then, again, we could use the same arguments to show that these are just sets B' which differ from an invariant B on a set of P-measure 0.

Once we have a concept of an invariant set for a stationary process, we also have a concept of ergodicity. Y is called *ergodic* if any invariant set $B \in \mathscr{A}'$ satisfies $P(B) = 0$ or $P(B^c) = 0$.

Proposition 4.3. *If $Y = (Y_0, Y_1, \dots)$ is stationary and ergodic and $f\colon \Omega^+ \to \tilde{E}$ is measurable, then the process $\tilde{Y} = (\tilde{Y}_0, \tilde{Y}_1, \dots)$ defined by $\tilde{Y}_i = f(Y_i, Y_{i+1}, \dots)$ is ergodic.*

Proof. We have $\tilde{Y} = Z \circ Y$. If $\tilde{Y}$ is not ergodic there exists an invariant B for the $\tilde{Y}$-process such that $0 < P(B) < 1$. For B there exists a shift-invariant set $\tilde{A}$ in $\tilde{\Omega}^+$ with $B = \tilde{Y}^{-1} \tilde{A}$. Put $A = Z^{-1} \tilde{A}$. It is straightforward to check that the invariance of $\tilde{A}$ under the shift in $\tilde{\Omega}^+$ implies the θ-invariance of A. Therefore $B = Y^{-1} A$ is invariant for the Y-process, a contradiction to the ergodicity of Y. □

Birkhoff's ergodic theorem, spelled out for stationary processes instead of endomorphisms τ, now has the following form:

Theorem 4.4. *If $Y_0, Y_1, \dots$ is a stationary real valued process and Y_0 is integrable, then* $\lim_{n \to \infty} n^{-1} \sum_{k=0}^{n-1} Y_k = E(Y_0 \mid \mathscr{I}')$.

This follows in the same way from Birkhoff's theorem as Kolmogorov's strong law.

We have discussed stationarity for a process $Y = (Y_j, j \in \mathbb{Z}^+)$ defined on an arbitrary probability space. Now note that $X = (X_j, j \in \mathbb{Z}^+)$ is the identity on Ω^+. Therefore the distribution $\mu = P \circ Y^{-1}$ agrees with the distribution of X. Hence, all probability statements on Y can be expressed in terms of X. For this reason X is called the *canonical representation* of processes with distribution μ.

2. Remote σ-algebras. The σ-algebra $\mathscr{A}_\infty = \bigcap_{n=0}^{\infty} \mathscr{A}(n, \infty)$ in Ω^+ and in $\Omega = E^{\mathbb{Z}}$ is called the (right) *remote σ-algebra*. (Also the term *tail σ-algebra* is in use). In Ω we can also define the left remote σ-algebra $\mathscr{A}_{-\infty} = \bigcap_{n=0}^{\infty} \mathscr{A}(-\infty, -n)$. $\mathscr{A}'(m, n) = Y^{-1}\mathscr{A}(m, n)$ is the smallest σ-algebra in Ω' in which the random variables Y_j $(m \leqq j \leqq n, j \neq \pm\infty)$ are measurable.

$$\mathscr{A}'_\infty = \bigcap_{n=0}^{\infty} \mathscr{A}'(n, \infty)\,(= Y^{-1}\mathscr{A}_\infty) \quad \text{and} \quad \mathscr{A}'_{-\infty} = \bigcap_{n=0}^{\infty} \mathscr{A}'(-\infty, -n)$$

are the right and left remote σ-algebras of Y. Events like $\{\omega' \in \Omega': \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} Y_k(\omega') \geqq \alpha\}$ or $\{\omega' \in \Omega': Y_k(\omega') \in B_k$ for infinitely many $k\}$ belong to $\mathscr{A}'_\infty$.

The σ-algebra $\mathscr{I}$ in Ω^+ is contained in $\mathscr{A}_\infty$ because of $A \in \mathscr{I} \Rightarrow A \in \mathscr{A} = \mathscr{A}(0, \infty) \Rightarrow A = \theta^{-n}A \in \mathscr{A}(n, \infty)$, $(n \geqq 0)$. Similarly $\mathscr{I}' \subset \mathscr{A}'_\infty$ for processes with index set $\mathbb{Z}^+$.

We call a σ-algebra *trivial* (with respect to a given measure) if it contains only nullsets and their complements. Ergodicity means that the σ-algebra of invariant sets is trivial. Therefore a stationary process $Y = (Y_j, j \in \mathbb{Z}^+)$ for which $\mathscr{A}'_\infty$ is trivial must be ergodic. If ν is a measure on a σ-algebra $\mathscr{B}$ and $\mathscr{B}_0, \mathscr{B}_1$ are sub-σ-algebras, $\mathscr{B}_0 \subset \mathscr{B}_1 \pmod{\nu}$ shall mean that for all $B_0 \in \mathscr{B}_0$ there exists a $B_1 \in \mathscr{B}_1$ with $\nu(B_0 \Delta B_1) = 0$.

Proposition 4.5. *If $(Y_j, j \in \mathbb{Z})$ is a bilateral stationary process, we have $\mathscr{I}' \subset \mathscr{A}'_\infty$ (mod P) and $\mathscr{I} \subset \mathscr{A}_\infty$ (mod μ).*

Proof. Because of $\mathscr{I} \subset \mathscr{A} = \mathscr{A}(-\infty, +\infty)$ there exists for any $A \in \mathscr{I}$ and for any $\varepsilon > 0$ an $n_\varepsilon \in \mathbb{N}$ and an $A_\varepsilon \in \mathscr{A}(-n_\varepsilon, +\infty)$ with $\mu(A_\varepsilon \Delta A) < \varepsilon$. For all n, $A = \theta^{-n}A$ and $\mu = \mu \circ \theta^{-n}$ imply $\mu(A \Delta \theta^{-n} A_\varepsilon) < \varepsilon$. But $\theta^{-n}A_\varepsilon \in \mathscr{A}(-n_\varepsilon + n, \infty)$, so that A can be approximated arbitrarily well by sets in $\mathscr{A}(m, \infty)$ where m is arbitrarily large. This implies $A \in \mathscr{A}_\infty$ mod μ. Applying Y^{-1} also $\mathscr{I}' \subset \mathscr{A}'_\infty$ mod P follows. □

These observations yield an alternative proof of the ergodicity of Bernoulli shifts.

One simply has to apply *Kolmogorov's zero-one-law* which asserts that $\mathscr{A}'_\infty$ is trivial for any process ($Y_j, j \in \mathbb{Z}$ or $\mathbb{Z}^+$) consisting of independent random variables Y_j. The following theorem of Blackwell and Freedman [1964] shows that the triviality of the remote σ-algebra is, in fact, equivalent to an asymptotic independence condition:

Theorem 4.6. *If* $(Y_j, j \in \mathbb{Z}^+)$ *is any sequence of random variables defined on* $(\Omega', \mathscr{A}', P)$, *the remote* σ*-algebra* $\mathscr{A}'_\infty$ *is trivial if and only if each* $B \in \mathscr{A}'$ *satisfies*

$$(4.5) \qquad \lim_{n \to \infty} \sup_{A \in \mathscr{A}'(n,\infty)} |P(A \cap B) - P(A)\,P(B)| = 0.$$

Proof. For any $A \in \mathscr{A}'_\infty$ and any $\varepsilon > 0$ there exists an n_ε and an $A_\varepsilon \in \mathscr{A}(n_\varepsilon, \infty)$ with $P(A \Delta A_\varepsilon) < \varepsilon$. As A belongs to each $\mathscr{A}'(n,\infty)$ we can apply (4.5) to $B = A_\varepsilon$ to get $P(A \cap A_\varepsilon) = P(A)\,P(A_\varepsilon)$. Passing with ε to zero we arrive at $P(A \cap A) = P(A)^2$, so that $P(A) \in \{0,1\}$. Conversely, suppose $\mathscr{A}'_\infty$ is trivial and fix $B \in \mathscr{A}'$. By the backward martingale convergence theorem (see Doob [1953: thm. 4.2, 382]) the sequence $E_P(1_B | \mathscr{A}'(n,\infty))$ of conditional expectations converges a.e. and in L_1-norm to $E_P(1_B | \mathscr{A}'_\infty)$, which equals $P(B)$ because of the triviality of $\mathscr{A}'_\infty$. For any $A \in \mathscr{A}'(n,\infty)$,

$$\begin{aligned} |P(A \cap B) - P(A)\,P(B)| &= \left|\int_A (1_B - P(B))\,dP\right| \\ &= \left|\int_A (E_P(1_B | \mathscr{A}'(n,\infty)) - P(B))\,dP\right| \\ &\leqq \| E_P(1_B | \mathscr{A}'(n,\infty)) - P(B)\|_1 . \end{aligned}$$

Hence (4.5) follows. □

3. Recurrence times. Let $Y = (Y_j, j \in \mathbb{Z}^+)$ be a stationary process and let A' be of the form $A' = \{Y \in A\} = Y^{-1}A$. For example, $A' = \{Y_0 \in A_0\}$ or $A' = \{Y_0 > Y_1 + Y_2\}$. The set A may depend on infinitely many coordinates. We assume $P(A') > 0$.

Let $P_A(B') = P(A')^{-1} P(A' \cap B')$ denote the conditional probability of B' given $\{Y \in A\}$. $P_A \circ Y^{-1}$ is the normalized restriction of $\mu = P \circ Y^{-1}$ to A.

By the recurrence theorem μ-a.e. point of A belongs to the set A_{inf} of points $\omega \in A$ returning to A infinitely often under θ. We delete the P-nullset $Y^{-1}(A \setminus A_{inf})$ from Ω'. This does not affect the distribution of Y. Then, for all $\omega' \in A'$ there are infinitely many $n \in \mathbb{N}$ with $(Y_n(\omega'), Y_{n+1}(\omega'), \ldots) \in A$, so that the random variables $R_0 = 0$,

$$R_{i+1}(\omega') = \inf\{\mathrm{n} > R_i(\omega') : (Y_n(\omega'), Y_{n+1}(\omega'), \ldots) \in A\}, \quad (i \geqq 0)$$

are finite in A'. The *recurrence times* are given by $T_i = R_i - R_{i-1}$, $(i \geqq 1)$. If $r_A(\omega) = \inf\{n \geqq 1 : \theta^n \omega \in A\}$ is the return time studied in section 1.3 and $\omega = Y(\omega')$, then $R_i(\omega')$ is the time of the i-th return of ω to A. Formally, we have $T_i(\omega')$

$= r_A(\theta^{R_{i-1}(\omega')}\omega)$. Therefore $(Y_{R_i}, Y_{R_i+1}, \ldots)(\omega')$ is the point $\theta_A^i(\omega)$, where θ_A is the transformation induced by the shift on A. Combining Theorem 3.6, proposition 3.8, corollary 4.2 and proposition 4.3 we have proved:

Theorem 4.7. *Let $Y = (Y_i, i \geqq 0)$ be an E-valued stationary process, and let f be a measurable map from Ω^+ into $\tilde{E}$. Assume $A \in \mathscr{A}$ and $\mu(A) > 0$. Then the process*

$$U_i = f(Y_{R_i}, Y_{R_i+1}, \ldots) \quad (i \geqq 0)$$

on $(A', A' \cap \mathscr{A}', P_A)$ is stationary. The process $(U_i, i \geqq 0)$ is ergodic if Y is ergodic. In particular the processes $T = (T_i, i \geqq 1)$ and $V = (V_i, i \geqq 0)$ with $V_i = Y_{R_i}$ are stationary, and they are ergodic if Y is ergodic.

It is clear that an analogous theorem holds for bilateral processes.

4. Bilateral extensions of unilateral processes. In a way, bilateral stationary processes are simpler than unilateral processes because the bilateral shift is invertible. It is therefore of interest that in all probabilistic questions on stationary processes we may assume that the process is bilateral if E satisfies very mild regularity assumptions.

Call $(E, \mathscr{F})$ a *Borel space* if there exists a bijective map ψ of E onto a Borel subset $E' \subset \mathbb{R}^1$ which is measurable in both directions. E.g., *Polish spaces* (i.e., complete separable metric spaces) and Borel subsets of such spaces with their Borel σ-algebra are Borel spaces.

Theorem 4.8. *If $Y = (Y_j, j \in \mathbb{Z}^+)$ is a unilateral stationary process defined on a probability space $(\Omega', \mathscr{A}', P)$ and taking values in a Borel space $(E, \mathscr{F})$, then there exists a θ-invariant probability measure μ on $\Omega = E^{\mathbb{Z}}$ such that the distribution of $(X_j, j \geqq 0)$ in $(\Omega, \mathscr{A}, \mu)$ agrees with that of Y.*

Proof. For any set $A \in \mathscr{A}(-n, +n)$ there exists a measurable subset $A^{(n)}$ of E^{2n+1} such that $A = \{\omega \in \Omega: (X_{-n}(\omega), \ldots, X_n(\omega)) \in A^{(n)}\}$. If we put $\mu(A) = P(\{\omega' \in \Omega': (Y_{k-n}(\omega'), \ldots, Y_{k+n}(\omega')) \in A^{(n)}\})$, then the stationarity of Y implies that this definition is independent of k as long as $k \geqq n$, and also that the same number is assigned to $\mu(A)$ if A is considered as an element of the σ-algebra $\mathscr{A}(-(n+1), n+1)$. Thus these definitions of the marginal distributions of $(X_{-n}, \ldots, X_n)$ are consistent. An application of the Daniell-Kolmogorov extension theorem (see e.g. Jacobs [1978]), completes the proof. □

It is a simple consequence of proposition 4.5 that the ergodicity of a unilateral stationary process is equivalent to the ergodicity of its bilateral extension.

5. Further examples of stationary processes. A class of examples of great interest in probability theory and in prediction theory is provided by the stationary