1960

The ergodic theorems.

Peterson, Richard Michael

Boston University

http://hdl.handle.net/2144/17838

Boston University
BOSTON UNIVERSITY

GRADUATE SCHOOL

Thesis

THE ERGODIC THEOREMS

by

RICHARD M. PETERSON
(A.B., Boston University, 1957)

Submitted in partial fulfilment of the requirements for the degree of
Master of Arts
1960
AM
1960
Pe
Copy 1

Approved by

First Reader... PHILIP T. MEKEL
Professor of Mathematics

Second Reader... THOMAS SCHID
Professor of Mathematics
Table of Contents

Introduction.............................................. i
1. The Individual Ergodic Theorem.................. I
2. The Mean Ergodic Theorem in $L_p$................. XVII
3. Examples of Stationary Ergodic Processes...... XXIV
4. Bibliography
Introduction

The development of ergodic theory has its roots in questions which arose in the field of statistical mechanics in the early part of the twentieth century. The problem as formulated in statistical mechanics is essentially concerned with a general dynamical system described by the differential equations \( \frac{dx_i}{dt} = X_i(x_1, x_2, \ldots, x_n), \ i = 1, 2, \ldots, n \) which hold for an \( n \) dimensional closed manifold \( M \). The functions \( X_i \) are to be analytic, but otherwise arbitrary. A model for the system is easily realized if we consider points of \( M \) to be moving according to the given equations of motion relative to a stationary \( n \) dimensional coordinate system. Then the velocity of a moving point of \( M \) in the \( x_i \) direction is given by \( \frac{dx_i}{dt} = X_i(x_1, x_2, \ldots, x_n) \) at the time when the point is at \( (x_1, x_2, \ldots, x_n) \) of the stationary coordinate system. Now the solution of this system of \( n \) differential equations would establish the path of a particular moving point corresponding to a particular choice of the arbitrary constants which arise in the solution of this system of equations. We call such a path the trajectory of the particular point and so the set of all trajectories will be the collection of all paths corresponding to all possible choices of these constants, i.e., the set of all trajectories followed by all the moving points.

The work of Poincare(1) which is the initial step in
the development of ergodic theory begins by considering an $(n-1)$ dimensional surface $\Sigma$ in the manifold $M$ in which the above system of equations holds without singularity. Also we assume that the volume integral $\int \cdots \int dx_1 dx_2 \cdots dx_n$ in $M$ is invariant with respect to choices of the coordinates $x_1, x_2, \ldots, x_n$ and that all trajectories of moving points in $M$ cut $\Sigma$ in one and the same sense at an angle $\theta \geq d > 0$.

The reader will note that it is indeed possible to choose $\Sigma$ such that the latter requirement is met. In particular we are interested in the trajectories of points which initially, i.e., at a certain time $t$ are points of $\Sigma$. If we allow time to increase indefinitely, points of $\Sigma$ initially will move to new positions along their trajectories in accordance with the above system of equations. After some lapse in time it is clear that some points initially of the surface $\Sigma$ will not belong to $\Sigma$ whereas others may at varying instances of time belong to $\Sigma$, i.e., the trajectories of these latter points again cut $\Sigma$.

The classic work of Poincare(1) is a statement to the effect that if time is allowed to increase or decrease indefinitely all points of $\Sigma$ initially, excepting a set of measure 0, will again belong to $\Sigma$ an infinite number of times. An alternative way of stating this result is that the trajectories of almost all points of $\Sigma$ initially will cut $\Sigma$ an infinite number of times with probability 1. If this is so, then the natural question arises: If a point $x$
returns to $\sigma$ an infinite number of times then what is the probability, if such a definite probability exists, that we will find $x \in \sigma$ at a particular instant of time $t$? That is, what proportion of time does $x$ spend in $\sigma$ as compared with the total time? The body of this paper is concerned with the various existing mathematical formulations of this problem which led to a solution, at least in some sense, of it.
We may expect in general given such a system of differential equations that if a subset $S$ of $M$ is measurable and invariant with respect to time then $S$ has measure $0$ or $S$ has measure equal to the volume $V$ of $M$. This would imply in terms of trajectories that every measurable set of complete trajectories in $M$ has the measure $0$ or $V$. This property of a dynamical system called strong transitivity is indeed very difficult to verify for a specific system. Simple examples of strongly transitive systems do exist, the first example being given by E. Hopf(2).

The work of Poincare was carried on by von Neumann(3) and G. D. Birkhoff(4). Because the work of Birkhoff follows in a natural sequence that of Poincare, we will reserve remarks concerning von Neumann's work until later. The statement that the trajectory of a point of $\mathcal{S}$ initially cuts $\mathcal{S}$ an infinite number of times raises the question of the existence of a definite frequency with which the trajectories cut $\mathcal{S}$. Or equivalently, we may ask for the existence of an average amount of time between successive crossings of $\mathcal{S}$ by a point $P$. Letting $t_n$ be the time lapse between the time that $P$ is on $\mathcal{S}$ initially and the time of the $n^{th}$ crossing of $\mathcal{S}$ by $P$, Birkhoff seeks to establish the existence of $\lim_{n \to \infty} t_n(P)$. Letting $T$ be the transformation which takes one point $P$ issuing from $\mathcal{S}$ into another point $\overline{P}$ where $\overline{P}$ is the...
intersection of $\sigma$ and the first crossing again of $\sigma$ by
the trajectory issuing from $P$ on $\sigma$, we find that $T$ is
one-one, since every point of $\sigma$, excepting a set of measure
0, again cuts $\sigma$ an infinite number of times and therefore a
first time according to the work of Poincare. The transfor-
mation $T^k(P)$ will be taken to mean $T(T^{k-1}(P))$. Moreover,
the volume generated during a period of time $dt$ by an
element of area $d\sigma$ which initially issued from $\sigma$ under
the transformation $T$ is the same regardless of the choice of
the time $t$ at which $t$ advances by $dt$. This result follows by
hypothesis that volumes are conserved. The reader will note
that this condition in the hypothesis is that of incompress-
ibility. We denote the volume generated during $dt$ by
$dV = dP \cdot dt$ where $P$ is a representative point of $d\sigma$. The
identity $t_n(P) = t(T^{n-1}(P)) + t_{n-1}(P)$ is easily observed to
hold and by repeated applications of this identity we have
$t_n(P) = t(T^{n-1}(P)) + t(T^{n-2}(P)) + ... + t(P)$ so that
\[
\int_\sigma t_n(P)dP = \int_\sigma t(T^{n-1}(P))dP + \int_\sigma t(T^{n-2}(P))dP + ... + \int_\sigma t(P)dP,
\]
where $dP$ is the volume element above. Now
$t(P)dP$ represents the volume generated by the trajectories
of points belonging to $d\sigma$ which move from $\sigma$ initially to
$\sigma$ again for the first time. It is therefore clear that
\[
\int_\sigma t(P)dP \text{ represents the volume generated by the trajectories of all points initially of } \sigma \text{ which move to } \sigma \text{ again for the first time. But assuming as Birkhoff does that the system is strongly transitive the measure of this set of}
complete trajectories must be \( V \), the total volume of \( M \), so
\[
\int_{\sigma} t(P)\,dP = V.\]
Replacing \( t(P) \) by \( t(T(P)) \) the same argument goes through again because \( T \) transforms \( \sigma \) into itself except for a set of measure 0, according to the work of Poincare. Thus we have
\[
\int_{\sigma} t(T^k(P))\,dP = \int_{\sigma} t(T^{k-1}(P))\,dP = \ldots = V\]
so that
\[
\int_{\sigma} t_n(P)\,dP = nV \quad \text{and dividing by } n \cdot \int_{\sigma} \,dP, \]
\[
\frac{\int_{\sigma} t_n(P)\,dP}{n \cdot \int_{\sigma} \,dP} = \frac{V}{\int_{\sigma} \,dP} = \alpha.\]
The average time of the \( n \)th crossing of \( \sigma \) is equal to the constant \( \alpha \). This does not, however, prove that there exists a mean time for \( n \) infinite.

To prove the existence of \( \alpha \) for \( n \) infinite Birkhoff considers the set \( S_\delta \) on \( \sigma \) of points \( P \) for which \( t_n(P) > n(\alpha + \delta) \) for an infinite number of values of \( n \) and \( \delta > 0 \). Imposing the condition of strong transitivity we find that the measure of this set of points is 0 or the measure of \( \sigma \), since the measure of any set of complete trajectories in \( M \) is either 0 or \( V \) and the measure of the set of complete trajectories issuing from points of \( S_\delta \) is \( \int_{S_\delta} t(P)\,dP \).

The above remark which follows from the condition of strong transitivity is also applicable to the set \( S_\delta' \) of point \( P \) which satisfy \( t_n(P) < n(\alpha + \delta) \) for an infinite number of values of \( n \). Birkhoff proceeds to break up the set \( S_\delta \) into a sequence \( \{S_\delta, k\} \) of disjoint sets \( S_\delta, k \) showing that if \( S_\delta \) has the measure \( \sigma \) then the limit \( S_\delta, \infty \) of this sequence of sets has measure \( \sigma \). The inequality
\[
\int_{S_\delta, k} t(P)\,dP > (\alpha + \delta) \int_{S_\delta, k} \,dP
\]
is established for sets \( S_\delta, k \) in the
sequence and therefore allowing $k$ to go to infinity the inequality $\int_{S_{\delta,\infty}} t(P) dP \geq (\alpha + \delta) \int_{S_{\delta,\infty}} dP$ results which is impossible for $\delta > 0$ because we know that the measure of $\sigma$ is equal to the measure of $S_{\delta,\infty}$ and that $\int_{\sigma} t(P) dP = \alpha \int dP$.

Having proved the existence of a mean for strongly transitive systems, Birkhoff(5) sets out to prove the more general result that a mean $\tau(P)$ exists for systems where the condition of strong transitivity does not necessarily hold. The basis of this proof is a lemma which states that if for any point $P$ of a measurable and invariant with the usual exception (with respect to $T$) set $S_\lambda$ it is true that

$$\lim_{n \to \infty} \sup \frac{t_n(P)}{n} > \lambda > 0,$$

then $\int_{S_\lambda} t_n(P) dP \geq \lambda \int_{S_\lambda} dP$.

This lemma is easily established if we note that in the statement given above concerning point $P$ which satisfied the inequality $t_n(P) > n(\alpha + \delta)$ it followed that

$$\int_{S_{\delta,k}} t(P) dP \geq (\alpha + \delta) \int_{S_{\delta,k}} dP$$

and that in the limit

$$\int_{S_{\delta,k}} t(P) dP \geq (\alpha + \delta) \int dP.$$ We may in the same way break up $S_\lambda$ into disjoint sets $S_{\lambda,k}$ for which it would follow in the same way for points $t_n(P) > n(\lambda - \epsilon)$, $\epsilon > 0$ arbitrary, that

$$\int_{S_{\lambda,k}} t(P) dP \geq (\lambda - \epsilon) \int_{S_{\lambda,k}} dP$$

and therefore

$$\int_{S_{\lambda,k}} t(P) dP \geq (\lambda - \epsilon) \int_{S_{\lambda,k}} dP.$$ The condition $t_n(P) > n(\lambda - \epsilon)$ is satisfied by all points of $S_\lambda$ so that for $\epsilon$ arbitrary we have $\int_{S_{\lambda,k}} t(P) dP \geq \lambda \int_{S_{\lambda,k}} dP$. 

and the lemma follows because $t_n(P) \geq t(P)$, $(n = 2, 3, \ldots, n)$.

To prove the existence of the mean we consider the measurable invariant sets $S_{\lambda}$ and $S'_{\lambda}$ on $\sigma$, where $S_{\lambda}$ consists of those points for which $t_n(P) \geq n \lambda$ for an infinite number of values of $n$ and $S'_{\lambda}$ consists of those points for which $t_n(P) < n \lambda$ for an infinite number of values of $n$. It is clear that $S_{\lambda}$ decreases when $\lambda$ increases and that $S'_{\lambda}$ increases when $\lambda$ increases. It is important to note also that for some particular values of $\lambda = \lambda_i$ a set of points of measure greater than 0 may be common to both $S_{\lambda}$ and $S'_{\lambda}$. Denoting this set of points by $S^*_i$, we see that this is possible if and only if $\limsup_{n \to \infty} t_n(P) \leq \lambda_i$ and $\liminf_{n \to \infty} t_n(P) \geq \lambda_i$ for all $P \in S^*_i$. This shows that for points common to both sets $S_{\lambda}$ and $S'_{\lambda}$ for definite values of $\lambda$, there exists a mean time.

For values of $\lambda$ other than these $\lambda_i$ we know then that at most a set of points of measure 0 can belong to both $S_{\lambda}$ and $S'_{\lambda}$. Let two such values of $\lambda$ be designated by $\lambda^* > \lambda^*$. The sets $S_{\lambda^*}$ and $S_{\lambda^*}$ are then clearly defined as is the set $S_{\lambda^*} - S_{\lambda^*}$ consisting of points $P$ for which $t_n(P) \geq n \lambda^*$, but $t_n(P) < n \lambda^*$ for an infinite number of values of $n$. It follows then that $\lambda^* \leq \limsup_{n \to \infty} t_n(P) \leq \lambda^*$ and that $\lambda^* \leq \liminf_{n \to \infty} t_n(P) \leq \lambda^*$ for points $P$ of $S_{\lambda^*} - S_{\lambda^*}$. This, however, would imply from the lemma above that $\int_{S_{\lambda^*} - S_{\lambda^*}} t_n(P) dP \geq \lambda^* \int_{S_{\lambda^*} - S_{\lambda^*}} dP$ and that
\[ \int_{S^{x^*}} t_n(P) dP \leq \lambda \int_{S^{x**}} dP \] which is impossible if \( x^{**} > x^* \) and \( S^{x^*} - S^{x**} \) is of measure greater than 0.

We have to conclude then that the measure of \( S^{x^*} - S^{x**} \) is 0 for all choices of \( x^{**} > x^* \) which allows only the possibility that \( x^{**} \) and \( x^* \) are not only equal but are equal to a definite value \( \lambda \), i.e., \( x^* = x^{**} = \lambda \). It follows then that the \( \lim_{n \to \infty} \sup \frac{t_n(P)}{n} = \lim_{n \to \infty} \inf \frac{t_n(P)}{n} = \lambda \) for any point \( P \) of \( \sigma \), excepting a set of measure 0.

In the proof of the ergodic theorem it is necessary to point out that a set of surfaces \( \sigma^* \) may be found such that the trajectory of every moving point cuts \( \sigma^* \). Consider now any volume \( G \) in \( M \) and denote by \( \bar{T}(P) \) the time spent by \( P \) issuing from \( \sigma^* \) in the volume \( G \) before \( P \) again for the first time cuts \( \sigma^* \), that is the time interval spent in \( G \) before \( P \) reaches \( T(P) \). It is clear that the functional relation \( \bar{T}(P) = \bar{T}(T^{-1}(P)) + \bar{T}_{n-1}(P) \) is satisfied by \( \bar{T}_n(P) \) and so by the development which resulted in the existence of the limit \( \lim_{n \to \infty} \bar{T}_n(P) = \bar{\tau}(P) \) we may prove the existence of \( \lim_{n \to \infty} \bar{\tau}_n(P) = \bar{\tau}(P) \). It is obvious that \( \bar{\tau}(P) \leq \bar{\tau}(P) \), so there is a definite probability \( p \leq 1 \) that a point \( P \) will be in the volume \( G \) at any particular instance of time. If the particular system under consideration is strongly transitive the probability \( p \) is equal to the volume of \( G \) divided by the volume of \( M \) because in this case the only proper subset of \( M \) which is invariant with respect to \( T \) is the set of measure 0. This is essentially the extent of G. D. Birkhoff's work in
ergodic theory.

It is important to realize that in proving this theorem Birkhoff assumed the system to be a reversible one. This is easily seen to be the case if we recall that the transformation $T$ which transformed points $P$ of $U$ into points $T(P)$ of $U$ was assumed to be one-one in $U$ except for a set of measure 0. That is, the work of Poincare and G. D. Birkhoff is concerned only with systems for which $T^{-1}$ exists and $T^{-1}(Tx) = x$. A generalization of Birkhoff's theorem which includes the irreversible case has been given by E. Riesz(6). We begin with a few definitions.

Let $X$ be an arbitrary non-empty point set on which the measure $\mu$ is defined and let $T$ be a measure preserving transformation of $X$ into itself. We do not assume that $T$ is invertible so that $T^{-1}(y)$ will be taken to mean the inverse image of $y$, i.e., we do not assume that $T^{-1}(y) = x$ when $Tx = y$. Furthermore, we will let $g(x) = f(Tx)$ where $g(x) = Uf(x)$ so that $U$ is the transformation of $f(x)$ to $f(Tx)$. The properties that $U$ is linear, that $f(x)$ nonnegative implies that $Uf(x)$ is nonnegative and that $|g(x)| = U(|f(x)|)$ are easily verified. Finally we will show that $U$ is norm preserving in the case that the norm of $f$ is defined by $\|f\|_p = \left( \int_X |f|^p d\mu \right)^{1/p}$ for any measurable $f(x)$ and $1 \leq p < \infty$. Let $E$ be any measurable subset of $X$ and let $f = \chi_E$, the characteristic function of $E$. If $g(x) = f(Tx)$ then $g(x) = \chi_{T^{-1}(E)}(x)$ since if
If $x \in E$ then $x \in T^{-1}(E)$. Substituting $\chi_{T^{-1}(E)}$ into the expression $\left( \sum \chi \right)^{1/p} = \|g\|_P$, we find $\|g\|_P^p = \mu(T^{-1}(E))$. But $T$ is measure preserving so that $\|g\|_P^p = \mu(T(E)) = \mu(E) = \|f\|_P^p$ for $f$ equal to the characteristic function $\chi_E$. It follows immediately by the definition of a norm and consideration of step functions that $\|g\|_P = \|f\|_P$ for every measurable $f$ assuming only a finite number of nonnegative values. We consider now any $f$ which is measurable and nonnegative. We know now from the theory of integration that if we take a sequence of measurable functions $f_n \uparrow f$ each assuming only a finite number of nonnegative values that $\|f_n\|_P = \|f\|_P$ as $n \to \infty$. But corresponding to the sequence $\{f_n\}$ there exists the sequence $\{g_n\}$ where $g_n = Uf_n$ and $\|g_n\|_P = \|f_n\|_P$ for every $n$. Therefore $\|g\|_P = \|f\|_P$ and this shows that the transformation $U$ is isometric.

In order to prove the individual ergodic theorem we will establish two preliminary results. First we will show that the sum of the $m$ leaders in a sequence $a_1, \ldots, a_n$ is nonnegative. A number $a_k$ in the sequence is called an $m$ leader if $a_k + a_{k+1} + \ldots + a_{k+p-1}$ is nonnegative for some $p$, $1 \leq p \leq m$. If there are no $m$ leaders then the sum of the $m$ leaders is $0$ and therefore nonnegative. If these are $m$ leaders let the first one be $a_k$ and let the shortest non-
negative sum for which $a_k$ is an $m$ leader be $a_k + \ldots + a_{k+p-1}$. Then we claim that every $a_i$, $i = k, \ldots, k + p - 1$ is an $m$ leader for if this is not so for such an $a_i$ then $a_i + a_{i+1} + \ldots + a_{k+p-1}$ would be negative and therefore $a_k + a_{k+1} + \ldots + a_{i-1}$ would be positive which is impossible since $a_k + \ldots + a_{k+p-1}$ is the shortest nonnegative sum for which $a_k$ is an $m$ leader. Next we consider the remaining terms $a_{k+p}, \ldots, a_n$ in the sequence and determine the first $m$ leader $a_1$ in this sequence. Repeating the above procedure we find that all terms in the shortest nonnegative sum for which this $m_1$ is an $m$ leader are also $m$ leaders. It follows that the sum of all $m$ leaders in the sequence $a_1, a_2, \ldots, a_n$ is a sum of nonnegative sums and therefore nonnegative.

We undertake now to prove the maximal ergodic theorem from which it will be possible to prove the individual ergodic theorem. Consider the sequence $f(x), f(Tx), f(T^2x), \ldots, f(T^{m+n-1}x)$, $f$ is $\mu$-summable, of functional values for the positive integers $m$ and $n$ fixed and let $D_k$ be the set of all points $x \in X$ for which $f(T^kx), 0 \leq k \leq n + m - 1$ is an $m$ leader. For a particular value of $x$ the sum of the $m$ leaders is given by $s(x) = \sum_{k=0}^{n+m-1} f_k(x) \chi_{D_k}(x)$ and this sum $s(x)$ is nonnegative for all $x \in X$ by the preceding result, so that $\int_X s(x)dx \geq 0$, or $\sum_{k=0}^{n+m-1} f_k(x) \chi_{D_k}(x)dy = \sum_{k=0}^{n+m-1} \int_{D_k} f_k(x) \geq 0$. For $k = 0$ we find that $D_k$ is the set $D_0$ of all $x$ for which $f(x)$ is an $m$ leader which means that
\( f(x) + f(Tx) + \ldots + f(T^{P-1}x) \geq 0 \) for some \( p \leq m \). Similarly \( D_1 \) is the set of all \( x \) for which \( f(Tx) + f(T^2x) + \ldots + f(T^{P}x) \geq 0 \). But the latter expression implies that \( Tx \in D_0 \) for all \( x \in D_1 \) so that \( x \in T^{-1}(D_0) \). Letting \( y = T^{k-1}(x) \) and \( Ty = T^kx \) we find in the same way that \( D_k = T^{-1}(D_{k-1}) \) provided \( k \) does not exceed \( n - 1 \) since we have only \( n + m - 1 \) terms in the sequence of functional values. Therefore, for these values of \( k \) we have \( D_k = T^{-k}(D_0) \). Recalling that the function \( U(f(x)) = f(Tx) \) is norm preserving we find that

\[
\int_{D_k} f(T^kx) \, d\mu = \int_{T^{-k}(D_0)} f(Tx) \, d\mu = \int_X f(T^kx) \chi_{D_0}(T^kx) \, d\mu = \int_X f(x) \chi_{D_0}(x) \, d\mu = \int_{D_0} f(x) \, d\mu.
\]

We denote the set \( D_0 \) of all \( x \) for which \( f(x) \) is an \( m \) leader by \( E_m \) so that \( \int_{D_k} f(T^kx) \, d\mu = \int_{E_m} f(x) \, d\mu \) for \( k = 0, 1, \ldots, n - 1 \). Now for the \( m \) leaders \( f(T^kx) \) where \( k = n, \ldots, n + m - 1 \) it is true that if \( x \in D_k \) then \( Tx \in D_{k-1} \); but these \( Tx \in D_{k-1} \) where \( x \in D_k \) do not exhaust \( D_{k-1} \) so that we may not write \( D_k = T^{-1}D_{k-1} \), however it is always true that

\[
\int_{D_k} f_k(x) \, d\mu \leq \int_X |f_k(x)| \, d\mu
\]

and by the property of preservation of the norm that \( \int_X |f(x)| \, d\mu = \int_X |f_k(x)| \, d\mu \). Substituting now into the expression for the sum of the \( m \) leaders in the sequence \( f(x), \ldots, fT^{n+m-1}(x) \) we have
\[
\sum_{k=0}^{n+m-1} \int_{D_k} f^k(x) d\mu = n \int_{E_m} f d\mu + m \int_X |f| d\mu \geq 0.
\]

Dividing by \( n \) and letting \( n \to \infty \) we find that \( \int_{E_m} f d\mu \geq 0 \).

If we now allow \( m \) to take all possible values, that is if we allow \( m \) to go to infinity, then it is clear that \( E_m \uparrow E \)
where \( E \) is the set of \( x \) for which \( f(x) \) is an \( m \) leader in the
infinite sequence \( f(x), f(Tx), \ldots, f(T^nx), \ldots \) . Equivalently we may state that if the set \( E \) is the set of all \( x \in X \)
where one at least of the sums \( \sum f_0(x) + \ldots + f_n(x) \) is nonnegative then \( \int_E f d\mu \geq 0 \).

With the aid of the maximal ergodic theorem we propose
now to prove the individual ergodic theorem for noninvertible transformations. The reader will note that the proof
is given along the lines used by Birkhoff(5) to prove the
theorem for the invertible case.

We assume that the function \( f(x) \) belongs to \( L_1(X,\mu) \),
i.e., \( f(x) \) is of \( \sigma \) finite measure, and that \( T \) is as before
a measure preserving transformation of \( X \) into itself. The
individual ergodic theorem establishes the existence the limit of \( s_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^kx) \) as \( n \to \infty \) for almost all \( x \in X \).

If \( f(T^kx) = \chi_{E}(T^kx) \) it is easy to see that the problem is
reduced to that considered by Birkhoff since in this case
we would be proving the existence of an average time spent
by \( x \) in the subset \( E \) as \( x \) is transformed an infinite number
of times. We begin by considering the subsets \( X_k = \{ x \mid f(T^k x) \neq 0 \} \) which are all of \( \sigma \)-finite measure since \( X_0 = \{ x \mid f(x) \neq 0 \} \) is of \( \sigma \)-finite measure and \( X_k = T^{-k}(X_0) \).

It follows that the sum \( X' = \sum_{k=0}^{\infty} X_k \) is of \( \sigma \)-finite measure and that on \( X - X' \) all \( f(T^k x) \) vanish so that all iterates

\[
s_n(x) = \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)
\]

vanish and therefore \( s_n(x) \to 0 \) on \( X - X' \). We may consider then for our proof the set \( X' \) instead of \( X \) since convergence is already established on \( X - X' \).

Let us suppose now that there exists a set \( Y \subset X' \) for which \( \lim \inf s_n(x) < a < b < \lim \sup s_n(x) \) where \( a < b \) are real numbers. The set \( Y \) is invariant since \( \lim \inf s_n(Tx) = \lim \inf s_n(x) \) and \( \lim \sup s_n(Tx) = \lim \sup s_n(x) \) and therefore \( Y = T^{-1}(Y) \) which means that \( Y \) may now replace \( X \) in any applications of the maximal ergodic theorem since it too, like \( X \), is invariant with respect to the transformation \( T \).

We may assume that \( b > 0 \) since otherwise \( a < 0 \) and the same proof using \( -f(x) \) and \( -a > 0 \) instead of \( f(x) \) and \( b \) would go through. This follows from the fact that the inequalities \( \lim \inf s_n(x) < a < b < \lim \sup s_n(x) \) would reverse upon multiplication by \( -1 \). We define the function

\[
g(x) = f(x) - b \chi_Z(x)
\]

on any arbitrary finite measurable subset \( Z \) of \( Y \). Now because \( \lim \sup s_n(x) > b \) for \( x \in Y \) it follows that for some \( n = n_0 \), \( s_n(x) > b \chi_Y(x) \) or

\[
\sum_{k=0}^{n_0-1} g(T^k x) = \sum_{k=0}^{n_0-1} \{ f(T^k x) - b \chi_Z(T^k x) \} =
\]
\[ \sum_{k=0}^{n-1} \{ f(T^k x) - b \chi_{T^{-k}(Z)}(x) \} \geq \sum_{k=0}^{n-1} \{ f(T^k x) - b \chi_{Y}(x) \} > 0 \]

since \( \chi_{T^{-k}(Z)}(x) \) is 0 at least as often as \( \chi_{Y}(x) \). Applying the maximal ergodic theorem we have

\[ \int (f - b \chi_{Z}) d\mu \geq 0 \]

so that \( \mu(Z) \leq \frac{1}{b} \int f d\mu \leq \frac{1}{b} \int f d\mu \).

We may now choose a sequence \( Z_k \subset Y \) of subsets each satisfying the above inequality which converge to the set \( Y \). It follows then that

\[ \mu(Y) = \lim \mu(Z_k) \leq \frac{1}{b} \int f d\mu \]

and for \( f \) real and finite, which we may assume without loss of generality in the proof, it turns out that \( \mu(Y) \) is finite. Since all conditions demanded of the subset \( Z \) are now fulfilled by \( Y \), we have, replacing \( Z \) by \( Y \),

\[ \int \{ f(x) - b \} d\mu \geq 0. \]

Carrying through the same argument with respect to the function \( g_1(x) = a - f(x) \) we would find that

\[ \int (a - f(x)) d\mu \geq 0 \]

which implies that \( \int (a - b) d\mu \geq 0 \) and this is impossible if \( b > a \). We therefore must conclude that \( \mu(Y) = 0 \).

It is true then that \( \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \) converges to a limit function \( f^*(x) \) for all \( x \in X \) with the usual exception. In addition it is evident that \( f^*(Tx) = f^*(x) \) almost everywhere and by use of Fatou's lemma we have

\[ \int f^* d\mu = \liminf \int |s_n| d\mu \leq \int |f| d\mu \]

since \( \int |s_n| d\mu = \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k x)| d\mu \), so

\[ \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k x)| d\mu \leq \frac{1}{n} \sum_{k=0}^{n-1} |f(T^k x)| d\mu = \int |f| d\mu, \]

that \( f^* \) is summable. To prove that

\[ \int f^* d\mu = \int f d\mu. \]
when \( \mu(X) \) is finite we note that the set \( X(p,n) = \{ x : \frac{p}{2^n} \leq f^* \leq \frac{p+1}{2^n} \} \) for \( p \) an arbitrary integer is invariant because \( f^* \) is invariant with respect to \( T \). Furthermore for any \( x \in X(p,n) \) and \( \epsilon > 0 \), it is true that one at least of the sums \( \sum_{k=0}^{n-1} \{ f(T^k x) - (\frac{p}{2^n} - \epsilon) \} \) is nonnegative since

\[
\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) \text{ converges to } f^* \text{ and therefore by the maximal ergodic theorem } \int_{X(p,n)} f \, d\mu \geq (\frac{p}{2^n} - \epsilon) \cdot \mu(X(p,n)) \text{ so that }
\]

\[
\int_{X(p,n)} f \, d\mu \geq \frac{p}{2^n} \cdot \mu(X(p,n)) \cdot \frac{p+1}{2^n} \, \mu \{ X(p,n) \} \text{ so that } \frac{p}{2^n} \cdot \mu \{ X(p,n) \} \leq \int_{X(p,n)} f \, d\mu \leq \frac{p+1}{2^n} \cdot \mu \{ X(p,n) \} . \text{ The same inequalities of course hold for } f^* \text{ and by subtraction we obtain } - \frac{1}{2^n} \mu \{ X(p,n) \} \leq \int_{X(p,n)} (f - f^*) \, d\mu \leq \frac{1}{2^n} \cdot \mu \{ X(p,n) \} . \text{ Summing now the above inequalities corresponding to different values of } p \text{ we have } \int_{X} (f - f^*) \, d\mu \leq 2^{-n} \mu(X). \text{ It follows that }
\]

\[
\int f^* \, d\mu = \int f \, d\mu . \text{ We may also show that }
\]

\[
\lim \int |f^* - s_n| \, d\mu = 0. \text{ We consider the case of } f(x)
\]
bounded separately from the case when \( f(x) \) is not bounded. Let \( f(x) \) be bounded then \( f(T^kx) \) is bounded so that
\[
\frac{1}{n} \sum_{k=0}^{n-1} f(T^k(x)) \text{ is bounded for any } n.
\]
By Lebesque's theorem on dominated convergence it follows then that
\[
\lim \int s_n d\mu = \int f^* d\mu \text{ if } \mu(X) \text{ is bounded.}
\]
If \( f(x) \) is not bounded we know that there exists a bounded function \( g(x) \) such that
\[
||f - g||_1 < \epsilon, \epsilon > 0, \text{ since } f(x) \text{ is summable.}
\]
Adding and subtracting \( g(T^kx) \) and \( g^*(x) \) in the above expression we have
\[
||s_n - f^*||_1 \leq ||\frac{1}{n} \sum_{k=0}^{n-1} (f(T^kx) - g(T^kx)) ||_1 +
\]
\[
||\frac{1}{n} \sum_{k=0}^{n-1} g(T^kx) - g^*||_1 + ||g^* - f^*||_1 \leq ||\frac{1}{n} \sum_{k=0}^{n-1} (g(T^kx) - g^*)||_1
\]
\[
+ 2 \epsilon.
\]
Choosing \( n \) so that
\[
||\frac{1}{n} \sum_{k=0}^{n-1} (g(T^kx) - g^*) ||_1 < \epsilon
\]
we have
\[
||s_n - f^*||_1 \leq 3 \epsilon \text{ and therefore } \lim \int |f^* - s_n| d\mu = 0.
\]
This completes the proof of the individual ergodic theorem and properties which characterize the function \( f^* \).

In order to prove the additional result that for strongly transitive systems the space mean is equal to the time mean we will show that in the case of strong transitivity every measurable invariant function \( f(x) \) is (almost) equal to a constant. We consider the set \( X(p,n); p/2^n \leq f(x) < \frac{p+1}{2^n} \) where \( p \) and \( n \) are defined as above and \( f(x) \) is measurable and invariant. Because \( f(x) \) is invariant it follows that the set \( X(p,n) \) is invariant with respect to \( T \). But by the condition of strong transitivity the measure
\[ \mu \{ X(p,n) \} = 0 \] or \[ \mu \{ X - X(p,n) \} = 0. \] Now for a fixed value of \( n \) we may write \( X = X_\infty + X_{-\infty} + \sum_{p=-\infty}^{\infty} X(p,n) \) where \( X_\infty = \{ x : f(x) = \infty \} \) and \( X_{-\infty} = \{ x : f(x) = -\infty \} \) and by the condition of strong transitivity all these sets which compose \( X \) are of measure 0 except one. We denote this one by \( X(p_n,n) \) since for different values of \( n \) it is conceivable that \( X(p_n,n) \) will represent different sets in \( X \). But if the measure of \( X(p,n) \) is not 0 then the measure of its complement is 0 and therefore the measure of the complement of \( X_\infty = \bigcup_{n=1}^{\infty} X(p_n,n) \) is also 0. By the definition of the sets \( X(p,n) \) it is obvious that \( f(x) \) is constant on \( X_\infty \) so that \( f(x) \) is constant almost everywhere on \( X \) since \( \mu(X - X_\infty) = 0. \)

The desired result for strongly transitive systems follows immediately. We found from the individual ergodic theorem that the limit function \( f^*(x) \) is summable and invariant and because of strong transitivity equal almost everywhere to a constant. For \( \mu(X) = \infty \) this constant is 0 for \( f(x) = 0 \) is the only summable constant in this case. For \( \mu(X) < \infty \), we may use the fact that \( \int_X f^* d\mu = \int_X f d\mu \) or \( f^* = \frac{\int_X f d\mu}{\int_X d\mu} \). It has become customary to call a system which is strongly or metrically transitive an ergodic system or more precisely to call the measure preserving transformation \( T \) in this case an ergodic transformation.
The Mean Ergodic Theorem in $L_p$

The mean or "statistical" ergodic theorem came as a direct result of the work of J. von Neumann (7) and M.H. Stone (8) on the spectral resolution of unitary operators in Hilbert space. The mean ergodic theorem establishes the existence of the limit of $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ as $n \to \infty$, where $f$ is a function in the Hilbert space $L_2(X, \mu)$. Convergence is of course required with respect to the $L_2$ norm $\|f\|_2 = \left( \int |f|^2 \, d\mu \right)^{1/2}$, i.e., we seek a function $\bar{f} \in L_2$ such that $\left( \int \left| \frac{1}{n} \sum_{k=0}^{n-1} f(T^k x) - \bar{f} \right|^2 \right)^{1/2} \to 0$ as $n \to \infty$.

The proof given by J. von Neumann (3) deals only with the special case of invertible transformations. The proof we will give here following Zaanen (9) is due to F. Riesz and B. Sz-Nagy (10) and is not confined to the special case of invertible transformations. We begin by recalling that in the work which resulted in the individual ergodic theorem we established the fact that the transformations $f(x) \to f(Tx)$ is norm preserving for any $f \in L_p$ provided $1 \leq p < \infty$. Then the transformation $f(x) \to f(Tx)$ is isometric for the particular case of $f \in L_2$ so that we can prove the mean ergodic theorem by showing the existence of the limit of $\frac{1}{n} \sum_{k=0}^{n-1} U^k f$ for any isometric transformation $U$ of $L_2$ into itself. We begin with a number of properties of functions in Hilbert space.

A Hilbert space $H$ is a Banach space with the norm
of elements \( x \in H \) defined by \( \| x \| = (x, x)^{1/2} \) where \( (x, y) \)
denotes the inner product of \( x \) and \( y \). If for two elements \( x, y \in H \) it is true that \( (x, y) = 0 \), we say that the elements \( x, y \) are orthogonal. Furthermore given a closed subspace \( M \subset H \) we will denote the closed subspace of all \( y \in H \) for which it is true that \( (x, y) = 0 \) for all \( x \in M \) by \( M^\perp \). \( M^\perp \)
is called the orthogonal complement of \( M \) and it is true that for each \( f \in H \) a unique decomposition \( f = f_1 + f_2 \) exists
where \( f_1 \in M \) and \( f_2 \in M^\perp \). The projection operator
\( P : Pf \rightarrow f_1 \) is called the orthogonal projection on \( M \) and obviously satisfies \( P^2 = P \) and \( \| Pf \| \leq \| f \| \). Considering
now the isometric transformation \( U \) we define the adjoint transformation of \( U \) as the transformation \( U^* \) which satisfies
\( (Ux, y) = (x, U^*y) \) for all \( x, y \in H \). It can be proved that \( U^* \)
is unique and that \( \| U^* \| = \| U \| \).

To prove the mean ergodic theorem we consider functions
\( f \in H \) of the form \( f = g - Ug \) where \( g \) is a function of \( H \).
For these functions \( f \) we have \( \sum_{k=0}^{n-1} U^k f = (g - Ug) + (Ug - U^2g) \)
\[ \vdots + (U^{n-1}g - U^ng) = g - U^ng, \] so that \( \| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \| \) =
\( \frac{1}{n} \| g - U^ng \| \). Because the transformation \( U \) is isometric
\( \| g - U^ng \| \leq \| g \| + \| U^ng \| = 2 \| g \| \), and letting \( n \rightarrow \infty \)
we find \( \| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \| \leq 2 \| g \| \rightarrow 0 \). The linear collection
of functions of the form \( f = g - Ug \) may not be closed since
a sequence of functions \( f_n = gn - Ugn \) may converge to a
function $f$ which is not of the above form. However, if we consider the closure $K$ of this collection of functions we would find that $\frac{1}{n} \sum_{k=0}^{n-1} U^k f \rightarrow 0$ for every $f \in K$. Let $f_p \rightarrow f$ where $f_p$ is of the form $g_p - Ug_p$. Then

$$
\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \right\| = \left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k \{ (f - f_p) + f_p \} \right\| \leq
$$

$$
\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k (f - f_p) \right\| + \left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f_p \right\| \leq \left\| f - f_p \right\| + \left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f_p \right\|
$$

since $\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k (f - f_p) \right\| \leq \left\| f - f_p \right\|$. Now since $f_p \rightarrow f$ we may choose $p$ so that $\left\| f_p - f \right\| < \varepsilon/2$ and also because $f_p$ is of the form $g_p - Ug_p$ we may choose $n$ large enough so that $\left\| \sum_{k=0}^{n-1} U^k f_p \right\| < \varepsilon/2$ which means that

$$
\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \right\| < \varepsilon/2 + \varepsilon/2 = \varepsilon
$$

for every $f \in K$. We have established that $\left\| \frac{1}{n} \sum_{k=0}^{n-1} U^k f \right\|$ converges to the null element for every $f \in K$.

We will now determine the orthogonal complement $K^\perp$ of $K$. A function $h$ is in $K^\perp$ if and only if $(h, g - Ug) = 0$ for all $f = g - Ug$. But $(h, g - Ug) = (h, Ug) = (h, g) - (U^* h, g) = (h - U^* h, g) = 0$ for all $g$, which implies that $h - U^* h = 0$ or $h = U^* h$. That is, functions $h \in K^\perp$ satisfy $U^* h = h$. Finally we will show that the condition $U^* h = h$ is equivalent with $Uh = h$. Expanding $\|Uh - h\|^2$ we have
\[ \|Uh - h\|^2 = (Uh - h, Uh - h) = \|Uh\|^2 - (h, Uh) - (Uh, h) \]

\[ + \|h\|^2 = 2\|h\|^2 - (U^*h, h) - (h, U^*h) \]

by the isometric property of \( U \). Assuming now that \( U^*h = h \) it follows that

\[ \|Uh - h\|^2 = 0 \]

so that \( Uh = h \). To prove the converse we consider \((Uh, Uh')\), \(h, h' \in H\). The equality \((U^*Uh, h') = (Uh, Uh') = (h, h') \) implies \((U^*Uh - h, h') = 0 \) for all \( h' \), so that \( U^*Uh = h \). Assuming that \( Uh = h \) we have immediately that \( U^*h = h \) and this completes the proof that the conditions \( U^*h = h \) and \( Uh = h \) are equivalent. Thus, the functions \( h \) in \( K^\perp \) satisfy the condition \( Uh = h \).

Using the theorem which states every function \( f \in H \) has the unique decomposition \( f = f_1 + f_2 \) where \( f_1 \in K^\perp \), \( f_2 \in K \) we see that

\[ \frac{1}{n} \sum_{k=0}^{n-1} U^k f = \frac{1}{n} \sum_{k=0}^{n-1} U^k (f_1 + f_2) = \]

\[ \frac{1}{n} \sum_{k=0}^{n-1} U^k f_1 + \frac{1}{n} \sum_{k=0}^{n-1} U^k f_2 \]

converges to \( \frac{1}{n} \sum_{k=0}^{n-1} U^k f_1 + 0 \). But \( f_1 \in K^\perp \) for which it is true that \( U^k f_1 = f_1 \) so that

\[ \frac{1}{n} \sum_{k=0}^{n-1} U^k f \]

converges to \( f_1 = Pf \) where \( P \) is the orthogonal projection on \( K \). This completes the proof of the mean ergodic theorem for isometric operators in Hilbert space.

It is possible to generalize the mean ergodic theorem by requiring that the transformations \( T \) satisfy a weaker condition than that of isometry and by extending the theorem to more general spaces. The proof which follows, originally due to G. Birkhoff(11), is an extension of the mean
ergodic theorem to any uniformly convex Banach space and to the more general linear transformations $T$ which satisfy $\|Tx\| \leq \|x\|$.

A uniformly convex space $X$ as defined by J. A. Clarkson(12) is one which satisfies the following conditions: If $x, y \in X$, $\|x\| \leq \|y\| \leq 1$ and $\|x - y\| \geq \varepsilon$ then $\|\frac{1}{2}(x + y)\| \leq \|y\| - u(\varepsilon)$ where $u(\varepsilon) \to 0$ if $\varepsilon \to 0$. Clarkson has shown that all $L_p(p > 1)$ spaces are uniformly convex.

Let $g_n$ denote the $n^{th}$ means $\frac{1}{n}(f + Tf + \ldots + T^{n-1}f)$ where $f \in X$ and let $M$ denote the infimum of the $\|g_n\|$ over all $n$. Then for some $n$ and $\varepsilon > 0$ it is true that $\|g_n\| < M + \varepsilon$. Consider now the translations of $g_n$ by $T^kn$ which are given by $\frac{1}{n}(T^{kn}f + T^{kn+1}f + \ldots + T^{kn+(n-1)}f)$, where $k$ is any positive integer or 0 and let $\alpha_k = \|T^{kn}g_n - g_n\|$. By the condition of uniform convexity it follows that $\|\frac{1}{2}(T^{kn}g_n + g_n)\| < M + \varepsilon - u(\alpha_k)$. Also since $\|Tf\| \leq \|f\|$ for $f \in X$ we have $\|\frac{1}{2}T^{hn}(T^{kn}g_n + g_n)\| < M + \varepsilon - u(\alpha_k)$ for all $h = 0, 1, 2, \ldots$. In particular we shall be interested in the inequality $\|\frac{1}{2}T^{hn}(T^{kn}g_n + g_n)\| < M + \varepsilon - u(\alpha_k)$ for $h = 0, 1, 2, \ldots, k - 1$. The mean of these $\frac{1}{2}T^{hn}(T^{kn}g_n + g_n)$ is given by $\frac{1}{2k} \sum_{h=0}^{k-1} \frac{1}{2}T^{hn}(T^{kn}g_n + g_n) = 1/2k((g_n + Tng_n + T^2g_n + \ldots + T^{(k-1)}g_n) + (T^{kn}g_n + T^{(k+1)}g_n + \ldots + T^{(2k-1)}g_n)) = g_{2kn}$ so that by
the above inequality \( \|g_{2kn}\| < M + \varepsilon - u(\alpha_k) \). But since 
M is the infimum of \( g_n \) for all \( n \), we also have \( \|g_{2kn}\| \geq M \)
and this is only possible if \( u(\alpha_k) \leq \varepsilon \). Then clearly
\( u(\alpha_k) \) depends upon \( \varepsilon \) for its value. For this reason we
write \( \|T^{kn}g_n - g_n\| \leq \omega(\varepsilon) \). Since now the mean \( g_{mn} \) of
the \( T^{kn}g_n \) for \( k = 0, \ldots, m - 1 \) is in the same way equal to
\( \frac{1}{m}(g_n + T^ng_n + \ldots + T^{m-1}g_n) \) it follows by subtracting \( g_n \)
from \( T^{kn}g_n \) and averaging these differences that
\[
\| \frac{1}{m} \sum_{k=0}^{m-1} (T^{kn}g_n - g_n) \| = \| g_{mn} - g_n \| \leq \omega(\varepsilon) \quad \text{for, by the}
\]
above each \( \|T^{kn}g_n - g_n\| \leq \omega(\varepsilon) \). We may write

\[
g_{r+i} = \frac{1}{r+1} \left( (f + T^{r-1}f) + (T^rf + \ldots + T^{r+i-1}f) \right)
\]
in the form \( g_{r+i} = \frac{r}{r+1} g_r + \frac{1}{r+1} (T^rf + \ldots + T^{r+i-1}f). \)

Subtracting \( g_r \) we find \( \|g_{r+i} - g_r\| = \)

\[
\| g_r \left( \frac{-i}{r+1} \right) + \frac{1}{r+1} (T^rf + \ldots + T^{r+i-1}f) \| \leq \)

\[
\frac{i}{r+1} \| g_r \| + \frac{i}{r+1} \| f \| \leq \frac{2i}{r+1} \| f \| \quad \text{since} \quad \| g_r \| = \)

\[
\frac{1}{r^n} \| f + T^{r-1}f \| \leq \frac{1}{r^n} (\| f \| + \| T^{r-1}f \| + \ldots + \| T^{r-1}f \|) \leq \)
\( \| f \|. \) Forming \( (g_{mn+i} - g_n) = (g_{mn+i} - g_{mn} + g_{mn} - g_n) \) we see that
\( \|g_{mn+i} - g_n\| \leq \|g_{mn+i} - g_{mn}\| + \| g_{mn} - g_n \|. \) But \( \|g_{mn+i} - g_{mn}\| \leq \)
\[
\frac{2i}{mn+1} \| f \| < \frac{1}{m+1} \| f \| < \frac{1}{m} \| f \|	ext{ if } 0 \leq i < n, \text{ and by the }
\]
above result \( \| g_{mn} - g_n \| < \omega(\varepsilon) \), so that \( \| g_{mn+i} - g_n \| < \frac{2}{m} \| f \| + \omega(\varepsilon) \). Letting \( \frac{2}{m} \| f \| = \xi \) we see that from a certain point on every \( g_j \) is within \( \xi + \omega(\varepsilon) \) of \( g_n \) and therefore within \( 2\xi + 2\omega(\varepsilon) \) of every other \( g_j \). We then clearly have a Cauchy sequence in the Banach space \( X \) which of course implies convergence.

It is important to note from the method of proof given above that replacing \( T^k \) by \( T^{k+p} \) does not alter the limit, i.e., \( \lim \frac{1}{n} \sum_{k=0}^{n-1} T^k f = \lim \frac{1}{n} \sum_{k=p}^{p+n-1} T^k f \). This is not necessarily true in the case of the individual ergodic theorem.

For, in the case of systems which are not strongly transitive you will recall that \( \lim_{n \to \infty} t_n(P) = \gamma(P) \) depended upon the particular point \( P \) initially on the surface \( \Sigma \). On the other hand, for strongly transitive systems the limit of \( t_n(P) \) does not depend on the particular point \( P \) and it follows that
\[
\lim_{n \to \infty} \sum_{k=0}^{n-1} T^k f = \lim_{n \to \infty} \sum_{k=p}^{p+n-1} T^k f \text{ for } f \in L_1.
\]
Examples of Stationary Ergodic Processes

It is perhaps appropriate to begin with a definition of a stationary random process. A random process may be defined as a sample space or ensemble composed of functions of time. In addition it is assumed that a probability distribution is defined over subsets of the sample space so that the occurrence of a specific event in the sample space has a definite probability even though in practice calculation of these probabilities may be almost impossible. As an example the ensemble of functions $A \cos(\omega t + \phi)$ form a random process. The sample space is three dimensional and each point of the space represents a particular function in the ensemble. Furthermore, the variables $A$, $\omega$ and $\phi$ satisfy a three dimensional probability distribution.

If it happens that the exterior conditions under which a random process functions do not change with time then evidently the probability distribution associated with subsets of the ensemble will not change. Then by definition, if this condition of invariance of probability is satisfied then the random process is stationary. It is of course implied that the ensemble of functions which constitute the process remains the same with respect to translations in time.

Given now an ensemble of functions $x(t)$ which constitute a stationary random process and any function $V(x(t))$ defined on the functions $x(t)$ of the ensemble we will
designate the average or expected value of \( V(x(t)) \) for all \( x(t) \) in the ensemble at a particular instant of time by \( \text{EV}(x(t)) \). Now it may be true that \( \text{EV}(x(t)) \) is equal to the average or expected value of \( V(x(t)) \) for any one particular function \( x(t) \) over an infinite period of time. If this property whereby ensemble averages at an instant are equal to averages of any member of the ensemble for time infinite holds, then the random stationary process is ergodic. A mathematical formulation would require that \( \text{EV}(x(t)) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} V(x(t + T)) dT \).

In conclusion, the following two examples of random processes which under suitable conditions exhibit the ergodic property are interesting. These examples are due to Lanning & Battin(l3) and G. D. Birkhoff(l4) respectively.

Example 1: The concept of the ergodic property was first introduced in classical statistical mechanics in the following way. Suppose that there are \( n \) molecules in a container. A total of \( 3n \) numbers are required to specify the simultaneous positions of all molecules, and another \( 3n \) numbers are required to specify all velocities or momenta. Thus the state of the system at any instant can be specified in terms of \( 6n \) numbers that can be considered as coordinates of a single point in a \( 6n \)-dimensional space called phase space. By this device the changing state of the system with time is visualized in terms of the motion through phase
space of the point representing the system.

Now consider an ensemble of systems of this nature with an assigned probability distribution, e.g., the collection of all systems possessing a total energy $E$ lying between fixed bounds $E_1 \leq E \leq E_2$. There will then exist a corresponding ensemble of points in phase space, and the motion of each such point with time will represent the progressive changes in the state of that system. Since any property or quantity associated with the system can be represented as a function of its $6n$ generalized coordinates, the successive values of such a quantity constitute a random function of time and the ensemble of such functions form a random process. Under suitable conditions the ergodic hypothesis can then be used to show that the average properties of a particular system with time are the same as the ensemble average at any instant.

Roughly speaking, the idea in this example is that of a continual mixing of the ensemble of points in phase space. This takes place in such manner that almost all points ultimately pass through every portion of the space an infinite number of times. In fact, if the initial distribution of points is properly chosen, then, with the exception of a set of points of zero probability, the average proportion of time spent in a given region by an arbitrary point is equal to the probability associated with that region in the initial distribution. Thus the behavior of any system in the
long run is the same as the behavior of any other, except for a collection of exceptional systems with special properties and with zero probability of occurrence.

Example 2: The kind of applications to dynamical systems which the ergodic theorem affords are exceedingly varied and interesting. Take the simple example of an idealized convex billiard table on which an idealized billiard ball $P$ moves with velocity 1. In the figure let $\phi = \text{arc } OA$, $\phi_1 = \text{arc } OA_1$, $L = AP$, $L^* = AA_1$. We have a transformation $(\phi_1, \theta_1) = T(\phi, \theta)$ defined over a rectangle $0 < \theta < \pi$; $0 \leq \phi \leq p$, ($p$ = perimeter of table) in the $\theta\phi$-plane, associated with the motion. It is not hard to prove that $T$ is measure preserving in the sense that the double integral $\int\int \frac{\sin \theta}{\sin \theta_1} \, d\theta \, d\phi$ has the same value when extended over any measurable part of this rectangle as over its image under $T$; indeed it would be possible to deform the rectangle so that, over the new region,
ordinary areas are preserved.

Furthermore it is clear that, if we associate with any "state of motion" of the billiard ball, as of P, the three coordinates $\theta, \phi, l$ then a steady flow $T_t$ is defined in the corresponding region of three-dimensional $\theta\phi l$-space:

$$0 < \theta < \pi \ ; \ 0 < \phi < \pi, \ 0 < l < l^*$$

in which the following volume integral is preserved:

$$\int \left( \int \frac{\sin \theta}{\sin \frac{\theta}{l}} d\theta d\phi \right) dL.$$  Thus the theorem applies to this flow.

Here are three obvious applications to this simple but typical dynamical problem:

(1) the average length of $n$ successive chords of the path tends to a definite limit, the same whether the time $t$ increases or decreases;

(2) the average angle $\theta$ at $n$ successive collisions tends to a definite limiting value;

(3) the billiard ball tends in the limit to lie in any assigned area of the table a definite proportion of the time.

M. H. Stone

Bibliography

G. Birkhoff

G. D. Birkhoff

J. A. Clarkson

E. Hopf
(2) Ergodentheorie, Berlin, 1937.

J. H. Lanning & R. H. Battin

J. von Neumann

H. Poincare

F. Riesz

F. Riesz & B. Sz-Nagy

M. H. Stone
A. C. Zaanen