Ergodic Theory

Paul C. Kettler, '63

Thesis Submitted in Partial Fulfillment
of the Requirements
for the Bachelor of Arts Degree

Department of Mathematics
Princeton University
1963

Contents

Chapter 1

Fundamentals

1.1 What is Ergodic Theory?

Briefly, ergodic theory is the study of the asymptotic behavior of certain classes of measure-preserving transformations from a measure space to itself. Unfortunately, such a simple definition gives little insight toward justifying any special interest in the subject. To understand why mathematicians began to develop the theory it is necessary to look into the subject's history.

Ergodic theory has its origins in statistical mechanics. Physicists discovered that the state of a contained gas, for example, could be described by giving the location of each particle along with its velocity. Thus in the case of n particles, the state of the system could be specified by giving $6n$ numbers, or coordinates — three position components and three velocity components — for each particle. Such a $6n$-tuple was considered to be a point in a $6n$-dimensional "phase" space. Assuming that phase space were assigned a metric, and that the states varied continuously with time, a continuous curve was described in the phase space. According to classical mechanics, the knowledge of one point on this curve would enable one to determine in theory the path of the curve as time continued. In practice, however, it is not possible to determine so accurately the state of a system. Hence, a more relevant question than, "Given the state at time t_0, what will the state be at time t_1?" is the question, "Given that the state at time t_0 lies in a subset U of phase space, what is the probability that at time t_1 the state lies in subset V of phase space?" An even more interesting question is, "What will probably happen to the points of U as t tends to infinity?"

Such considerations led Liouville to investigate what happens to a subset of phase space when the space changes (flows) continuously with time according to accepted physical laws. An important theorem of his in statistical mechanics states that if phase space is assigned the Euclidean metric, then the Lebesgue measure of a subset equals the Lebesgue measure of its transform.

To the mathematician a number of theoretical questions arose immediately. One of these was, "What can be said concerning Liouville's theorem if the number of particles is taken to be infinite, or if the containing space is many dimensional or infinite dimensional?" Another was, "What happens if the transformations of a space are arbitrary, and not restricted to one-parameter groups of transformations such as phase space flows?" With these questions before them, the mathematicians took their basic space to be an abstract space with a measure assigned, took their transformations to be simple functions from the space to itself, and with appropriate simplifications as necessary for convenience began to examine the properties of measure-preserving transformations as applied repeatedly to the basic space.

This disquisition owes undying gratitude to the seminal contributions of pioneers of this and related fields, as represented particularly in these works which are otherwise not cited, but whose influence pervades the entire book (Kakutani 1950; Riesz and Sz.-Nagy 1955; Halmos 1956; Kolmogorov and Fomin 1961).

1.2 Some Definitions

Before plunging directly into the theory it is necessary to be familiar with the several concepts defined below.

Definition 1.1. A *sigma-algebra of sets* is a collection of sets that is closed under countable unions, intersections, symmetric differences, complements, and combinations thereof. In particular, the union of all the sets is an element in the algebra.

Definition 1.2. A *measure* is a non-negative, countably additive set function.

Definition 1.3. A *measure space* is a sigma-algebra of sets, together with a measure defined on all elements of the algebra.

Definition 1.4. A *sigma-finite measure space* is a measure space with the restriction that the entire space is the union of countably many sets of finite measure. We will consider only sigma-finite measure spaces.

Definition 1.5. A *measurable transformation* is a function from a space to a measure space such that the inverse image of a measurable set is a measurable set.

Definition 1.6. A transformation T mapping a measure space X onto a measure space Y is called *invertible* if there exists a transformation S such that $ST(x) = x$ for all x in X. It follows that $TS(y) = y$ for all y in Y, and we are therefore justified in calling S the *inverse* of T, written T^{-1}.

1.3 Recurrence — a Simple Asymptotic Property

For a measure-preserving transformation from a measure space of finite measure to itself a general theorem regarding the recurrence of points under repeated application of the transformation can be stated.

Theorem 1.7. *If U is a measurable subset of a space X of positive measure, and T is a measure-preserving transformation from X to itself, then for almost every $x \in U$, $Y^n(x) \in U$ for some n.*

Proof. The proof is indirect. Look at the set of points $U' \subset U$ such that for any $x' \in U'$, $T^n(x') \notin U$ for any n. Following from the fact that the transform of any point in X is either in U or in $X \setminus U$, we know that U' is always measurable. Explicitly, it is the countable intersection of measurable sets given by

$$U' = U \cap T^{-1}(X \setminus U) \cap T^{-2}(X \setminus U) \cap \cdots \cap T^{-n}(X \setminus U) \cap \cdots ,$$

where $T^{-n}(X \setminus U)$ means the set of points mapped into $X \setminus U$ by T^n, *etc.* Assume that the measure of U' is positive. From the construction of U', $T^{-n}(U')$ is disjoint from U' for all n, for if that were not the case then there would be a point in U' such that its transform under T^n were also in U', contradicting the assumption. Furthermore, the whole sequence $\{T^n(U')\}$ is mutually disjoint, because

$$T^{-(n+k)}(x) \in U' \iff T^{-n}(x) \in T^{-k}(U'),$$

implying that

$$T^{-n}(U') \cap T^{-(n+k)}(U') = T^{-n}\left(U' \cap T^{-k}(U')\right),$$

which is empty. Inasmuch as T is measure preserving and the measure of X is finite, a contradiction has been established. □

Theorem (1.7) can actually be strengthened by replacing "some n" by "infinitely many n."

Theorem 1.8. *Given the hypothesis of Theorem* (1.7)*, for almost every $x \in U$, $T^n(x) \in U$ for infinitely many n.*

Proof. Let U_n be the set of points which never return to U under repeated applications of T^n, and let V be the set of all x such that $x \in U \setminus (U_1 \cup U_2 \cup \cdots)$. Since $x \in U \setminus U_1$, $T^N(x)$ is in U for some N. Since $x \in U \setminus U_N$, $T^{MN}(x) \in U$ for some M. The argument can be continued indefinitely, giving an infinite sequence of recurrence times. □

Chapter 2

Functions on a Measure Space

2.1 Isometries

Given a measure space X and a measure subspace U, define a characteristic function $f : X \mapsto \mathbb{R}$ as follows. Let

$$f(x) = \begin{cases} 1 & \text{for } x \in U \\ 0 & \text{for } x \in X \setminus U \end{cases}$$

Then using this definition, Theorem (1.8) could be restated, "Given the hypothesis of Theorem (1.7), for almost every $x \in U$, the series $\{S_n\}$, where $S_n = \sum_{i=0}^{n-1} f\big(T^i(x)\big)$, diverges." At this point, an important generalization can be made. We need not restrict consideration to the characteristic function. Any non-negative, measurable function will work as well. Recall:

Definition 2.1. a *measurable function* from a measure space to the complex numbers is a function constructed so that the set of points $\{x\}$ such that $|f(x)| < r$ is a measurable set for any real r. Then we have:

Theorem 2.2. *Given a measure space X and a non-negative (real-valued) measurable function f, then for almost every $x \in U$, where U is the set of all x with $f(x) \geq 0$, the sequence $\{S_n\}$, where $S_n = \sum_{i=0}^{n-1} f(T^i)(x)$, diverges.*

Proof. Let U_j be the set of all x such that $f(x) \geq 1/j$. Then $\cup_{j=1}^{\infty} U_j = U$, and by Theorem (1.8) almost every $x \in U_j$ returns to U_j infinitely often. For every $x \in U$ we can pick a j so that x is in U_j. For almost every x in this U_j the number of terms of $\{S_n\}$ greater than $1/j$ increases without bound as n goes to infinity. Since no term of $\{S_n\}$ is negative, $\{S_n\}$ diverges. The conclusion follows from the fact that the set of points for which $\{S_n\}$ possibly does not diverge has measure zero in each U_j, and U is a countable union of the $\{U_j\}$. □

Knowing that almost every point of a subspace U of a measure space X returns to U infinitely often under the repeated applications of a measure-preserving transformation

T leads one to consider the following question: "As n gets large, for what fraction of the spaces $T^i(U)$, $0 \le i \le n-1$, is the point $T^i(x)$ contained in $T^i(U)$, $x \in U$? Furthermore, does this fraction tend to a limit as n goes to infinity?" If we again define a characteristic function f on X such that f on points of U is 1 and f on other points of X is 0, the fraction under consideration is $(1/n)\sum_{i=1}^{n-1} f\bigl(T^i(x)\bigr)$. Actually, with the problem in this state the analysis involves the tricky problems of pointwise convergence, left for later. At this stage, it is more instructive to examine first certain relationships among functions of X. Suppose we begin to consider not just real-valued functions, but complex-valued functions, also, and suppose we define $g(x)$ equal to $f\bigl(T(x)\bigr)$. Then knowing f gives us g, *i.e.*, g can be written as a function of f, say $R(f)$. Then if we let $h(x) = g\bigl(T(x)\bigr) = f\left[T\bigl(T(x)\bigr)\right]$, written $f\bigl(T^2(x)\bigr)$, we have $h = R(g) = R\bigl(R(f)\bigr)$, written $R^2(f)$. Continuing the process inductively, we see that T^i corresponds, in the above manner, to R^i for all positive i. We find, then, that the transformation R has some familiar properties. First of all it is linear. Given complex numbers a and b, $R(af + bg) = aR(f) + bR(g)$. This fact follows directly from the definition of R.

$$R(af+bg)(x) = (af+bg)\bigl(T(x)\bigr) = af\bigl(T(x)\bigr) + bg\bigl(T(x)\bigr) = aR(f)(x) + bR(g)(x),$$

for all x. Furthermore, we have an important theorem regarding the action of R on a certain space of functions. Before giving the theorem, I list the following definitions.

Definition 2.3. The space of complex-valued functions on a measure space X such that $\left(\int_X |f(x)|^k \,\mathrm{d}\mu\right)^{1/k}$ is finite is called the L_k *space*, for positive integral k. (μ is the measure defined on the measure space.)

Definition 2.4. The L_k *norm* of a function in the L_k space is the quantity $\left(\int_X |f(x)|^k \,\mathrm{d}\mu\right)^{1/k}$, written $\|f\|_k$.

Definition 2.5. An *isometry* on the L_k space is a transformation S such that for $f \in L_k$, also $Sf \in L_k$, and $\|Sf\|_k = \|f\|_k$.

Theorem 2.6. *Given a measure-preserving transformation T on a measure space X, the function R defined by $Rf(x) = f\bigl(T(x)\bigr)$ for all $x \in X$ is an isometry on L_1.*

Proof. Take the characteristic function g of a subset U of finite measure. Since $Rg(x) = g\bigl(T(x)\bigr)$, $R(g)$ is the characteristic function of $T^{-1}(U)$. In this case $\|g\|_1$ equals the measure of U. Then $\|R(g)\|_1$, which equals the measure of $T^{-1}(U)$, also equals $\|g\|_1$ inasmuch as T is measure preserving. Since R is linear, it is an isometry on all such combinations of characteristic functions. Not all functions in L_1, however, are finite linear combinations of characteristic functions. For example, take the unit interval of real numbers, $0 \le r < 1$, with Lebesgue measure for the measure space, and the function to be $h(x) = x$. Nevertheless, any non-negative (real-valued) function f_+ can be represented as the pointwise limit of pointwise increasing functions $\{f_n\}$ that are finite linear combinations of characteristic functions. (In the example, a sequence of functions satisfying the criterion is given by the representative function

$$f_n = \frac{k}{n} \leq x < \frac{k+1}{n},\ k = 1, 2, \ldots, n-1,$$

as n goes through the index set $\{2^0, 2^1, 2^2, \ldots, 2^m, \ldots\}$.) Since $Rf_n(x) = f_n\big(T(x)\big)$, which goes to $f_+\big(T(x)\big) = Rf_+(x)$, for all $x \in X$, the monotonicity of the $\{f_n\}$ implies that $\lim_{n\to\infty} \|f_n\|_1 = \|f_+\|_1$, and hence that $\lim_{n\to\infty} \|Rf_n\|_1 = \|Rf_+\|_1 \cdot \|Rf_+\|_1 = \|f_+\|_1$ because $\|f_n\|_1 = \|Rf_n\|_1$ for each n. The conclusion for any (complex-valued) function f follows from the fact that $\|f\|_1 = \||f|\|_1$. □

Corollary 2.7. *R is an isometry on the L_k space if $\mu(X) < \infty$.*

Proof. Any function f in L_k is also in L_1, and the L_k norm of f is simply the k^{th} root of the L_1 norm of f^k. R is therefore an isometry on L_k because it preserves the L_1 norm. □

Remark. An arbitrary isometry on L_k is not necessarily an isometry on L_1 because L_k does not contain all of L_1.

2.2 The Mean Ergodic Theorem

Consider now the case where T has an inverse. One can define a function S on L_1 using T^{-1} just as was done with T, *i.e.*, $Sf(x) = f\big(T^{-1}(x)\big)$. S has all the properties possessed by R, and in fact S is the inverse of R, for

$$RSf(x) = R\big(Sf(x)\big) = Sf\big(T(x)\big) = f\big(T^{-1}T(x)\big) = f(x)$$

Let a scalar product on L_2 be defined as follows:

Definition 2.8. The *scalar product* of two functions f and g in L_2, written (f, g), is the real number given by $\int_X f(x)\overline{g(x)}\,\mathrm{d}\mu$, where X is the measure space and μ is the measure. It is a theorem that whenever f and g are in L_2 the scalar product exists. The norm of f, then, is the square root of the scalar product of f with itself. The L_2 space together with the scalar product constitute a Hilbert space (definition omitted,) and it is a theorem that for Hilbert space (all Hilbert spaces are isomorphic) an isometry with an inverse is a unitary operator.

Definition 2.9. A *unitary operator* Q is a function on a Hilbert space such that for any f and g in the space, $(Qf, Qg) = (f, g)$. Note that any unitary transformation is an isometry, because $\|Qf\|_2 = (Qf, Qf)^{1/2} = (f, f)^{1/2} = \|f\|_2$.

Now, returning to the problem of recurrence, we see that the problem of interest has changed considerably. Instead of being left with the uninteresting problem of pointwise convergence of $\{(1/n)\sum_{i=1}^{n-1} f(T^i(x))\}$ (f is the characteristic function,) we have the problem of convergence in L_2 norm of the sequence $\{(1/n)\sum_{i=0}^{n-1} R^i f\}$ for a function f in a Hilbert space and a unitary operator R on that space.

To get an intuitive idea of what this sequence means for operators on complex Hilbert spaces it is instructive to look at the sequence for unitary operators on finite-dimensional complex vector spaces. In the case of complex dimension one a unitary operator is simply multiplication by a complex number c of absolute value one. If $c = 1$, then the sequence converges to $1 \cdot x$ for all x in the space. If $c \neq 1$, then the n^{th} term of the sequence, which equals

$$\frac{1}{n}(1 + c + c^2 + \cdots + c^{n-1}) \cdot x = \frac{1 - c^n}{n(1-c)} \cdot x$$

is less in absolute value than $2/(n(1-c)) \cdot x$, which goes to zero. In an arbitrary finite complex dimension a unitary transformation can be represented as a diagonal matrix C with entries of absolute value 1, by proper choice of coordinate system. Then the sequence converges to $P(x)$ for all x in the space, where P is a matrix with 1's and 0's corresponding to the entries equal to 1 and different from 1, respectively in C, *i.e.*, P is a projection onto the subspace of vectors with $C(x) = x$. For the Hilbert space case it would be convenient if we could also say that the sequence converges to the projection of the function onto the subspace of functions invariant under the unitary operator R. This indeed is the case. The infinite-dimensional case, as usual, requires its own proof.

Definition 2.10. The *adjoint operator* R^* of a unitary operator R is the transformation defined by the condition that $(f, R^*g) = (Rf, g)$ must hold for every f and g in the underlying space.

Following is an important theorem.

Theorem 2.11 (Mean Ergodic). *If R is a unitary transformation on a Hilbert space H, and if P is the projection onto all elements invariant under R, then the elements $\{f_n = (1/n)\sum_{i=0}^{n-1} R^i f\}$ converge to Pf for every f in the space.*

Proof. The idea of the proof is to separate every f in the space into the sum of two orthogonal elements, one of which is sent to zero in the sequence, the other of which is sent to itself. The orthogonality of the two elements will show that the limit transformation is a projection, and the invariance of the second element will imply that that transformation maps onto the subspace of invariant elements.

Let the subspace H' be the completion of the space of all elements of the form $g - Rg$, for all $g \in H$. Then if f' is an element of H', $f' = g - Rg + g'$ for some g and g' such that

$\|g'\|_2 < \epsilon$ for an arbitrary small positive real number ϵ. Then

$$f'_n = \frac{1}{n}\sum_{i=0}^{n-1} R^i(g - Rg + g') = \frac{1}{n}\sum_{i=0}^{n-1}\left(R^i g - R^{i+1}g + R^i g'\right)$$

$$= \frac{1}{n}\sum_{i=0}^{n-1}\left[\left(R^0 g + R^1 g + R^2 g + \cdots + R^{n-1}g\right) - \left(R^1 g + R^2 g + \cdots + R^n g\right)\right] + \frac{1}{n}\sum_{i=0}^{n-1} R^i g'$$

$$= \frac{1}{n}(g - R^n g) + \frac{1}{n}\sum_{i=1}^{n} R^i g'$$

It follows that $\|f'_n\|_2 \leq (1/n)\cdot 2\,\|g\|_2 + \epsilon$. For all sufficiently large n the first term of this last sum will also be smaller than ϵ, hence f_n goes to zero in norm.

Let the subspace H'' be composed of all elements f'' of H that are invariant under R, *i.e.*, such that $Rf'' = f''$. For these elements, $f''_n = f''$ for all n, implying that the limit element is the original element itself.

The proof will be complete if we can show that H'' is the orthogonal complement of H', because then it will be possible to express every element f of H as the sum of an element in H' and an element in H''. Since the limit function restricted to H' is the zero transformation, and restricted to H'' is the identity, it will follow that it is the projection of all elements of H into the space of invariant functions. Since the invariant elements also belong to H it will be trivial that the limit function is onto (surjective.) We have that

$$(g - Rg, h) = (g, h) - (Rg, h) = (g, h) - (g, R^*h) = (g, h - R^*h)$$

from the linearity of the scalar product and the definition of the adjoint transformation. This equation says the set of elements orthogonal to elements of the form $g - Rg$ (*i.e.*, such that the scalar product with $g - Rg$ is zero) for all $g \in H$ consists precisely of the elements invariant under R^* (*i.e.*, such that $h - R^*h = 0$.) This statement is also true for the limits of the elements $g - Rg$ because then we would have $(g - Rg + g', h) = (g, h - R^*h) + (g', h)$, and because $\|g'\|_2$ can be made arbitrarily small, (g', h) can be made arbitrarily small, and thus be neglected. It is a theorem that for unitary transformations $R^* = R^{-1}$. Therefore, the elements invariant under R^* are the same as the elements invariant under R^{-1}, and hence invariant under R, *i.e.*, the elements of H''. □

2.3 The Statistical, Individual, and Maximal Ergodic Theorems

This theorem can be generalized by noting that for the proof the only facts we needed to know about R were that R be defined everywhere and linear, that $\|Rf\|_2 \leq \|f\|_2$, and that the invariant elements of R be the same as the invariant elements of R^*. Actually, the third property is a consequence of the first two. I shall make this statement as a theorem after listing two definitions.

Definition 2.12. a *contraction* R of a Hilbert space H is a linear transformation defined on the whole space H such that $\|Rf\|_2 \le \|f\|_2$ for each $f \in H$.

Definition 2.13. The *norm of a transformation* R, written $\|R\|_2$ on a Hilbert space H is the supremum of quantities $\|Rf\|_2 / \|f\|_2$ for all non-zero $f \in H$.

Theorem 2.14. *If R is a contraction of a Hilbert space H then the invariant elements of R and R^* are the same.*

Proof. From Definitions 2.9 and 2.10 we have that $\|R\|_2 \le 1$. It is a theorem for Hilbert space that $\|R\|_2 = \|R^*\|_2$, hence $\|R^*\|_2 \le 1$. For every element f invariant under R,

$$\|f\|_2^2 = (f, f) = (Rf, f) = (f, R^*f) \le \|f\|_2 \cdot \|R^*f\|_2 \text{ (Schwarz inequality) } \le \|f\|_2^2,$$

i.e., the strict inequalities cannot hold, giving

$$(f, R^*f) = \|f\|_2 \cdot \|R^*f\|_2 \text{ and } \|R^*f\|_2 = \|f\|_2$$

Therefore,

$$\|f - R^*f\|_2^2 = \|f\|_2^2 - (f, R^*f) - (R^*f, f) + \|R^*f\|_2^2 = 0,$$

giving that $R^*f = f$. Since the argument proceeds in the same manner if R and R^* are interchanged, the conclusion follows. □

Theorem 2.14 allows us to sharpen Theorem 2.11 by generalizing the hypothesis. I give the result as a new theorem.

Theorem 2.15. *If R is a contraction of a Hilbert space H, and if P is the projection onto all elements invariant under R, then the elements $\left\{f_n = (1/n)\sum_{i=0}^{n-1} R^i f\right\}$ converge to Pf for every f in the space.*

Proof. The proof proceeds exactly as the proof of Theorem 2.11. □

The so-called Statistical Ergodic Theorem of John von Neumann is a corollary of Theorem 2.11 (and therefore also of Theorem 2.15.)

Corollary 2.16 (Statistical Ergodic). *If T is a measure-preserving transformation on the measure space X, if T^{-1} exists, if $\mu(X) < \infty$, and if $f \in L_2$, then there exists a subsequence of $\{f_n(x)\} = \left\{(1/n)\sum_{i=0}^{n-1} f\big(T^i(x)\big)\right\}$ which converges almost everywhere to a function $Pf(x)$, invariant with respect to T almost everywhere. In other words, $Pf(x) = Pf\big(T(x)\big)$ except on a set of measure zero.*

Proof. If R is the unitary operator defined by $Rf(x) = f\big(T(x)\big)$, then Theorem 2.11 says that the sequence $\left\{f_n = (1/n)\sum_{i=0}^{n-1} R^i f\right\}$ converges to a function Pf which is invariant under R. It is a theorem that such a sequence has a subsequence convergent almost everywhere to Pf. Remembering that Pf represents an equivalence class of functions that are the same except possibly on a set of measure zero, we have that $R(Pf)(x) = Pf(x) = Pf\big(T(x)\big)$ for almost every $x \in X$. □

After studying the cases in L_2 it is natural to try to see if similar results can be obtained for the space L_1. Proofs similar to those given for L_2 are not possible without complications because of the absence of a Hilbert space construction. However, by using methods of pointwise convergence instead of mean convergence it is possible to prove the so-called Individual Ergodic Theorem. For finite measure spaces convergence in the L_1 norm follows from almost everywhere convergence, thus the theorem is not markedly weaker than the corresponding L_2 theorems, and in some respects is stronger. After stating the Theorem I will give a definition and then prove two lemmas concerning sequences of real numbers and real-valued functions. The proof of the theorem will follow.

Theorem 2.17 (Individual Ergodic). *Given a measure-preserving transformation T on a measure space X, a function $f \in L_1$, and the isometry R defined by $Rf(x) = f(Tx)$ for all $x \in X$, the series $\left\{(1/n)\sum_{i=0}^{n-1} T^i(x)\right\}$ converges almost everywhere to a function $f^* \in L_1$ such that $f^*(Tx) = f^*(x)$ almost everywhere. If the measure of X is finite, then $\|f^*\|_1 = \|f\|_1$, and the sequence $\left\{(1/n)\sum_{i=0}^{n-1} R^i f\right\}$ converges in L_1 norm to f^*, such that $Rf^* = f^*$.*

Proof. Deferred

Definition 2.18. Given $m \leq n$, an *m-leader* a_i of a sequence of real numbers $\{a_0, \dots, a_n\}$ is an element of the sequence such that for some p with $1 \leq p \leq m$, the sum $a_i + a_{i+1} + \cdots + a_{i+p-1} \geq 0$.

Lemma 2.19. *For any finite sequence of real numbers the sum of the m-leaders is non-negative.*

Proof. If no m-leaders exist then the lemma is true. If a_i is the first m-leader and $a_i + a_{i+1} + \cdots + a_{i+p-1}$ the shortest sequence with non-negative sum that it leads, then it follows that each a_j of that sequence is also an m-leader because $a_j + a_{j+1} + \cdots + a_{i+p-1} \geq 0$. In fact, if this were not the case, then for some j the sum $a_1 + a_2 + \cdots + a_{j-1}$ would be positive, contradicting the choice of p. Continue inductively by examining $a_{i+p} + a_{i+p+1} + \cdots + a_n$ for m-leaders. The sum of these shortest sequences of m-leaders is non-negative, and there are no other m-leaders. □

The next lemma, also called the Maximal Ergodic Theorem, was first proved for one-to-one measure-preserving transformations by K. Yosida and S. Kakutani (Yosida and Kakutani 1939). F. Riesz simplified the statement of the theorem slightly, however managed to give a proof without needing the one-to-one assumption (Riesz 1944). It is essentially his idea that I follow below.

Lemma 2.20 (Maximal Ergodic Theorem). *If $U = \left\{x \mid \sum_{i=0}^{n-1} f(T^i x) \geq 0\right\}$ for a measure-preserving transformation T and a real-valued function f on X, then $\int_U f(x)\,\mathrm{d}\mu \geq 0$.*

Proof. Let $U_m = \left\{x \mid \sum_{i=0}^{n-1} f(T^i x) \geq 0\right\}$ for some $p \leq m$. Then $U_1 \subset U_2 \subset \cdots$ and $\cup_{m=0}^{\infty} U_m = U$. Therefore, it is sufficient to prove that $\int_{U_m} f(x)\,\mathrm{d}\mu \geq 0$ for each m.

For each point $x \in X$ consider the m-leaders of the sequence $f(x), f(Tx), \ldots, f(T^{m+n-1}x)$ for any fixed positive n. Let $s(x)$ be their sum. Let $V_j = \{x \mid f(T^j x) \text{ is an } m\text{-leader of } f(x), f(Tx), \ldots, f(T^{m+n-1}x)\}$. Define

$$g_j(x) = \begin{cases} 1 & \text{for } x \in V_j \\ 0 & \text{for } x \notin V_j \end{cases}$$

Then $s(x) = \sum_{j=0}^{n+m-1} f(T^j x) \cdot g_j(x)$ is defined for all $x \in X$, and by Lemma 2.19, $s(x) \geq 0$. Since each V_j is measurable (each is a subset of a measure space) and $s(x) < \infty$ by virtue of the finite number of terms being summed, we have

$$\int_X s(x)\,\mathrm{d}\mu = \int_X \sum_{j=0}^{n+m-1} f(T^j x) \cdot g_j(x)\,\mathrm{d}\mu = \sum_{j=0}^{n+m-1} \int_{V_j} f(T^j x)\,\mathrm{d}\mu \geq 0 \tag{2.1}$$

For $j = 1, 2, \ldots, n-1$ and given

$$x, Tx \in V_{j-1} \iff x \in V_j,$$

because the left side means $f\big(T^{j-1}(Tx)\big)$ is an m-leader, and the right side means $f\big(T^j(x)\big)$ is an m-leader. Thus $V_{j-1} = T(V_j)$, and repeating the argument, $V_0 = T^j(V_j)$ for $1 \leq j \leq n-1$. Therefore,

$$\int_{V_j} f(T^j x)\,\mathrm{d}\mu = \int_{V_0} f(x)\,\mathrm{d}\mu$$

by replacing $T^j(x)$ with x. Hence, the first n terms of Equation (2.1) above are equal. Furthermore, $V_0 = U_m$; a point is in one or the other if and only if $f(x)$ is an m-leader. This gives by Equation (2.1) that

$$n \int_{U_m} f(x)\,\mathrm{d}\mu + m \int_X |f(x)|\,\mathrm{d}\mu \geq 0$$

by replacing the last m terms with $\|f\|_1$, obviously greater than or equal to any integral on a subset. Divide through by n, and since n was chosen arbitrarily, let it go to infinity. The conclusion:

$$\int_{U_m} f(x)\,\mathrm{d}\mu \geq 0$$

□

Proof of Theorem 2.17. There is no loss of generality if we consider only real-valued functions, because $\{(1/n)\sum_{i=0}^{n-1} f(T^i x)\}$, f complex valued, converges pointwise if and only if $\{(1/n)\sum_{i=0}^{n-1} |f(T^i x)|\}$ converges.

Let $a < b$ be real numbers, and let

$$Y(a,b) = \left\{ x \mid \liminf \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) < a < b < \limsup \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right\}$$

Since every $x \in X$ can be tested for inclusion in $Y(a,b)$, $Y(a,b)$ is measurable, and clearly, $Y = T^{-1}Y$ because an n-fold average is not changed significantly by the addition of another term as n gets large. The plan of the proof is to prove successively that

1. the measure of $Y(a,b)$ is finite, and
2. the measure of $Y(a,b)$ is zero.

It then will follow that $\left\{(1/n)\sum_{i=0}^{n-1} f(T^i x)\right\}$ converges almost everywhere.

Assume $b > 0$, because if $b < 0$, then $a < 0$, and the argument could be carried out using $-f$ and the real numbers $(-b, -a)$. Let $C \subset Y(a,b)$ be an arbitrary measurable subset with finite measure. Let

$$g(x) = \begin{cases} 1 & \text{for } x \in C \\ 0 & \text{otherwise} \end{cases}$$

Let

$$F = \left\{ x \in X \mid (f - bg)(x) + (f - bg)(Tx) + \cdots + (f - bg)(T^{n-1}x) \geq 0 \text{ for some } n \right\}$$

Then Lemma 2.20 implies that $\int_F [f(x) - bg(x)]\,\mathrm{d}\mu \geq 0$. If $x \in Y(a,b)$ then infinitely many of the averages $\left\{(1/n)\sum_{i=0}^{n-1} f(T^i x)\right\}$ are greater than b. Therefore $(1/n)\sum_{i=0}^{n-1} f(T^i x) \geq nb$ for some n, and since $0 \leq g(x) \leq 1$ for all $x \in X$, $\sum_{i=0}^{n-1}[f(T^i x) - bg(T^i x)] > 0$ for that n, implying that $x \in F$, *i.e.*, $Y \subset F$. Since $\int_F f(x)\,\mathrm{d}\mu \geq \int_F bg(x)\,\mathrm{d}\mu$ it follows that $\int_X |f(x)|\,\mathrm{d}\mu \geq b\,\mathrm{m}(C)$ ($\mathrm{m}(\cdot)$ is the measure function.) Thus if $Y(a,b)$ has a subset of finite measure, then the measure of that subset cannot exceed $(1/b)\,\|f\|_1$. We are assuming sigma finiteness on X, therefore $\mathrm{m}\big(Y(a,b)\big) \leq (1/b)\,\|f\|_1$.

Now consider applications of Lemma 2.20 to the space $Y(a,b)$. This can be done because $Y(a,b)$ is invariant under T. Look at the function $\big(f(x) - b\big)$ defined on $Y(a,b)$. Then if

$$G = \left\{ x \in Y \mid \big(f(x) - b\big) + \big(f(Tx) - b\big) + \cdots + \big(f(T^{n-1}x) - b\big) \geq 0 \text{ for some } n \right\},$$

we see that $x \in G$ if $(1/n)\sum_{i=0}^{n-1} f(Tx) \geq b$ for some n. However, every point $x \in Y(a,b)$ so qualifies. Therefore, $G = Y$. We then have by Lemma 2.20 that $\int_Y \big(f(x) - b\big)\,\mathrm{d}\mu \geq 0$. Similarly, $\int_Y \big(a - f(x)\big)\,\mathrm{d}\mu \geq 0$. Adding these two gives $\int_Y (a - b)\,\mathrm{d}\mu \geq 0$, but we know $a < b$. Therefore, $\mathrm{m}\big(Y(a,b)\big)$ must be zero.

Apply this result to the (countable) set of all pairs of rational numbers $a < b$. It follows that

$$\mathrm{m}\left\{ x \mid \liminf \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \neq \limsup \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right\} = 0,$$

i.e., the averages converge almost everywhere.

Call the limit function f^*. Then

$$\int_X \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right| \,\mathrm{d}\mu \leq \frac{1}{n}$$

and

$$\int_X \left| \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right| \,\mathrm{d}\mu = \int_X |f(x)| \,\mathrm{d}\mu = \|f\|_1$$

because T is measure preserving. Inasmuch as $\|f\|_1$ is finite, the first term, which converges to $\int_X |f^*(x)| \,\mathrm{d}\mu = \|f^*\|_1$, *i.e.*, $f^* \in L_1$. For almost all $x \in X$,

$$f^*(Tx) - f^*(x) = \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=1}^{n} f(T^i x) - \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) \right] = \lim_{n\to\infty} [f(T^n x) - f(x)] = 0$$

Now consider the case where $\mathrm{m}(x)$ is finite. To prove that $\int_X f(x) \,\mathrm{d}\mu = \int_X f^*(x) \,\mathrm{d}\mu$ first look at two relations between f and f^* given by Lemma 2.20. If $f^*(x) \geq a$ everywhere, then given any $\epsilon > 0$, $(1/n) \sum_{i=0}^{n-1} f(T^i x) > a - \epsilon$ for each $x \in X$ and choice of n sufficiently large. $\sum_{i=0}^{n-1} [f(T^i x) - (a - \epsilon)] > 0$ under those conditions, implying that $\int_X f(x) \,\mathrm{d}\mu \geq (a - \epsilon) \,\mathrm{m}(X)$ for each ϵ, thus $\int_X f(x) \,\mathrm{d}\mu \geq a \cdot \mathrm{m}(X)$. Likewise, if $f^*(x) \leq b$ everywhere, $\int_X f(x) \,\mathrm{d}\mu \leq b \cdot \mathrm{m}(X)$.

Let

$$X(k,n) = \left\{ x \in X \;\middle|\; \frac{k}{2^n} \leq f^*(x) \leq \frac{k+1}{2^n} \right\}$$

Then $X(k,n)$ is invariant under T because $f^*(Tx) = f^*(x)$ for all $x \in X$. Therefore, T restricted to $X(k,n)$ is measure preserving, and

$$\frac{k}{2^n} \,\mathrm{m}\big(X(k,n)\big) \leq \int_{X(k,n)} f(x) \,\mathrm{d}\mu \leq \frac{k+1}{2^n} \,\mathrm{m}\big(X(k,n)\big)$$

Because of the bounds of f^* on $X(k,n)$ this inequality is also true for f^*, and

$$\left| \int_{X(k,n)} f(x) \,\mathrm{d}\mu - \int_{X(k,n)} f^*(x) \,\mathrm{d}\mu \right| \leq \frac{1}{2^n} \,\mathrm{m}\big(X(k,n)\big)$$

Since $X = \cup_k \big(X(k,n)\big)$, and the $X(k,n)$ are disjoint with respect to k (except for sets of measure zero,) summing over k gives

$$\left| \int_X f(x) \,\mathrm{d}\mu - \int_X f^*(x) \,\mathrm{d}\mu \right| \leq \frac{1}{2^n} \,\mathrm{m}(X)$$

Let n go to infinity, giving

$$\int_X f(x)\,\mathrm{d}\mu = \int_X f^*(x)\,\mathrm{d}\mu$$

To prove that $\left\{(1/n)\sum_{i=0}^{n-1} R^i f\right\}$ converges in L_1 norm to f^*, the pointwise limit function, it is necessary to show that $\lim_{n\to\infty}\left\|(1/n)\sum_{i=0}^{n-1} R^i f - f^*\right\|_1 = 0$. An equivalent statement is that

$$\lim_{n\to\infty}\int_X \left|\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) - f^*(x)\right| \mathrm{d}\mu = 0$$

If f is bounded, then the statement is true because for any $\epsilon > 0$ the integrand can be majorized by $\epsilon \cdot \mathrm{m}(X)$ inasmuch as the averages converge almost everywhere. If f is not bounded, then the following inequality holds for any bounded function g by the triangle inequality.

$$\int_X \left|\frac{1}{n}\sum_{i=0}^{n-1} f(T^i x) - f^*(x)\right| \mathrm{d}\mu$$

$$\leq \int_X \left|\frac{1}{n}\sum_{i=0}^{n-1}\big(f(T^i x) - g(T^i x)\big)\right| \mathrm{d}\mu + \int_X \left|\frac{1}{n}\sum_{i=0}^{n-1} g(T^i x) - g^*(x)\right| \mathrm{d}\mu + \int_X |g^*(x) - f^*(x)|\,\mathrm{d}\mu$$

The first term on the right side is not greater than $\int_X |f(x) - g(x)|\,\mathrm{d}\mu$ because T is measure preserving; the last term equals $\int_X |g(x) - f(x)|\,\mathrm{d}\mu$ because $g^*(x) - f^*(x) = (g - f)^*(x)$, and $\int_X (g - f)^*(x)\,\mathrm{d}\mu = \int_X (g - f)(x)\,\mathrm{d}\mu$. Since the second term concerns the bounded function, choosing n large enough will make it small. Inasmuch as $f \in L_1$, given any $\epsilon > 0$ we can always choose a bounded g so that the first and third terms on the right side are each less than $\epsilon/3$. Then by choosing n large enough the second term can be made smaller than $\epsilon/3$. Thus the left side is less than ϵ, *i.e.*, $\left\{(1/n)\sum_{i=0}^{n-1} R^i f\right\}$ converges to f^* for all $f \in L_1$ on a finite measure space X.

The fact that $Rf^* = f^*$ is just a special case of the situation for arbitrary X that $f^*(Tx) = f^*(x)$ almost everywhere. For, if this latter statement is true, then $Rf^*(x) = f^*(Tx) = f^*(x)$ almost everywhere, and functions that differ only on a set of measure zero are not distinguished. □

Chapter 3

Ergodic Transformations

Up to this point there have been no assumptions made about the transformations on a measure space X, with the exception that they be measure preserving. Turning our attention to these transformations, it can be noted that if a transformation T is invariant on a subspace $U \subset X$, *i.e.*, if $T^{-1}(U) \cup U \setminus T^{-1}(U) \cap U$ has measure zero, the study of X under T can be split into the separate studies of U under T and $X - U$ under T. The transformations of interest, therefore, are those that are invariant only on the whole space on which they act. Such transformations are called ergodic. I give the precise definition.

Definition 3.1. A transformation T on a measure space X is *ergodic* if and only if T invariant on U implies either $\mathrm{m}(U) = 0$ or $\mathrm{m}(X \setminus U) = 0$

In order to apply techniques similar to those used in the foregoing chapter it is necessary to look at complex-valued functions on a measure space, and to see what happens to them under ergodic transformations.

Definition 3.2. An *invariant* function on X is a function f such that given T, $f(Tx) = f(x)$ for almost every $x \in X$. Consequently, if $Rf(x) = f(Tx)$, $Rf = f$ in L_2.

Theorem 3.3. *A transformation T on X is ergodic if and only if every invariant measurable function is constant almost everywhere.*

Proof. If X has measure zero, the theorem is trivially true; therefore, assume $\mathrm{m}(X) > 0$. Then, given that T is ergodic, form the sets

$$X(k, n) = \left\{ x \mid \frac{k}{2^n} \le |f(x)| < \frac{k+1}{2^n} \right\}$$

for a measurable, invariant function f.

$X(k, n)$ if invariant because if $x \in T^{-1}\big(X(k, n)\big)$, then $f(Tx) = f(x)$ almost everywhere implies, since T is measure preserving, that

$$T^{-1}\big(X(k, n)\big) \cup X(k, n) \setminus T^{-1}\big(X(k, n)\big) \cap X(k, n)$$

has measure zero, conforming to the definition.

Since T is ergodic, for any given n only one of the X(k,n) has measure greater than zero. Call this set $X(n)$. Since $\mathrm{m}\big(X(n)\big) = \mathrm{m}(X)$ for all n, it follows that $|f(x)| = k$, a constant almost everywhere. Otherwise one could choose n large enough so that $\mathrm{m}\big(X(n)\big) < \mathrm{m}(X)$.

We can think of f as being a function from X to the circle of radius k in the complex plane. If f were not a constant almost everywhere, then there would exist disjoint subsets W_1 and W_2 in this circle such that $\mathrm{m}\big(f^{-1}(W_1)\big) > 0$ and $\mathrm{m}\big(f^{-1}(W_2)\big) > 0$. But for almost every point $x \in T^{-1}\big(f^{-1}(W_1)\big)$, $f(x) = f(Tx) \in W_1$, implying that $x \in f^{-1}(W_1)$. Since T is measure preserving, it follows that $f^{-1}(W_1)$ is invariant under T. Similarly, $f^{-1}(W_2)$ is invariant under T. Since both of these sets have measure greater than zero, and since they are disjoint (from the single-valuedness of a function,) T cannot be ergodic, contrary to hypothesis. The only possible conclusion is that f is constant almost everywhere.

Conversely, if all invariant functions are constant, there can be no invariant set Y under T such that $0 < \mathrm{m}(Y) < \mathrm{m}(X)$. If such a Y did exist, the function equal to 1 on Y and equal to 0 on $X \setminus Y$ would be invariant under T, and would, of course, be non-constant. Therefore T is ergodic. □

Corollary 3.4. *If* $\mathrm{m}(X) < \infty$*,* T *is ergodic if and only if every invariant function in* L_p *is constant almost everywhere.*

Proof. If T is ergodic then the theorem implies that every invariant function in L_p is constant almost everywhere because every function in L_p is measurable. If T is not ergodic then there is a non-trivial invariant subset of X. The function equal to one on this subset and zero elsewhere is in L_p, but is not constant. □

Now that a concept of erdogicity has been developed, one can look at some of the earlier theorems to see what they imply if T is ergodic.

Theorem 2.17 tells that if $f \in L_1$, $f^*(x) = \lim_{n\to\infty}(1/n)\sum_{i=0}^{n-1} f(T^i x)$ is invariant almost everywhere. If T is ergodic, it follows that $f^*(x) = k$, a constant almost everywhere. In the case $\mathrm{m}(X)$ is infinite, then $f^*(x) = 0$ because f^* is integrable; in the case $\mathrm{m}(X)$ is finite,

$$\int_X f^*(x)\,\mathrm{d}\mu = k \cdot \mathrm{m}(X) = \int_X f(x)\,\mathrm{d}\mu$$

If $f \in L_2$, Theorem 2.11 gives that $\big\{(1/n)\sum_{i=0}^{n-1} R^i f\big\}$, where $Rf(x) = f(Tx)$ converges in L_2 norm to $f^* = k$, without the restriction that $\mathrm{m}(X)$ be finite.

If T is ergodic, then with certain topologies defined on measure spaces it is possible to generalize and strengthen the recurrence Theorems 1.7 and 1.8. In fact, given a point $x \in U \subset X$, a selected measure space, it can be demonstrated that x returns under T not only to U infinitely often, but to any set of positive measure infinitely often. The theorem follows a few definitions. I assume that the reader is familiar with the concept of a topological space.

Definition 3.5. A *countable basis* for a topological space is a denumerable collection of open subsets such that any open set in the space can be represented as the union of basis sets.

Definition 3.6. A topologized measure space is called *non-atomic* if every non-empty open set has positive measure.

Definition 3.7. The *orbit* of a point $x \in X$, an arbitrary set, under the transformation T is the sequence $\{T^n(x)\}$.

Theorem 3.8. *Given an ergodic, measure-preserving transformation T on a topologized non-atomic measure space X with countable basis $\{B_i\}$, the orbit of $x \in X$ is everywhere dense,* i.e., *$T^j x \in B_i$ for some j, given any i.*

Proof. The orbit of x is not dense if and only if for some B_i, $x \in \cap_{j=0}^{\infty}(X \setminus T^{-j}B_i) = \overline{B}_i$. Every such x is mapped into $\overline{B}_i$ by T, and since T is measure preserving, it follows that $\overline{B}_i$ is invariant. But $\overline{B}_i \cap B_i = \varnothing$, hence T is ergodic, and $\mathrm{m}(B_i) > 0$ implies that $\mathrm{m}(\overline{B}_i) = 0$. Since $\{B_i\}$ are countable the set of all $x \in X$ such that the orbits are not dense, $\cup_i \overline{B}_i$, has measure zero. □

When T is ergodic the facts that only trivial sets are invariant and that under suitable conditions the orbit of almost every point passes through every set of positive measure suggest an intuitive idea of mixing. Another way to look at mixing is from the viewpoint of probability. If one normalizes a space X of finite measure by setting $\mathrm{m}'(U) = \mathrm{m}(U)/\,\mathrm{m}(X)$ for $U \subset X$ then m' is a probability measure. We know that

$$f^*(x) = \lim_{n\to\infty} \sum_{i=0}^{n-1} f(T^i x) = \frac{1}{\mathrm{m}(X)} \int\limits_X f(x)\,\mathrm{d}\mu$$

If

$$f(x) = \begin{cases} 1 & \text{for } x \in U \\ 0 & \text{for } x \in X \setminus U \end{cases}$$

then $f^*(x) = \mathrm{m}(U)/\,\mathrm{m}(X) = \mathrm{m}'(U)$. In other words, the average time the images of x under iterations of T spend in U tends in the limit to $\mathrm{m}'(U)$. If in addition

$$g(x) = \begin{cases} 1 & \text{for } x \in V \subset X \\ 0 & \text{for } x \in X \setminus V \end{cases}$$

then

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} f(T^i x) g(x) = f^*(x) g(x) = \frac{\mathrm{m}(U)}{\mathrm{m}(X)} g(x)$$

can be integrated term by term (by the Lebesgue bounded convergence theorem given $0 \leq f(T^i x)g(x) \leq 1$ for all i and x.) The result,

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \mathrm{m}(T^i U \cap V) = \frac{\mathrm{m}\, U \,\mathrm{m}\, V}{\mathrm{m}\, X} = \mathrm{m}'(U)\,\mathrm{m}'(V)\,\mathrm{m}(X),$$

implies when divided by $\mathrm{m}(X)$, that the average probability that $T^n x$ is in V for $x \in U$ tends in the limit to the product of the probabilities of U and V. The limit depends in no way on the specific choice of U or V. Conversely, if the above limit holds for any pair of measurable sets U and V, then T is ergodic. Given a set W which is invariant under T, let $U = V = W$. Then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathrm{m}'(T^{-i} W \cap W) = (\mathrm{m}'\, W)^2$$

thus $\mathrm{m}'\, W = (\mathrm{m}'\, W)^2 \Longrightarrow \mathrm{m}'\, W = 0$ or 1, and hence T is ergodic.

To conclude this chapter I mention briefly two concepts somewhat stronger than ergodicity. As we have just seen, a measure-preserving transformation T acting on a space of finite measure is ergodic if and only if the sequence $\left\{(1/n)\sum_{i=0}^{n-1} \mathrm{m}'(T^{-i} U \cap V)\right\}$ converges to $\mathrm{m}'(U)\,\mathrm{m}'(V)$ for any two sets U and V. The two other concepts, called weak mixing and strong mixing, I list as definitions.

Definition 3.9. A transformation T is called *weakly mixing* if for any two measurable sets U and V, $\lim_{n\to\infty}(1/n)\sum_{i=0}^{n-1} |\mathrm{m}'(T^{-i} U \cap V) - \mathrm{m}'(U)\,\mathrm{m}'(V)| = 0$.

Definition 3.10. A transformation T is called *strongly mixing* if for any two measurable sets U and V, $\lim_{n\to\infty} \mathrm{m}'(T^{-n} U \cap V) = \mathrm{m}'(U)\,\mathrm{m}'(V)$.

To show that strong mixing implies weak mixing, which in turn implies ergodicity, it is sufficient to prove that if $\{a_n\}$ is a sequence of real numbers,

$$\lim_{n\to\infty} a_n = a \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i - a| = 0 \Longrightarrow \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} a_i = a$$

Given $\lim_{n\to\infty} a_n = a$, then for any $\epsilon > 0$ there exists an N such that $|a_i - a| < \epsilon$ for $i > N$. Hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |a_i - a| \leq \lim_{n\to\infty} \left[\frac{1}{n} \sum_{i=0}^{N} |a_i - a| + \frac{1}{n}(n - N)\epsilon\right] = 0 + \epsilon$$

Since ϵ is arbitrary, the minorant term is zero. Also, given $\lim_{n\to\infty}(1/n)\sum_{i=0}^{n-1} |a_i - a| = 0$, then $\lim_{n\to\infty} \left|(1/n)\sum_{i=0}^{N} a_i - a\right| = 0$ because the absolute value of a sum is less than or equal to the sum of the absolute values, and all terms are non-negative. And finally, $\lim_{n\to\infty}(1/n)\sum_{i=0}^{n-1} a_i = a$.

Chapter 4

The Ulam – von Neumann Random Ergodic Theorem

Theorem 2.17, the Individual Ergodic Theorem, stated certain conditions under which the sequence $\left\{(1/n)\sum_{i=0}^{n-1} f(T^i x)\right\}$ converges. A possible generalization of this theorem involves looking not only at the simple iterations $T^i x$, but at the successive points $S^i x$, where S^i is obtained from S^{i-1} by composition with some $\overline{S^i}$ chosen "at random" from some family $\{\overline{S^j}\}$ of measure-preserving transformations. In other words the question of interest is to determine the set of all x (or at least the measure of the set of all x) such that the convergence of $\left\{(1/n)\sum_{i=0}^{n-1} f(S^i x)\right\}$ does or does not take place, where $S^i = \overline{S^i} \circ S^{i-1}$. The first necessity, of course, is to define precisely what a random choice from a family of measure-preserving transformation means.

One way to look at the problem of randomness is to take a measure space Y with normalized measure p, *i.e.*, $\mathrm{p}(Y) = 1$, and examine the space Y^* of all sequences $\{(y_0, y_1, y_2, \dots)\}$, where $y_i \in Y$. Give Y^* the infinite product measure p^*. This means that if $U^* \subset Y^*$, $\mathrm{p}^*(U^*) = \prod_{i=0}^{\infty} \mathrm{p}$ (projection: $U^* \to i^{\text{th}}$ component,) giving $\mathrm{p}^*(Y^*) = 1$. Then given a function which assigns a measure-preserving transformation $\overline{S}_y$ on a measure space X for each $y \in Y$, a typical element $y^* \in Y^*$ allows one to choose a random sequence of $\{S^i_{Y^*}\}$. Simply let $S^i_{Y^*} = \overline{S}_{y_i} \circ \overline{S}_{y_{i-1}} \circ \cdots \circ \overline{S}_{y_0}$. The problem now more explicitly stated is to determine $\mathrm{p}^*(V^*)$, where $V^* = \left\{y_* \in Y^* \;\middle|\; (1/n)\sum_{i=0}^{n-1} f(S^i_{y^*} x)\right\}$ converges. As usual $f \in L_1$ on X.

Now let $Q : Y^* \to Y^*$ be given by a unit shift of index; $Q(y_0, y_1, y_2, \dots) = (y_1, y_2, y_3, \dots)$. Q is measure preserving with respect to the p^* measure because given $U^* \subset Y^*$, $\mathrm{p}\{y_0 \in Y \mid (y_0, y_1, y_2, \dots) \in Q^{-1}(U^*)\} = 1$. Then $\overline{S}_{y_k} = \overline{S}_{y_0^{(Q^k y^*)}}$. By using Q, therefore, $S^i_{y^*}$ becomes $\overline{S}_{y_0^{(Q^i y^*)}} \circ \overline{S}_{y_0^{(Q^{i-1} y^*)}} \circ \cdots \circ \overline{S}_{y_0^{(y^*)}}$. In words this says that the transformations $\{\overline{S}\}$ are selected randomly by taking a specific measure-preserving transformation (Q) on a normalized measure space (Y^*) and defining $\overline{S}$ as a function of the first coordinate. Then all one does to determine $S^i_{y^*}$ is to choose a $y^* \in Y^*$, apply the iterations of Q, select the $\overline{S}_{y_0^{(Q^k y^*)}}$ by the predetermined function, and form the composition.

The next step in reasoning is to generalize Q to be any measure-preserving transformation on an arbitrary normalized measure space Y^*. Let $\overline{S}_{y^*} : X \to X$ be a measure-preserving transformation for each $y^* \in Y^*$, and let $S^i_{y^*} = \overline{S}_{Q^i y^*} \circ \overline{S}_{Q^{i-1} y^*} \circ \cdots \circ \overline{S}_{y^*}$. The only compatibility restriction required on the $\{\overline{S}_{y^*}\}$ is that the mapping from $X \times Y^*$ to X defined by $(x, y^*) \to \overline{S}_{y^*}(x)$ be measurable with respect to the product measure of m (on X) and p* (on Y^*). [Note that the example used in the preceding paragraph conforms to the compatibility restriction. In that case $\overline{S}_{y^*}$ depended only on y_0, which is measurable under p.] With these hypotheses I give the Random Ergodic Theorem (and drop the asterisks to simplify notation.)

Theorem 4.1 (Random Ergodic). *If $f \in L_1$ on X, then for almost every $y \in Y$, $\big\{(1/n)\sum_{i=0}^{n-1} f(S^i_y x)\big\}$ converges for almost every $x \in X$. The limit function f'_y is in L_1.*

Proof. Let $R : X \times Y \to X \times Y$ be given by $R(x, y) = (\overline{S}_y x, Qy)$. Then R is measurable because $(x, y) \to \overline{S}_y x$ and $(x, y) \to Qy$ are measurable. Also, R is measure preserving because Q preserves Y measure, and for any fixed Qy_0, $\overline{S}y_0$ preserves X measure. If $g \in L_1$ on $X \times Y$, then Theorem 2.17 gives that $(1/n)\sum_{i=0}^{n-1} g\big(R^i(x, y)\big)$ converges almost everywhere to $g' \in L_1$, where

$$\int\limits_{X\times Y} g'(x, y)\,\mathrm{d}\,\mathrm{m}(x)\,\mathrm{d}\,\mathrm{p}(y) = \int\limits_{X\times Y} g(x, y)\,\mathrm{d}\,\mathrm{m}(x)\,\mathrm{d}\,\mathrm{p}(y),$$

and m and p indicate the X and Y measures, respectively. By induction, $R^i(x, y) = (S^{i-1}_y x, Q^i y)$. Let $g(x, y) = f(x)$. (Since the measure of Y is 1, $g(x, y) \in L_1$.) Then, $g\big(R^i(x, y)\big) = f(S^{i-1}_y)$. Substituting into the above integration implies the conclusion. □

Although not used in the above proof, Theorem 2.17 also implies that $g'\big(R(x, y)\big) = g'(x, y)$ for almost all (x, y). Then since $f'_y(x) = g'(x, y)$, we have that $f'_y(\overline{S}_y x) = f'_y(x)$, *i.e.*, f'_y is invariant on each $\overline{S}_y$.

As a special case let $Y = [0, 1)$, and express each $y \in Y$ by its binary expansion. For dyadically rational numbers, choose one of the two possible expansions. (For our purpose we could actually consider these two expansions to be distinct because the Lebesgue measure of such points is zero; however, the simplification makes the identification one-to-one.) Let $Q(.y_1 y_2 y_3 \cdots) = (.y_2 y_3 y_4 \cdots)$ be the measure preserving transformation on Y, and let

$$\overline{S}_y = \begin{cases} T_1 & \text{for } 0 \le y < .1 \text{ (one half)} \\ T_2 & \text{for } .1 \le y < 1 \end{cases}$$

where T_1 and T_2 are ergodic transformations. Then $(x, y) \to \overline{S}_y(x)$ is obviously measurable. Under these conditions, Theorem 4.1 gives that for almost all $y \in Y$, *i.e.*, for almost every sequence $(\cdots \circ T_{i_m} \circ \cdots \circ T_{i_3} \circ T_{i_2} \circ T_{i_1})$, $\big\{(1/n)\sum_{i=0}^{n-1} f(S^i_y x)\big\}$ converges for almost every $x \in X$. Since the limit function f'_y is invariant on the ergodic transformations T_1 and T_2, it is a constant almost everywhere. If X happens to have infinite measure, the integrability of f'_y implies that $f'_y = 0$ almost everywhere.

Chapter 5

Some Examples of Ergodic Transformations

Example 5.1. Let X be the space of integers with $\mathrm{m}(U) =$ (number of integers in U), $U \subset X$. If T_1 is the unit translation defined by $T_1(x) = x + 1$, then T_1 is ergodic because the only invariant sets are the empty set and X. The translation T_n defined by $T_n(x) = x+n$ for $n > 1$ is not ergodic, however, because the integers modulo n are invariant.

Example 5.2. As another example let X be the circle group of complex numbers of absolute value one. For a constant $c \in X$, look at the multiplicative transformations defined by $Tx = cx$. If $c^n = 1$ for some positive integer n, then T is not ergodic because $f(x) = x^n = c^n x^n = f(Tx)$ is a non-constant measurable invariant function. If $c^n \neq 1$ for any positive integer n, then T is ergodic. Let $f_n(x) = x^n$ for all integral n. The proof follows readily. Let m be the normalized Lebesgue measure on X, *i.e.*, $\mathrm{m}(X) = 1/(2\pi) \oint |\,\mathrm{d}x| = 1$; then in L_2,

$$(f_n, f_n) = \frac{1}{2\pi} \oint x^n \overline{x^n}\,|\,\mathrm{d}x| = \frac{1}{2\pi} \cdot 2\pi = 1$$

Also, $(f_m, f_n) = 1/(2\pi) \oint x^m \overline{x^n}\,|\,\mathrm{d}x| = 0$ if $m \neq 0$ because $x_i \overline{x_i} = 1$ for all i and $\oint x^i\,|\,\mathrm{d}x| = \oint \overline{x^i}\,|\,\mathrm{d}x| = 0$. Thus $\{f_n\}$ constitutes an orthonormal system in L_2. This system is complete because (I omit the proof) any function can be approximated in the L_2 norm as closely as desired by polynomial functions. Any function $f \in L_2$, therefore, can be represented by $f = \sum_{i=0}^{\infty} a_i f_i$, where the sum converges in L_2 norm. Then given $Rf(x) = f(Tx)$ as the induced unitary operator in L_2, $Rf_n(x) = f_n(Tx) = f_n(cx) = c^n x^n = c^n f_n(x)$. Assume now that f is invariant on X under T. This assumption implies $Rf = \sum_{i=0}^{\infty} a_n c^n f_n = f = \sum_{i=0}^{\infty} a_n f_n$. Thus $a_n c^n = a_n$ for all n. Since $c^n \neq 1$ for $n > 0$, $a_n = 0$ for $n > 0$, *i.e.*, f is a constant. By Corollary 3.4, T is ergodic.

Example 5.3. Consider now the space of sequences $Y^* = \{(y_o, y_1, y_2, \dots)\}$, where $y_i \in Y$, a normalized measure space with measure p. As in the discussion of the Ulam – von Neumann Theorem, let p^* be the infinite product measure on Y^* derived from p, and let Q be the measure-preserving transformation on Y^* defined by $Q(y_0, y_1, y_2, \dots) = (y_1, y_2, y_3, \dots)$. If Q is not ergodic, then there exists a subset $U^* \subset Y^*$ such that $0 < \mathrm{p}(U^*) < 1$ and

$Q(U^*) = U^*$ up to sets of zero measure. In particular, this means that there is a subset $U \subset Y$ such that $U^* = (U, U, U, \dots)$. But if $\mathrm{p}(U) < 1$, then $\mathrm{p}^*(U^*) = 0$, and if $\mathrm{p}(U) = 1$, then $\mathrm{p}^*(U^*) = 1$. Therefore Q is ergodic.

Example 5.4. As a final example I take an ergodic transformation on the real half line. First it is necessary to give a special construction of a space which is a one-to-one measure-preserving copy of the half real line. Let $(a_0, a_1, a_2, \dots)$ be a sequence of positive real numbers such that $a_0 = 1$, $a_i > a_{i+1}$ for all i, and $\sum_{i=0}^{\infty} a_i$ diverges. The harmonic series $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \dots)$ is of this type. Let $X_i = \{(x, y) \in \mathbb{R}^2 \mid 0 \le x < a_i \text{ and } y = i\}$, and let $X = \cup_{i=0}^{\infty} X_i$. In words, X is the real half line broken into half open segments of length a_i, and stacked perpendicular to the positive y-axis. Suppose T_0 is an ergodic transformation on X_0. Define T on X as follows: $T(x, y) = (x, y + 1)$ for all $(x, y) \in X_i$ such that $(x, y + 1) \in X_{i+1}$, and $T(x, y) = T_0(x, 0)$ otherwise. T is measure preserving because T_0 and the projection of $X_i \setminus \{(x, y) \mid a_i \le x < a_{i+1} \text{ and } y = i\}$ onto X_{i+1} are measure preserving.

Let U be an invariant set of X. Look at the set $U_0 = U \cap X_0$. If $(x, 0) \in U_0$, then for each i if $(x, i) \in X_i$, it follows that $(x, i) \in U$. Conversely, if $(x, i) \in U$, then $(x, 0) \in U_0$. Since by hypothesis $a_i > a_{i+1}$ for all i, $T_0(x, 0) = T(x, i)$ for each $(x, 0) \in U_0$, and some i. Inasmuch as U is invariant under T, and T_0 is ergodic, the conclusion that $T_0(x, 0) \in U_0$ for all $(x, 0) \in U_0$ implies that the measure of U_0 or $X_0 \setminus U_0$ is zero. From the relation between U and U_0 it follows immediately that the measure of either U or $X \setminus U$ is zero, hence T is ergodic.

Chapter 6

Conclusion

From its humble beginning as a study of fluid dynamics, ergodic theory has thus extended into a significant field of pure mathematical concern. The general theory, however, is by no means a closed subject. There are many intriguing problems as yet unsolved. Of personal concern are those involving probability and topology in the analysis. The Ulam – von Neumann Theorem was a significant start incorporating probability into the theory. Continuing in that direction, what can be said about the ergodicity of T^n if T is ergodic? This is certainly not a trivial problem. For example, take a space X with three points, each of measure $\frac{1}{3}$. Let $T(a) = b$, $T(b) = c$, $T(c) = a$. T is obviously ergodic as is T^2; however, T^3, the identity, is not. Therefore it is seen that a power of an ergodic transformation may or may not be ergodic. For reasons such as these a group structure, and specifically a Lie group structure, is absent. Other questions can be asked regarding the topology, however. In the above example, of the given measure-preserving transformations two are ergodic, and one is not. This raises the question of denumerability or non-denumerability of ergodic or non-ergodic transformations on spaces of an infinite number of points. No general topological discussion of the spaces of ergodic transformations has yet been written. Questions such as these keep the field active.

Needless to say, the theory as I have presented it here has not included all work done to date. I omitted some theorems because they did not fit into the continuity, and others because they were only combinatoric varieties of those I have given. Where I thought embellishment was instructive I added some results of my own work.

My aim was to give a unified presentation of the fundamentals, while at the same time, give a selective view of the whole theory. The goal will be achieved if through reading this paper others will be stimulated in the pursuit of this subject.

Bibliography

Halmos, P. R. (1956). *Lectures on Ergodic Theory*. New York: Chelsea.

Kakutani, S. (1950). Ergodic theory. In *Proceedings of the International Congress of Mathematicians*, Volume 2, Cambridge, pp. 128–142.

Kolmogorov, A. N. and S. V. Fomin (1961). *Measure, Lebesgue Integrals, and Hilbert Space*. New York: Academic. Translated from the Russian by Natascha Artin Brunswick and Alan Jeffrey.

Riesz, F. (1944, Dec.). Sur la théorie ergodique. *Commentarii Mathematici Helvetici 17*(1), 221–239.

Riesz, F. and B. Sz.-Nagy (1955). *Functional Analysis* (2nd ed.). New York: Frederick Ungar. Translated from the French 2nd edition of *Leçons d'Analyse Fonctionelle* by Leo F. Boron.

Yosida, K. and S. Kakutani (1939). Birkhoff's ergodic theorem and the maximal ergodic theorem. In *Proceedings of the Imperial Academy*, Volume 15, Tokyo, pp. 165–168.

Colophon

In August 2007 the author recomposed this thesis from his original typescript of April 1963. He used the standard LaTeX book class and the $\mathcal{AMS}$ packages of the American Mathematical Society; he compiled the bibliography with BibTeX, employing the University of Chicago Press style. The author also included his own macros for specialized notation and convenience.

The original composition was set in the 11pt Computer Modern fonts of Donald Knuth for an A4 formatted page, the standard for this printing.