

RECENT PROGRESS IN ERGODIC THEORY¹

PAUL R. HALMOS

Prologue. In 1948, at the November meeting of the Society in Chicago, I delivered an address entitled *Measurable transformations*. In the twelve years that have elapsed since then, ergodic theory (of which the theory of measurable transformations is the greatest part) has been spectacularly active. The purpose of today's address is to report some of the developments of those twelve years; its title might well have been *Measurable transformations revisited*. The subjects I chose for this purpose are: some new ergodic theorems, information theory and its connection with ergodic theory, and the problem of invariant measure.

The stage on which most ergodic performances take place is a measure space consisting of a set X and of a measure μ defined on a specified σ -field of measurable subsets of X . At the most trivial level X consists of a finite number of points, every subset of X is measurable, and μ is a mass distribution on X (which may or may not be uniform). At a more useful and typical level X is the real line $(-\infty, +\infty)$, or the unit interval $[0, 1]$, measurability in either case is interpreted in the sense of Borel, and μ is Lebesgue measure. Another possibility is to consider a measure space having a finite number of points with total measure 1 and to let X be the Cartesian product of a countably infinite number of copies of that space with itself; measurability and measure in this case are interpreted in the customary sense appropriate to product spaces. This latter example is easily seen to be measure-theoretically isomorphic to the unit interval, as also are most of the normalized measure spaces (measure spaces with total measure 1) that ever occur in honest analysis. The only measure spaces I shall consider in this report are the ones isomorphic to one of the spaces already mentioned. The expert will know just how little generality is lost thereby, and the casual passer-by, quite properly, will not care.

A transformation T from a measure space X into a measure space Y is called *measurable* if the inverse image $T^{-1}E$ (in X) of each measurable set E (in Y) is again a measurable set. A measurable transformation T is *measure-preserving* if, for every measurable set E , the sets E and $T^{-1}E$ have the same measure. A measurable (but not

An address delivered before the Summer Meeting of the Society in East Lansing on September 1, 1960, by invitation of the Committee to Select Hour Speakers for Summer and Annual Meetings; received by the editors September 1, 1960.

¹ The work was done while the author had a National Science Foundation grant.

necessarily measure-preserving) transformation $T: X \rightarrow Y$ is *invertible* if there exists a (necessarily unique) measurable transformation $T^{-1}: Y \rightarrow X$ such that each of the composites $T^{-1}T$ and TT^{-1} is equal to the identity on its domain. If T is measure-preserving and invertible, then T^{-1} is measure-preserving also. Most of the transformations to be considered in this report are transformations from a measure space X into itself, or, in the customary phrase, transformations on X .

A measure-preserving transformation on a finite set with uniform distribution is simply a permutation. A familiar example of a measure-preserving transformation on the real line is given by a translation, $Tx = x + c$; the same equation, interpreted modulo 1, gives an interesting example of a measure-preserving transformation on the unit interval. An important transformation on the product of countably many copies of a finite normalized measure space is obtained as follows. Let the index set used in the formation of the Cartesian product be the set of all integers (positive, negative, or zero), so that a point x of the space X under consideration is a two-way infinite sequence (x_n) ; for each x in X let Tx be the sequence y such that $y_n = x_{n+1}$. Any transformation defined this way is called a *shift*, or, more precisely, the shift based on the given finite measure space. In case that space has k points and the original measure on it is the uniform distribution, the resulting shift is completely determined by the number k . The shift so determined is called the *k-shift*; in what follows it will be denoted by S_k .

Ergodic theorems. Every transformation T of a set X into itself induces a functional operator U that acts on functions whose domain is X ; by definition $(Uf)(x) = f(Tx)$. The transformations that always have been and still are of central interest in ergodic theory are the measure-preserving transformations. The restriction to any of the Lebesgue spaces L_p of the functional operator U associated with a measure-preserving transformation T turns out to be an isometry; the preservation of norm is an easy consequence of the preservation of measure. If T is invertible, then U is invertible. On L_2 , the pleasantest of all function spaces, an invertible isometry is a unitary operator, to which the extensive and powerful spectral theory can be applied. The first result of such an application is von Neumann's mean ergodic theorem; it studies the norm convergence in L_2 of averages such as $(1/n) \sum_{i=0}^{n-1} U^i f$. The methods of the study extend immediately to all unitary operators on L_2 , including the ones that are not induced by measure-preserving transformations. This fact motivated the subse-

quent extension of the theory to wider and wider classes of transformations on wider and wider classes of Banach spaces.

The finite special case of von Neumann's mean ergodic theorem is amusing and instructive. If, for instance, $X = \{1, 2, 3\}$, then L_2 (over X) is three-dimensional Euclidean space; if T is the cyclic permutation (1, 2, 3), then the induced unitary operator U is the one whose matrix (with respect to the usual and obvious basis) is

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

The original mean ergodic theorem asserts the convergence of the averages of the powers of such permutation matrices. The first and most natural generalization is to the convergence of the averages of the powers of arbitrary unitary matrices.

The so-called individual ergodic theorem of G. D. Birkhoff has the function space L_1 for its natural habitat; the assertion is that for every integrable function f the averages $(1/n) \sum_{i=0}^{n-1} U^i f$ converge almost everywhere. It is natural to try to extend this result to functional operators that may not be induced by measure-preserving transformations. The spirit of the general mean ergodic theorems and the spirit of the generalization desired here are quite different. The former have a wider domain of applicability (Banach spaces) with a necessarily weaker conclusion (norm convergence); the latter has a sharp conclusion (almost everywhere convergence) for a necessarily more special structure (the function space L_1). Recent progress of the latter kind (based on the pioneering work of Doob, Kakutani, and others) was made by Eberhard Hopf. He assumes that $\mu(X) < \infty$, and that U is a positive linear operator on L_1 subject to the following two conditions:

$$\int (Uf) d\mu = \int f d\mu$$

for all f in L_1 , and

$$U1 = 1.$$

(Positiveness means, of course, that if f is in L_1 and $f \geq 0$ a.e., then $Uf \geq 0$ a.e.) The conclusion is that for each f in L_1 the averages $(1/n) \sum_{i=0}^{n-1} U^i f$ converge almost everywhere. It is obvious that if U is the functional operator induced by a measure-preserving transformation, then Hopf's conditions are satisfied; the assertion is, in other words, a bona fide generalization of the classical individual ergodic theorem.

A glance at the finite special case will clarify the meaning of Hopf's conditions. In the finite case, the linear transformation U corresponding to a matrix (u_{ij}) is positive if and only if the entries of the matrix are positive ($u_{ij} \geq 0$). It is a simple and pleasant exercise to prove that, in the finite case, the first Hopf condition requires exactly that each column sum ($\sum_i u_{ij}$) be equal to 1, and the second one that each row sum ($\sum_j u_{ij}$) be equal to 1. The extent to which Hopf generalizes Birkhoff is now clear. Birkhoff treats matrices with exactly one 1 in each row and in each column, all other entries being 0, and Hopf treats matrices with positive entries whose row sums and column sums are equal to unity.

The theorem of Hopf was generalized by Dunford and Schwartz in one direction and by Chacon and Ornstein in another. The assumption that $\mu(X) < \infty$ is absent from both generalizations. The Dunford-Schwartz assumptions are that

$$\|U\|_1 \leq 1 \quad \text{and} \quad \|U\|_\infty \leq 1.$$

The first of these conditions is clear. It requires that the norm of the operator U on L_1 do not exceed 1, i.e., that $\int |Uf| d\mu \leq \int |f| d\mu$ for every integrable function f . The second condition requires that U map each essentially bounded function in L_1 onto an essentially bounded function, without increasing the essential supremum. It is noteworthy that Dunford and Schwartz do not assume that U is positive. The conclusion is, as before, that for each f in L_1 (or, for that matter, in any L_p , with $1 \leq p < \infty$) the averages $(1/n) \sum_{i=0}^{n-1} U^i f$ are almost everywhere convergent.

The L_1 norm of a finite matrix (u_{ij}) is the largest of the absolute column sums ($\sum_i |u_{ij}|$) and the L_∞ norm is the largest of the absolute row sums ($\sum_j |u_{ij}|$). This observation indicates the extent to which the Dunford-Schwartz result generalizes Hopf's: positiveness is dropped and the equations involving column sums and row sums become inequalities.

Chacon and Ornstein assume, as Hopf does, that U is a positive linear operator on L_1 , and they assume, as Dunford and Schwartz do, that $\|U\|_1 \leq 1$. Instead of adding further assumptions, they alter the form of the conclusion. The new conclusion asserts that if f and g are in L_1 and if $g \geq 0$ almost everywhere, then the ratios

$$\sum_{i=0}^{n-1} U^i f \bigg/ \sum_{i=0}^{n-1} U^i g$$

converge at almost every one of the points at which the denominator is ultimately different from zero. The Chacon-Ornstein result implies, in particular, that if for some g (for instance $g=1$) the

averages $(1/n) \sum_{i=0}^{n-1} U^i g$ converge to 1 almost everywhere, then the averages $(1/n) \sum_{i=0}^{n-1} U^i f$ converge almost everywhere; this exhibits the extent to which their result generalizes Hopf's.

In the finite case, the coordinates of the vector $U^n 1$ are the row sums of the matrix corresponding to U^n . Accordingly, a special case of the Chacon-Ornstein theorem applies to matrices with positive entries, column sums equal to 1, and row sums converging to 1 in the sense of Cesàro. The last qualification replaces Hopf's requirement that the column sums be equal to 1.

Information theory. If $\mu(X) = 1$, the measurable subsets of the space X may be viewed as the possible outcomes of some random process, and, from this point of view, the measure of a measurable set E becomes the probability that a randomly chosen point of X belong to E . It is well known that all the intuitive concepts of probability theory can be given a rigorous description in measure-theoretic language. Thus, for instance, an experiment with a finite number of possible outcomes is (corresponds to) a finite measurable partition \mathfrak{A} . (This means, of course, a finite disjoint collection of measurable sets whose union is X .)

How much information does an experimenter obtain from an experiment? If, to be slightly more precise, the random point of X that the experiment specifies belongs to the measurable set E , how much more is known after that specification than before? There are several heuristic arguments that tend to show that a reasonable measure of the quantity of information is $-\log \mu(E)$. It follows that associated with each finite measurable partition \mathfrak{A} there is a simple function $I(\mathfrak{A})$, which may be called the information conveyed by the performance of the experiment \mathfrak{A} ; the value of $I(\mathfrak{A})$ throughout each set E of \mathfrak{A} is $-\log \mu(E)$.

The formation of the successive iterates of some particular measure-preserving transformation T may be regarded as the action of the passage of time observed at equally spaced intervals, say once a day. Thus if " $x \in E$ " is read as " x belongs to E today," then " $Tx \in E$ " may be read as " x will belong to E tomorrow." Since " $Tx \in E$ " means the same as " $x \in T^{-1}E$," it follows that if the mathematical model of some event today is the measurable set E , then the model of the same event tomorrow will be the measurable set $T^{-1}E$.

Suppose now that \mathfrak{A}_0 is an experiment (with a finite number of possible outcomes) and suppose that the experiment \mathfrak{A}_1 consists of the performance of \mathfrak{A}_0 both today and tomorrow. In mathematical language, the partition \mathfrak{A}_1 is the least common refinement of \mathfrak{A}_0 and $T^{-1}\mathfrak{A}_0$, where the sets of $T^{-1}\mathfrak{A}_0$ are the sets of the form $T^{-1}E$ with E

in \mathfrak{A}_0 . In the standard concise lattice-theoretic notation $\mathfrak{A}_1 = \mathfrak{A}_0 \vee T^{-1}\mathfrak{A}_0$. If, more generally,

$$\mathfrak{A}_n = \mathfrak{A}_0 \vee \dots \vee T^{-(n-1)}\mathfrak{A}_0, \quad n = 1, 2, 3, \dots,$$

then \mathfrak{A}_n is (corresponds to) the experiment that consists of performing \mathfrak{A}_0 for n successive days starting today.

What can be said about the average amount of information obtained in a day by repeated performances of \mathfrak{A}_0 ? In mathematical terms the question is one about the asymptotic behavior of the sequence of functions $(1/n)I(\mathfrak{A}_n)$. The question appears to call for, and the answer is, a theorem of ergodic type; the proof, indeed, makes use of the classical individual ergodic theorem. Motivated by Shannon's work on Markov chains, McMillan proved that the sequence $(1/n)I(\mathfrak{A}_n)$ converges in the mean of order 1, and, later, Breiman asserted that it converges almost everywhere. McMillan and Breiman treat ergodic transformations only; an elegant simple proof of the general theorem (whose spirit is substantially the same as that of Breiman's proof) was recently given by Alexandra Ionescu-Tulcea. (Recall that T is ergodic in case, for each measurable set E that is invariant under T , either $\mu(E) = 0$ or $\mu(X - E) = 0$.)

In the present state of science the information ergodic theorem of Shannon-McMillan-Breiman looks more like a special consequence of ergodic methods and results than like a statement of comparable generality with the ergodic theorem. The special theorem has, however, some powerful applications, and, I think, a promising future.

The conjugacy problem. The main outstanding problem of ergodic theory was (and is) to find usable necessary and sufficient conditions for the conjugacy of two transformations. To say that two measure-preserving transformations T_1 and T_2 are conjugate means that there exists an invertible measure-preserving transformation S such that $ST_1 = T_2S$ modulo sets of measure zero. A long outstanding test problem was the conjugacy of the 2-shift and the 3-shift (defined above, in the prologue). The first decisive recent step along these lines was taken by Kolmogorov and the next one by Sinai. The conclusion is that if $n \neq m$, then S_n and S_m are not conjugate.

The Kolmogorov-Sinai method is to introduce a new conjugacy invariant $h^*(T)$ (called *entropy*) associated with a measure-preserving transformation T on a normalized measure space. The invariant $h^*(T)$ can be computed in many cases, and, in particular, it turns out that $h^*(S_n) = \log n$.

The function h^* is defined in two steps, via two other functions H and h (both of which are also called entropy). The domain of the first

auxiliary function H consists of finite measurable partitions. If \mathfrak{A} is such a partition, then, by definition, $H(\mathfrak{A}) = - \sum_{E \in \mathfrak{A}} \mu(E) \log \mu(E)$, or, in terms of the notation used before, $H(\mathfrak{A}) = \int I(\mathfrak{A}) d\mu$. The number $H(\mathfrak{A})$ is, intuitively speaking, the average amount of information obtainable from one performance of the experiment \mathfrak{A} .

It is a trivial consequence of McMillan's theorem that, for each finite measurable partition \mathfrak{A} the sequence of numbers

$$\frac{1}{n} H(\mathfrak{A} \vee \dots \vee T^{-(n-1)}\mathfrak{A})$$

converges to a finite limit, say $h(T, \mathfrak{A})$. The number $h(T, \mathfrak{A})$ is, intuitively speaking, the average amount of information obtainable *per day* from repeated performances of the experiment.

Now that h has been defined, the definition of h^* is a simple matter; write $h^*(T) = \sup_{\mathfrak{A}} h(T, \mathfrak{A})$, where the supremum is extended over all finite measurable partitions \mathfrak{A} . Intuitively speaking, the entropy $h^*(T)$ is the greatest quantity of information obtainable about the universe per day by repeated performances of experiments with a finite (but possibly unbounded) number of possible outcomes. The value of $h^*(T)$ is a real number between 0 and ∞ ; both extremes can occur.

The computation of the supremum $h^*(T)$ is often made easy by a penetrating result of Sinai. The assertion is that if T is invertible and if \mathfrak{A} is a finite measurable partition such that every measurable set belongs to the σ -field generated by the elements of the partitions $T^i\mathfrak{A}$ ($i = 0, \pm 1, \pm 2, \dots$), then the supremum $h^*(T)$ is attained, and, in fact, it is equal to $h(T, \mathfrak{A})$. In intuitive language: if knowledge of the entire history of performances of the experiment \mathfrak{A} (past, present, and future) entails knowledge of the exact state of the universe, then the average amount of information obtainable per day by repeated performances of \mathfrak{A} is as large as it can be for any experiment.

Having shown that the concept of entropy can be used to solve a problem of ergodic theory, the Russian school is proceeding with enthusiasm to build a theory around the theorem. Sample question: what is the relation between the entropy of T and the spectral properties of the induced unitary operator U ? Partial answer by Rokhlin: if U has pure point spectrum, then $h^*(T) = 0$. Sample question: if T is an automorphism of a compact group, what can be said about the entropy of T ? Partial answer by Sinai: if T is an automorphism of the torus, then T is in a sense the dual of a unimodular matrix M ; if α and β are the proper values of M , with $|\alpha| > |\beta|$, then $h^*(T) = \log |\alpha|$. Partial answer by Abramov: if T is an automorphism of

the character group of the additive group of rational numbers, then T is in a sense the dual of a rational multiplier r ; if α and β are (not necessarily respectively) the numerator and denominator of r in lowest form, with $|\alpha| > |\beta|$, then $h^*(T) = \log |\alpha|$.

In the early days of ergodic theory it was possible to conjecture that two invertible measure-preserving transformations on normalized measure spaces are conjugate if and only if the functional operators they induce are unitarily equivalent. That conjecture died long ago; the Kolmogorov-Sinai result buries it forever. The answer to the following question is, however, not at all obvious. If two invertible measure-preserving transformations on normalized measure spaces induce unitarily equivalent functional operators and if they have the same entropy, does it follow that they are conjugate? (This question was raised by Fomin.) It is instructive to examine it in the light of the following specific subquestion. If two shifts based on finite sets with not necessarily uniform distributions have the same entropy, does it follow that they are conjugate? (The subquestion was raised by Billingsley.) An affirmative answer to Fomin's question would imply an affirmative answer to Billingsley's. (The functional operators induced by any two shifts are unitarily equivalent.) Following Billingsley, I conjecture that the truth is the other way around; the answer to Billingsley's question is probably no.

Non-invariant measures. The preceding considerations had to do mostly with measure-preserving transformations. How much loss of generality does this involve? How likely is a transformation to preserve some measure? There are several possible ways to formulate a precise question along these lines; experimentation has revealed that the most fruitful formulation runs as follows. Suppose that T is an invertible measurable transformation on, say, the unit interval (there is no essential loss of generality so far) such that if E is a measurable set of measure zero, then both $T^{-1}E$ and TE have measure zero; does there exist a finite or possibly σ -finite measure ν equivalent to μ and invariant under T ? (Equivalence here means that μ and ν vanish on the same sets.) Several solutions have been offered for the problem of finite invariant measure in the course of the years, and, incidentally, it has been known for some time that the solution of that problem is not always affirmative.

The most recent solution of the problem of finite invariant measure is that of Hajian and Kakutani. The condition in terms of which that solution is stated is an elegant generalization of a well known condition. A set E is a *wandering set* (for a transformation T) if the sets $E, T^{-1}E, T^{-2}E, \dots$ are pairwise disjoint, and the transformation T

is *conservative* if every measurable wandering set has measure zero. These definitions are old; the new concepts of Hajian and Kakutani are defined as follows. A set E is a *weakly wandering set* (for T) if the sequence of sets $E, T^{-1}E, T^{-2}E, \dots$ has an infinite pairwise disjoint subsequence, and T is *weakly conservative* if every measurable weakly wandering set has measure zero. (Hajian and Kakutani do not actually use the latter term, but it seems natural to do so.) The new result is that a transformation T (satisfying the conditions stated before) preserves some finite invariant measure equivalent to the given one if and only if it is weakly conservative.

The general problem of (σ -finite) invariant measure has stood unsolved for a long time. There were, to be sure, some known necessary and sufficient conditions, but none of them was usable. In particular, for all anyone knew, the solution of the problem was always affirmative. The solution, when it finally came, turned out to be negative; Ornstein constructed an ingenious example of a transformation that satisfies all the stated conditions but does not preserve any equivalent σ -finite measure.

Ornstein defines his transformation on the unit interval; there is some value in an alternative approach via another measure space, which, however, is easily shown to be isomorphic to the interval. The construction uses a sequence of integers m_0, m_1, m_2, \dots , such that $m_k > 1$ for all k , and such that the quotients $m_{k+1}/m_0 \cdots m_k$ are unbounded. (Example: $m_k = 2^{3^k}$. In this case the quotients even tend to ∞ with k .) These conditions are sufficient to guarantee the truth of the assertions that follow. Since, however, those assertions are not proved here, it is not necessary to keep the exact form of the conditions in mind.

Let X_k be the additive group of integers modulo m_k ($k = 0, 1, 2, \dots$), and let μ_k be the (non-uniform) measure in X_k that assigns the weight $1/2$ to 0 and distributes the weight $1/2$ uniformly among the remaining points. Let X be the Cartesian product of the spaces X_k and let μ be the product of the measures μ_k . For an intuitive insight into the role that the elements of X play, think of them as generalized (infinite) integers written in a generalized (infinite) decimal system. An ordinary non-negative integer x has an expansion

$$x = x_0 + x_1 \cdot 10 + x_2 \cdot 10^2 + x_3 \cdot 10^3 + \dots,$$

where the digits x_k can take the values $0, 1, \dots, 9$, and where $x_k \neq 0$ for finitely many values of k only. The generalization replaces the k th 10 by m_k , and, at the same time, removes the finitely-non-zero restriction. The result is the consideration of formally infinite expan-

sions of the form

$$x_0 + x_1 \cdot m_0 + x_2 \cdot m_0 m_1 + x_3 \cdot m_0 m_1 m_2 + \cdots,$$

where the digits x_k can take the values $0, 1, \dots, m_k - 1$.

What happens to the sequence (x_0, x_1, x_2, \dots) of decimal digits of an integer x when x is replaced by $x+1$? The answer is that, usually, x_0 is replaced by x_0+1 . If, however, $x_0=9$, then x_0 is replaced by 0, and x_1 is replaced by x_1+1 ("carry one")—unless, that is, $x_1=9$ also, in which case both x_0 and x_1 are replaced by 0, and x_2 is replaced by x_2+1 —unless, etc., etc. In these terms it is easy to define the promised transformation T on X : it is the obvious generalization of adding 1. If

$$x = (x_0, x_1, x_2, \dots)$$

is an element of x , and if $x_0 \neq m_0 - 1$, then

$$Tx = (x_0 + 1, x_1, x_2, \dots).$$

If, however, $x_0 = m_0 - 1$, and if, in fact, k is the smallest index such that $x_{k+1} \neq m_{k+1} - 1$, then

$$Tx = (0, \dots, 0, x_{k+1} + 1, x_{k+2}, \dots),$$

where the number of initial zeros is exactly $k+1$ (one each for x_0, \dots, x_k). If $x_k = m_k - 1$ for all k , then

$$Tx = (0, 0, 0, \dots).$$

Intuitively X may be thought of as an infinite adding machine; the action of T on the numbers in X is simply to add 1. The proof that T is not even potentially measure-preserving is a complicated combinatorial argument.

Epilogue. Ergodic theory is very much alive these days; there are new results and there are new problems. Now that it is known, for instance, that there exist measure-theoretically interesting transformations with no invariant measure of the proper sort, the conjugacy problem for such transformations and the applicability of the concept of entropy to them become worthy of consideration. The same thing is true for the operators on L_1 that enter into the generalized ergodic theorems discussed above. Other old results and techniques also deserve to be extended to the newly important transformations and operators; this is especially true of the topological studies sometimes pursued under the name of approximation theorems.

Even according to the narrowest known definition of ergodic theory there are some parts of the subject that were not mentioned in this

report. There has been, for instance, some recent progress in the theory of conservative transformations and in the spectral theory of measure-preserving transformations. The latter subject, especially, constitutes a far from closed chapter. Since, in particular, asymptotic results (ergodic theorems) are usually true for both finite and infinite measure spaces, it is sometimes tacitly assumed that the same is true for spectral results. In fact, however, knowledge of the possible spectral behavior of measure-preserving transformations is meager for finite measure spaces and almost nil for infinite ones.

I hope that in the near future, in the course of the next twelve years, say, humanity learns sufficiently many new answers to these fascinating old questions to warrant another Society address on the subject. I should like to hear that address so that I may discover how everything came out and who did it.

REFERENCES

1. L. M. Abramov, *The entropy of an automorphism of a solenoidal group*, Teor. Veroyatnost. i Primenen. vol. 4 (1959) pp. 249–254.
2. L. Breiman, [a] *The individual ergodic theorem of information theory*, Ann. Math. Statist. vol. 28 (1957) pp. 809–811. [b] *A correction to the individual ergodic theorem of information theory*, Ann. Math. Statist. To appear.
3. R. V. Chacon and D. S. Ornstein, *A general ergodic theorem*, Illinois J. Math. vol. 4 (1960) pp. 153–160.
4. N. Dunford and J. T. Schwartz, *Convergence almost everywhere of operator averages*, J. Rat. Mech. Anal. vol. 5 (1956) pp. 129–178.
5. A. Hajian and S. Kakutani, *Weakly wandering sets and invariant measures*. To appear.
6. E. Hopf, *The general temporally discrete Markoff process*, J. Rat. Mech. Anal. vol. 3 (1954) pp. 13–45.
7. A. Ionescu-Tulcea, *Contributions to information theory for abstract alphabets*. To appear.
8. A. N. Kolmogorov, [a] *A new metric invariant of transitive dynamical systems and automorphisms of Lebesgue spaces*, Dokl. Akad. Nauk SSSR vol. 119 (1958) pp. 861–864. [b] *Entropy per unit time as a metric invariant of automorphisms*, Dokl. Akad. Nauk SSSR vol. 124 (1959) pp. 754–755.
9. D. S. Ornstein, *On invariant measures*. Bull. Amer. Math. Soc. vol. 66 (1960) pp. 297–300.
10. B. McMillan, *The basic theorems of information theory*, Ann. Math. Statist. vol. 24 (1953) pp. 196–219.
11. V. A. Rokhlin, *Entropy of metric automorphism*, Dokl. Akad. Nauk SSSR vol. 124 (1959) pp. 980–983.
12. Y. Sinai, *On the concept of entropy for a dynamic system*, Dokl. Akad. Nauk SSSR vol. 124 (1959) pp. 768–771.

UNIVERSITY OF CHICAGO