Ergodic Theory and its Applications

Christy Tom Mathews, Mily Roy
Department of Mathematics, Deva Matha College Kuravilangad, Kerala, India
christy.tommathews@gmail.com
milyroy2913@gmail.com

Abstract

This paper provides an indepth knowledge about ergodicity, its definition, examples and certain characterization theorems. It discusses about the major ergodic theorems and the invariant measures for continuous transformations. It also deals with the number theoretic applications of ergodic theory.

Key Words
Ergodicity, transformations, mixing, weak mixing, strong mixing, invariant measures, normal numbers

INTRODUCTION

In its broad interpretation ergodic theory is the study of the qualitative properties of actions of groups on spaces. The space has some structure, that is, the space may be a measurable space, or a topological space, or a smooth manifold and each element of the group acts as a transformation on the space and preserves the given structure (for example, each element acts as a measure-preserving transformation, or a continuous transformation, or a smooth transformation).

The word “ergodic” was introduced by Boltzmann to describe a hypothesis about the action of \( \{T_t | t \in \mathbb{R} \} \) on an energy surface \( H^{-1}e \) when the Hamiltonian \( H \) is of the type that arises in statistical mechanics. Boltzmann had hoped that each orbit \( \{T_t(x) | t \in \mathbb{R} \} \) would equal the whole energy surface \( H^{-1}e \) and he called this statement the ergodic hypothesis. The word “ergodic” is an amalgamation of the Greek words ergon (work) and odos (path). Boltzmann made the hypothesis in order to deduce the equality of the time means and phase means which is a fundamental algorithm in statistical mechanics. The ergodic hypothesis, as stated above is false. The property the flow needs to satisfy in order to equate time means and phase means of real-valued functions is what now called ergodicity.

It is common to use the name ergodic theory to describe only the qualitative study of actions of groups on measure spaces. The actions on topological spaces and manifolds are often called topological dynamics and differentiable dynamics. This measure theoretic study began in the early 1930’s and the ergodic theorems of Birkhoff and von Neumann were proved then. During recent years ergodic theory had been used to give important results in many branches of mathematics.
2. ERGODICITY

2.1 DEFINITION AND CHARACTERISATION

Definition 2.1:
Let \((X, \mathcal{B}, m)\) be a probability space. A measure-preserving transformation \(T\) of \((X, \mathcal{B}, m)\) is called ergodic if the only members \(B\) of \(\mathcal{B}\) with \(T^{-1}B = B\) satisfy \(m(B) = 0\) or \(m(B) = 1\).

Theorem 2.1.1:
If \(T: X \to X\) is a measure-preserving transformation of the probability space \((X, \mathcal{B}, m)\) then the following statements are equivalent:

(i) \(T\) is ergodic.
(ii) The only members \(B\) of \(\mathcal{B}\) with \(m(T^{-1}B \Delta B) = 0\) are those with \(m(B) = 0\) or \(m(B) = 1\).
(iii) For every \(A \in \mathcal{B}\) with \(m(A) > 0\) we have \(m(\bigcup_{n=1}^{\infty} T^{-n}A) = 1\).
(iv) For every \(A, B \in \mathcal{B}\) with \(m(A) > 0, m(B) > 0\) there exists \(n > 0\) with \(m(T^{-n}A \cap B) > 0\).

Theorem 2.1.2:
If \((X, \mathcal{B}, m)\) is a probability space and \(T: X \to X\) is measure preserving then the following statements are equivalent:

(i) \(T\) is ergodic.
(ii) Whenever \(f\) is measurable and \(f \circ T_x = f_x \forall x \in X\) then \(f\) is constant a.e.
(iii) Whenever \(f\) is measurable and \(f \circ T_x = f_x \) a.e. then \(f\) is constant a.e.
(iv) Whenever \(f \in L^2(m)\) and \((f \circ T)_x = f_x \forall x \in X\) then \(f\) is constant a.e.
(v) Whenever \(f \in L^2(m)\) and \((f \circ T)_x = f_x \) a.e. then \(f\) is constant a.e.

Proof:
Trivially we have (iii)\(\Rightarrow\)(ii), (ii)\(\Rightarrow\)(iv), (v)\(\Rightarrow\)(iv), and (iii)\(\Rightarrow\)(v). So it remains to show (i)\(\Rightarrow\)(iii) and (iv)\(\Rightarrow\)(i).

Claim: (i)\(\Rightarrow\)(iii).
Let \(T\) be ergodic and suppose \(f\) is measurable and \(f \circ T = f\) a.e. We can assume that \(f\) is real-valued for if \(f\) is complex-valued we can consider the real and imaginary parts separately. Define for \(k \in \mathbb{Z}\) and \(n > 0\),

\[
X(k, n) = \left\{ x \mid \frac{k}{2^n} \leq f(x) < \frac{(k + 1)}{2^n} \right\}
\]
\[ f^{-1}\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right]\right). \]

We have

\[ T^{-1}X(k,n)\Delta X(k,n) \subset \{x|(f \circ T)(x) \neq f(x)\} \]

And hence

\[ m(T^{-1}X(k,n)\Delta X(k,n)) = 0 \]

So that by (ii) of theorem 2.1.1, \( m(X(k,n)) = 0 \) or 1.

For each fixed \( n \) \( \bigcup_{k \in \mathbb{Z}} X(k,n) = X \) is a disjoint union and so there exists a unique \( k_n \) with \( m(X(k_n,n)) = 1 \). Let \( Y = \bigcap_{n=1}^\infty X(k_n,n) \). Then \( m(Y) = 1 \) and \( f \) is constant on \( Y \) so that \( f \) is constant a.e.

(iv)\( \Rightarrow \) (i): Suppose \( T^{-1}E = E, E \in \mathcal{B} \). Then \( X_E \in L^2(m) \) and

\[
(X_E \circ T)(x) = X_E(x) \forall x \in X \text{ so, by (iv), } X_E \text{ is constant a.e. Hence } X_E = 0 \text{ a.e. or } X_E = 1 \text{ a.e. and } m(E) = \int X_E \, dm = 0 \text{ or } 1.
\]

**Remarks 2.1.1:**

1. A similar characterisation in terms of \( L^p(m) \) functions (for any \( p \geq 1 \)), is true, since in the last part of the proof \( X_E \) is in \( L^p(m) \) as well as \( L^2(m) \).

2. Another characterisation of ergodicity of \( T \) is: whenever \( f:X \to \mathbb{R} \) is measurable and \( f(Tx) \geq f(x) \) a.e. then \( f \) is constant a.e. This is a stronger statement than (iii).

The next result is useful to analyse which of the examples are ergodic and it also relates the idea of measurable theoretic dense orbits (theorem 2.1.1 (iii) & (iv)) to the usual notion of dense orbits for continuous maps.

We have the following theorem concerning rotations of the unit circle \( K \).

**Theorem 2.1.3:**

The rotation \( T(z) = az \) of the unit circle \( K \) is ergodic (relative to Haar measure \( m \)) iff \( a \) is not a root of unity.
Proof:

Suppose \( a \) is root of unity, then \( a^p = 1 \) for \( p \neq 0 \). Let \( f(z) = z^p \). Then \( f \circ T = f \) and \( f \) is not constant a.e. Therefore \( T \) is not ergodic by theorem 2.1.2 (iii). Conversely, suppose \( a \) is not a root of unity and \( f \circ T = f \), \( f \in L^2(m) \). Let \( f(z) = \sum_{n=-\infty}^{\infty} b_n z^n \) be its Fourier series. Then

\[
f(a^z) = \sum_{n=-\infty}^{\infty} b_n a^n z^n
\]

and hence \( b_n (a^n - 1) = 0 \) for each \( n \). If \( n \neq 0 \) then \( b_n = 0 \), and so \( f \) is constant a.e. Theorem 2.1.2(v) gives that \( T \) is ergodic.

Theorem 2.1.4:

Let \( G \) be a compact group and \( T(x) = ax \), a rotation of \( G \). Then \( T \) is ergodic iff \( \{ a^n \}_{n=-\infty}^{\infty} \) of \( G \). In particular, if \( T \) is ergodic, then \( G \) is abelian.

Proof:

Suppose \( T \) is ergodic. Let \( H \) denote the the closure of the subgroup \( \{ a^n \}_{n=-\infty}^{\infty} \) of \( G \). If \( H \neq G \) then by the result “if \( H \) is a closed subgroup of \( G \) and \( H \neq G \) there exists a \( \gamma \in \widehat{G}, \gamma \neq 1 \) such that \( \gamma(h) = 1 \forall h \in H \)” there exists \( \gamma \in \widehat{G}, \gamma \neq 1 \) such that \( \gamma(h) = 1 \forall h \in H \). Then

\[
\gamma(Tx) = \gamma(ax) = \gamma(a) \gamma(x) = \gamma(x),
\]

and this contradicts ergodicity of \( T \). Therefore \( H = G \). Conversely, suppose \( \{ a^n \}_{n \in \mathbb{Z}} \) is dense in \( G \). This implies \( G \) is abelian. Let \( f \in L^2(m) \) and \( f \circ T = f \). By the result “if \( G \) is compact, the members of \( \widehat{G} \) form an orthonormal basis for \( L^2(m) \) where \( m \) is a Haar measure” \( f \) can be represented as a Fourier series \( \sum_i b_i \gamma_i \) where \( \gamma_i \in \widehat{G} \). Then

\[
\sum_i b_i \gamma_i(a) \gamma_i(x) = \sum_i b_i \gamma_i(x)
\]

so that if \( b_i \neq 0 \) then \( \gamma_i(a) = 1 \) and, since \( \gamma_i(a^n) = (\gamma_i(a))^n = 1, \gamma_i \equiv 1 \). Therefore only the constant term of the Fourier series of \( f \) can be non-zero, i.e., \( f \) is constant a.e. Theorem 2.1.2(v) gives that \( T \) is ergodic.
**Theorem 2.1.5:**

The two-sided $(p_0, p_1, \ldots, p_{k-1})$ shift is ergodic.

**Proof:**

Let $\mathcal{A}$ denote the algebra of all finite unions of measurable rectangles. Suppose $T^{-1}E = E, E \in \mathcal{B}$. Let $\varepsilon > 0$ be given, and choose $A \in \mathcal{A}$ with $m(E \Delta A) < \varepsilon$. Then

$$|m(E) - m(A)| = |m(E \cap A) + m(E/A) - m(A \cap E) - m(A/E)| < m(E/A) + m(A/E) < \varepsilon$$

Choose $n_0$ so large that $B = T^{-n_0}A$ depends upon different coordinates from $A$. Then $m(B \cap A) = m(B)m(A) = m(A)^2$ because $m$ is a product measure. We have

$$m(E \Delta B) = m(T^{-n}E \Delta T^{-n}A) = m(E \Delta A) < \varepsilon$$

and since $E \Delta (A \cap B) \subset (E \Delta A) \cup (E \Delta B)$ we have $m(E \Delta (A \cap B)) < 2\varepsilon$.

Hence $|m(E) - m(A \cap B)| < 2\varepsilon$.

And $|m(E) - m(E)^2| \leq |m(E) - m(A \cap B)| + |m(A \cap B) - m(E)^2| < 2\varepsilon + |m(A)^2 - m(E)^2| \leq 2\varepsilon + m(A)|m(A) - m(E)| + m(E)|m(A) - m(E)| < 4\varepsilon$, since $\varepsilon$ is arbitrary $m(E) = m(E)^2$ which implies $m(E) = 0$ or $1$.

**Theorem 2.1.6:**

If $T$ is the $(p, P)$ Markov shift (either one-sided or two sided) then $T$ is ergodic iff the matrix $P$ is irreducible (i.e. $\forall i, j \exists n > 0$ with $p_{ij}^{(n)} > 0$ where $p_{ij}^{(n)}$ is the $(i,j)$-th entry of the matrix $P^n$).

**Note:**

Ergodic decomposition of a given transformation can be formulated as follows:

Ergodic transformations are the "irreducible" measure-preserving transformations and we would like every measure-preserving transformation to be built of ergodic ones. To get some idea how this ergodic decomposition of a given transformation may be formulated consider a map $T$ of a cylinder $X = [0,1] \times K$ given by $T(x, z) = (x, az)$ where $a \in K$ is not a root of unity. In other words
T is the direct product of the identity map, I, of [0,1] and the rotation \( S_z = a \cdot z \) of \( K \). So \( T \) maps each circular vertical section of the cylinder to itself and acts on each section by \( S \).

Now \( [0,1] \times K \) can be partitioned into the sets \( \{ x \} \times K \) and we can consider the direct product measure \( m_1 \times m_2 \), where \( m_1 \) is the Lebesgue measure on \([0,1]\) and \( m_2 \) is the Haar measure on \( K \), as being decomposed into copies of \( m_2 \) on each element \( \{ x \} \times K \) of the partition.

In other words we have a partition \( \zeta \) of \( X \) on each element of which \( m_1 \times m_2 \) induces a probability measure and \( T \) induces a transformation which is ergodic relative to the induced measure. This is the ergodic decomposition of \( T \). It turns out that this procedure can always be performed when \((X,\mathcal{B},m)\) is a nice measure space, namely a Lebesgue space.

Now if \( X \) is a complete separable space, \( m \) is a probability measure on the \( \sigma \)-algebra \( \mathcal{B}(X) \) of Borel subsets of \( X \) and \( \mathcal{B} \) is the completion of \( \mathcal{B}(X) \) by \( m \) then \((X,\mathcal{B},m)\) is a Lebesgue space. Suppose \( T \) is a measure-preserving transformation of a Lebesgue space \((X,\mathcal{B},m)\). Let \( J(T) \) be the \( \sigma \)-algebra consisting of all measurable sets \( B \) with \( T^{-1}B = B \). The theory of Lebesgue spaces determines a partition \( \zeta \) of \((X,\mathcal{B},m)\) such that if \( \mathcal{B}(\zeta) \) denotes the smallest \( \sigma \)-algebra consisting of all members of \( \zeta \) then \( \mathcal{B}(\zeta) = J(T) \) in the sense that each element of one \( \sigma \)-algebra differs only by a null set from an element of the other. Moreover \( TC \subset C \) for each element \( C \) of \( \zeta \). Also, the measure \( m \) can be decomposed into probability measures \( m_C \) on the elements \( C \) of \( \zeta \). The transformation \( T_C \) turns out to be ergodic relative to the measure \( m_C \). This transformation is essentially unique.

**Example of Ergodic Transformation:**

**Irrational rotations**

Consider \(([0,1],\mathcal{B},\lambda)\), where \( \mathcal{B} \) is the Lebesgue \( \sigma \)-algebra, and \( \lambda \) Lebesgue measure. For \( \theta \in ]0,1[ \), consider the transformation \( T_\theta : [0,1] \to [0,1] \) defined by \( T_\theta x = x + \theta \mod 1 \).

\( T_\theta \) is measure-preserving with respect to \( \lambda \). If \( \theta \) is rational then \( T_\theta \) is not ergodic. Consider an example, let \( \theta = \frac{1}{4} \), then the set
Claim: \( T_\theta \) is ergodic iff \( \theta \) is irrational.

Proof of Claim:

Suppose \( \theta \) is irrational, and let \( f \in L^2(X, \mathcal{B}, \lambda) \) be \( T_\theta \)–invariant. \( f \) can be represented in Fourier series as

\[
f \in \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}.
\]

Since \( f(T_\theta x) = f(x) \), then

\[
f(T_\theta x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n (x+\theta)}
\]

\[
= \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x} e^{2\pi i n \theta}
\]

\[
= f(x)
\]

\[
= \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}
\]

Hence, \( \sum_{n \in \mathbb{Z}} a_n (1 - e^{2\pi i n \theta}) e^{2\pi i n x} = 0 \). By the uniqueness of the Fourier coefficients, we have

\( a_n (1 - e^{2\pi i n \theta}) = 0 \) \( \forall n \in \mathbb{Z} \). If \( n \neq 0 \), since \( \theta \) is irrational we have \( 1 - e^{2\pi i n \theta} \neq 0 \). Thus, \( a_n = 0 \) \( \forall n \neq 0 \), and therefore \( f(x) = a_0 \) is a constant. Then \( T_\theta \) is ergodic.

2.2 THE ERGODIC THEOREM

The first major result in ergodic theory was proved in 1931 by G. D. Birkhoff. We shall state it for measure-preserving transformation of a \( \sigma \)-finite measure space. A \( \sigma \)-finite measure on a measurable space \( (X, \mathcal{B}) \) is a map \( m: \mathcal{B} \to \mathbb{R}^+ \cup \{\infty\} \) such that \( m(\emptyset) = 0 \), \( m(\bigcup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} m(B_n) \) whenever \( \{B_n\} \) is a sequence of members of \( \mathcal{B} \) which are pairwise disjoint subsets of

\[
A = \left[ 0, \frac{1}{8} \right] \cup \left[ \frac{1}{4}, \frac{3}{8} \right] \cup \left[ \frac{1}{2}, \frac{5}{8} \right] \cup \left[ \frac{3}{4}, \frac{7}{8} \right]
\]
is \( T_\theta \) – invariant but \( \mu(A) = \frac{1}{2} \).
X, and there is a countable collection of \( \{A_n\}_{n=1}^{\infty} \) of elements of \( \mathcal{B} \) with \( m(A_n) < \infty \) \( \forall n \) and \( \bigcup_{n=1}^{\infty} A_n = X \). The Lebesgue measure on \((\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))\) provides an example of a \( \sigma \)-finite measure.

Of course any probability measure is \( \sigma \)-finite.

**Theorem 2.2.1 (Birchoff Ergodic Theorem):**

Suppose \( T: (X, \mathcal{B}, m) \to (X, \mathcal{B}, m) \) is measure-preserving (where we allow \((X, \mathcal{B}, m)\) to be \( \sigma \)-finite) and \( f \in L^1(m) \). Then \( \left(\frac{1}{n}\right) \sum_{i=0}^{n-1} f(T^i(x)) \) converges a.e. to a function \( f^* \in L^1(m) \). Also \( f^* \circ T = f^* \) a.e. and if \( m(X) < \infty \), then

\[ \int f^* \, dm = \int f \, dm. \]

**Remark 2.2.1:**

If \( T \) is ergodic then \( f^* \) is constant a.e. and so if \( m(X) < \infty \) \( f^* = \left(\frac{1}{m(X)}\right) \int f \, dm \) a.e. If \((X, \mathcal{B}, m)\) is a probability space and \( T \) is ergodic we have \( \forall f \in L^1(m) \)

\[ \lim_{n \to \infty} \left(\frac{1}{n}\right) \sum_{i=0}^{n-1} f(T^i(x)) = \int f \, dm \text{ a.e.} \]

**Corollary 2.2.1 (L^p Ergodic Theorem of Von Neumann):**

Let \( 1 \leq p < \infty \) and let \( T \) be a measure-preserving transformation of the probability space \((X, \mathcal{B}, m)\). If \( f \in L^p(m) \) there exists \( f^* \in L^p(m) \) with \( f^* \circ T = f^* \) a.e.

\[ \left\| \left(\frac{1}{n}\right) \sum_{i=0}^{n-1} f(T^i(x)) - f^*(x) \right\|_p \to 0. \]

**Corollary 2.2.2:**

Let \((X, \mathcal{B}, m)\) be a probability space and let \( T: X \to X \) be a measure-preserving transformation.

Then \( T \) is ergodic iff \( \forall A, B \in \mathcal{B} \)
\( \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} m(T^{-1}A \cap B) \to m(A)m(B). \)

**Proof:**

Suppose \( T \) is ergodic. Putting \( f = \chi_A \) in theorem 2.2.1 gives

\[ \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} \chi_A T^i(x) \to m(A) \text{a.e.} \]

Multiplying by \( \chi_B \) gives

\[ \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} \chi_A T^i(x) \chi_B \to m(A) \chi_B. \]

and the dominated convergence theorem implies

\[ \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} m(T^{-1}A \cap B) \to m(A)m(B). \]

Conversely, suppose the convergence property holds. Let \( T^{-1}E = E, E \in \mathcal{B}. \) Put \( A = B = E \) in the convergence property to get

\[ \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} m(E) \to m(E)^2. \]

Hence \( m(E) = m(E)^2 \) and \( m(E) = 0 \) or 1.

**Theorem 2.2.2 (Maximal Ergodic Theorem):**

Let \( U: L^1_B(m) \to L^1_B(m) \) be a positive linear operator with \( \|U\| \leq 1. \) Let \( N > 0 \) be an integer and let \( f \in L^1_B(m). \) Define \( f_0 = 0, f_n = f + Uf + U^2f + \cdots + U^{n-1}f \) for \( n \geq 1, \) and \( F_N = \max_{0 \leq n \leq N} f_n \geq 0. \) Then \( \int_{\{x|F_N(x) > 0\}} f dm \geq 0. \)

**3. INVARIANT MEASURES FOR CONTINUOUS TRANSFORMATIONS**

**3.1 MIXING**

**Definition:**

Let \( T \) be a measure-preserving transformation of the probability space \((X, \mathcal{B}, m)\).
(i) $T$ is weak-mixing if $\forall A, B \in \mathcal{B}$

$$\lim_{n \to \infty} \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} \left| m(T^{-i}A \cap B) - m(A)m(B) \right| = 0.$$ 

(ii) $T$ is strong-mixing if $\forall A, B \in \mathcal{B}$

$$\lim_{n \to \infty} m(T^{-n}A \cap B) = m(A)m(B).$$

Remarks:

1. Every strong-mixing transformation is weak-mixing and every weak-mixing transformation is ergodic. This is because if $\{a_n\}$ is a sequence of real numbers then $\lim_{n \to \infty} a_n = 0$ implies

$$\lim_{n \to \infty} \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} |a_i| = 0$$

And this second condition implies

$$\lim_{n \to \infty} \left( \frac{1}{n} \right) \sum_{i=0}^{n-1} a_i = 0.$$ 

2. Intuitive descriptions of ergodicity and strong-mixing can be given as follows. To say $T$ is strong-mixing means that for any set $A$ the sequence of sets $T^{-n}A$ becomes, asymptotically, independent of any other set $B$. Ergodicity means $T^{-n}A$ becomes independent of $B$ on the average, for each pair of sets $\mathcal{A}(J)$

Examples:

1. An example of an ergodic transformation which is not weak-mixing is given by a rotation $T(z) = az$ on the unit circle $K$. If $A$ and $B$ are two small intervals on $K$ then $T^{-i}A$ will be disjoint from $B$ for at least half of the values of $i$ so that

$$\left( \frac{1}{n} \right) \sum_{i=0}^{n-1} \left| m(T^{-i}A \cap B) - m(A)m(B) \right| \geq \frac{1}{2} m(A)m(B)$$

for large $n$. From this one sees that, intuitively, a weak-mixing transformation has to do some “stretching”.
2. There are examples of weak-mixing $T$ which are not strong-mixing. Kakutani has an example constructed by combinatorial methods and Maruyama constructed an example using Gaussian processes. Chacon and Katok and Stepin also have examples. If $(X, \mathcal{B}, m)$ is a probability space, let $\tau(X)$ denote the collection of all invertible measure-preserving transformations of $(X, \mathcal{B}, m)$. If we topologize $\tau(X)$ with the “weak” topology, the class of weak-mixing transformations is of second category while class of strong-mixing transformations is of first category. So from the viewpoint of this topology most transformations are weak-mixing but not strong-mixing.

The following theorem gives a way of checking the mixing properties for examples by reducing the computations to a class of sets we can manipulate with. For example, it implies we need only consider measurable rectangles when dealing with the mixing properties of shifts.

**Theorem 3.1.1:**

Let $(X, \mathcal{B}, m)$ be a measure space and let $\mathcal{I}$ be a semi-algebra that generates $\mathcal{B}$. Let $T: X \rightarrow X$ be a measure-preserving transformation. Then

(i) $T$ is ergodic iff $\forall A, B \in \mathcal{I}$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n}\sum_{i=0}^{n-1} m(T^{-i}A \cap B) = m(A)m(B), \right.$$  

(ii) $T$ is weak-mixing iff $\forall A, B \in \mathcal{I}$

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n}\sum_{i=0}^{n-1} |m(T^{-i}A \cap B) - m(A)m(B)| = 0, \right.$$  

(iii) $T$ is strong-mixing if $\forall A, B \in \mathcal{B}$

$$\lim_{n \rightarrow \infty} m(T^{-n}A \cap B) = m(A)m(B).$$

**Proof:**

Since each member of the algebra $\mathcal{A}(\mathcal{I})$, generated by $\mathcal{I}$ can be written as a finite disjoint union of members of $\mathcal{I}$ it follows that if any of the three convergence properties hold for all members of $\mathcal{I}$ then they hold for all members of $\mathcal{A}(\mathcal{I})$.

Let $\varepsilon > 0$ be given and let $A, B \in \mathcal{B}$. Choose $A_o, B_o \in \mathcal{A}(\mathcal{I})$ with

$$m(A \Delta A_o) < \varepsilon \text{ and } m(B \Delta B_o) < \varepsilon.$$  

For any $i \geq 0$,

$$\left(T^{-i}A \cap B\right) \Delta \left(T^{-i}A_o \cap B_o\right) \subset \left(T^{-i}A \cap T^{-i}A_o\right) \cup \left(B \Delta B_o\right),$$
so we have \( m\left((T^{-n}A \cap B)\Delta(T^{-i}A_0 \cap B_0)\right) < 2\varepsilon \), and therefore

\[
\left|m(T^{-i}A \cap B) - m(T^{-i}A_0 \cap B_0)\right| < 2\varepsilon.
\]

Therefore

\[
\left|m(T^{-i}A \cap B) - m(A)m(B)\right| \leq \left|m(T^{-i}A \cap B) - m(T^{-i}A_0 \cap B_0)\right| + \left|m(T^{-i}A_0 \cap B_0) - m(A_0)m(B_0)\right| + \left|m(A_0)m(B_0) - m(A)m(B)\right| < 4\varepsilon + \left|m(T^{-i}A_0 \cap B_0) - m(A_0)m(B_0)\right|.
\]

This inequality together with the known behaviour of the right hand term proves (ii) and (iii). To prove (i) one can easily obtain

\[
\left|(\frac{1}{n})\sum_{i=0}^{n-1} m(T^{-i}A \cap B) - m(A)m(B)\right| < 4\varepsilon + \left|(\frac{1}{n})\sum_{i=0}^{n-1} m(T^{-i}A_0 \cap B_0) - m(A_0)m(B_0)\right|
\]

and then use the known behaviour of the right hand side.

As an application of this result we shall prove the result about ergodicity of Markov shifts. To do this we shall use the following:

**Lemma 3.1.1:**

Let \( P \) be a stochastic matrix, having a strictly positive probability vector \( p \) with \( pP = p \). Then \( Q = \lim_{N \to \infty} \left(\frac{1}{N}\right)\sum_{n=0}^{N-1} P^n \) exists. The matrix \( Q \) is also stochastic and \( QP = PQ = Q \). Any eigenvector of \( P \) for the eigen value 1 is also an eigenvector of \( Q \). Also \( Q^2 = Q \).
Theorem 3.1.2:

Let $T$ denote the $(p, P)$ Markov shift (either one-sided or two-sided). We can assume $p_i > 0$ for each $i$ where $p = (p_0, p_1, ..., p_{k-1})$. Let $Q$ be the matrix obtained in Lemma. The following are equivalent:

(i) $T$ is ergodic.
(ii) All rows of the matrix $Q$ are identical.
(iii) Every entry in $Q$ is strictly positive.
(iv) $P$ is irreducible.
(v) $1$ is a simple eigenvalue of $P$.

3.2 INVARIANT MEASURES FOR CONTINUOUS TRANSFORMATIONS

The space $X$ denotes a compact metrisable space and $d$ will denote a metric on $X$. The $\sigma$-algebra of Borel set by $\mathcal{B}(X)$. So $\mathcal{B}(X)$ is the smallest $\sigma$-algebra containing all open subsets of $X$ and smallest $\sigma$-algebra containing all closed subsets of $X$. $M(X)$ denote the collection of all probability measures defined on the measurable space $(X, \mathcal{B}(X))$ and the members of $M(X)$ are called Borel probability measures on $X$. Each $x \in X$ determines a member $\delta_x$ of $M(X)$ defined by

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

$M(X)$ is a convex set where $pm + (1 - p)\mu$ is defined by

$$(pm + (1 - p)\mu)(B) = pm(B) + (1 - p)\mu(B) \text{ if } p \in [0, 1].$$

Theorem 3.2.1

A Borel probability measure $m$ on a metric space $X$ is regular.

Corollary 3.2.1:

If $T: X \rightarrow X$ is a continuous transformation of a compact metric space $X$ then $M(X, T)$ is non-empty.

The following are some properties of $M(X, T)$.
Theorem 3.2.2:

If $T$ is a continuous transformation of the compact metric space $X$ then

(i) $M(X,T)$ is a compact subset of $M(X)$.

(ii) $M(X,T)$ is convex.

(iii) $\mu$ is an extreme point of $M(X,T)$ iff $T$ is an ergodic measure-preserving transformation of $(X, \mathcal{B}(X), \mu)$.

(iv) If $\mu, m \in M(X,T)$ are both ergodic and $\mu \neq m$ then they are mutually singular.

Remark 3.2.2:

Since $M(X,T)$ is a compact convex set each member of $M(X,T)$ can be expressed in terms of ergodic members of $M(X,T)$. If $E(X,T)$ denote the set of extreme points of $M(X,T)$ then for each $\mu \in M(X,T)$ there is a unique measure $\tau$ on the borel subsets of the compact metrisable space $M(X,T)$ such that $\tau(E(X,T)) = 1$ and $\forall f \in C(X)$

$$\int_X f(x) d\mu(x) = \int_{E(X,T)} \left( \int_X f(x) dm(x) \right) d\tau(m).$$

We write $\mu = \int_{E(X,T)} md\tau(m)$ and call this the ergodic decomposition of $\mu$. Hence every $\mu \in M(X,T)$ is a generalized convex combination of ergodic measures.

3.3 INTERPRETATION OF ERGODICITY AND MIXING

Let $T: X \to X$ be a continuous transformation of a compact metric space. We say $\mu \in M(X,T)$ is ergodic or weak-mixing or strong-mixing if the measure-preserving transformation has corresponding property. $\mu$ is ergodic iff $\forall f, g \in L^2(\mu)$

$$\frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i(x)) g(x) d\mu(x) \to \int f d\mu \int g d\mu.$$

Lemma 3.3.1:

Let $\mu \in M(X,T)$. Then

(i) $\mu$ is ergodic iff $\forall f \in C(X) \forall g \in L^1(\mu)$

$$\frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i(x)) g(x) d\mu(x) \to \int f d\mu \int g d\mu.$$

(ii) $\mu$ is strong-mixing iff $\forall f \in C(X) \forall g \in L^1(\mu)$
\[ \int f(T^1(x))g(x) \, d\mu(x) \to \int f \, d\mu \int g \, d\mu. \]

(iii) \( \mu \) is weak-mixing iff there is a set \( J \) of natural numbers of density zero such that

\[ \forall f \in C(X) \forall g \in L^1(\mu), \]

\[ \lim_{J \ni n \to \infty} \int f(T^1(x))g(x) \, d\mu(x) \to \int f \, d\mu \int g \, d\mu. \]

**Proof:**

(i) Suppose the convergence condition holds and let \( F, G \in L^2(\mu) \). Then \( G \in L^2(\mu) \) so, \( \forall f \in C(X) \)

\[ \frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i(x))g(x) \, d\mu(x) \to \int f \, d\mu \int g \, d\mu. \]

Now approximate \( F \) in \( L^2(\mu) \) by continuous functions to get

\[ \frac{1}{n} \sum_{i=0}^{n-1} \int F(T^i(x))g(x) \, d\mu(x) \to \int F \, d\mu \int G \, d\mu. \]

Now suppose \( \mu \) is ergodic. Let \( f \in C(X) \). Then \( f \in L^2(\mu) \) so if \( h \in L^1(\mu) \) we have

\[ \frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i(x))h(x) \, d\mu(x) \to \int f \, d\mu \int h \, d\mu. \]

If \( g \in L^1(\mu) \) then by approximating \( g \) in \( L^1(\mu) \) by \( h \in L^2(\mu) \) we obtain

\[ \frac{1}{n} \sum_{i=0}^{n-1} \int f(T^i(x))g(x) \, d\mu(x) \to \int f \, d\mu \int g \, d\mu. \]

Similarly we can prove part (ii) and (iii).

**Theorem 3.3.1:**

Let \( T \) be a continuous transformation of a compact metric space. Let \( \mu \in M(X,T) \).

(i) \( \mu \) is ergodic iff whenever \( m\in M(X) \) and \( m \ll \mu \) then

\[ \frac{1}{n} \sum_{i=0}^{n-1} T^i m \to \mu. \]

(ii) \( \mu \) is strong-mixing iff whenever \( m\in M(X) \) and \( m \ll \mu \) then \( T^n m \to \mu \).

(iii) \( \mu \) is weak-mixing iff there is a set \( J \) of natural numbers of density zero such that whenever \( m\in M(X) \) and \( \mu \) is ergodic then

\[ \lim_{J \ni n \to \infty} T^n m \to \mu. \]

**Proof:**

(i) We use the ergodicity condition of the above lemma.

Let \( \mu \) be ergodic and suppose \( m \ll \mu, m\in M(X) \). Let \( g = \frac{dm}{d\mu} \in L^1(\mu) \). If \( f \in C(X) \) then
\[
\int f \ d \left( \frac{1}{n} \sum_{i=0}^{n-1} T^i m \right) = \frac{1}{n} \sum_{i=0}^{n-1} \int f \circ T^i \ dm \\
= \frac{1}{n} \sum_{i=0}^{n-1} \int f \left( T^i (x) \right) g(x) \ d\mu(x) \\
\rightarrow \int f \ d\mu \int g \ d\mu \\
= \int f \ d\mu \int 1 \ dm \\
= \int f \ d\mu.
\]

Therefore, \( \frac{1}{n} \sum_{i=0}^{n-1} T^i m \rightarrow \mu \)

We now show the converse.

Suppose the convergence condition holds. Let \( g \in L^1(\mu) \) and \( g \geq 0 \). Define \( m \in M(X) \) by

\[
m(B) = c \int g \ d\mu
\]

where \( c = \frac{1}{\int g \ d\mu} \).

Then if \( f \in C(X) \) we have, by reversing the above reasoning,

\[
\frac{1}{n} \sum_{i=0}^{n-1} \int f \left( T^i (x) \right) g(x) \ d\mu(x) \rightarrow \int f \ d\mu \int g \ d\mu.
\]

If \( g \in L^1(\mu) \) is real-valued then \( g = g_+ - g_- \) where \( g_+, g_- \geq 0 \) and we apply the above to \( g_+ \) and \( g_- \) to get the desired condition of the lemma for \( g \). The case of the complex-valued \( g \) follows. Similarly, we can prove part (ii) and (iii).

### 3.4 UNIQUE ERGODICITY

**Definition 3.4.1:**

A continuous transformation \( T : X \rightarrow X \) is a compact metrisable space \( X \) is uniquely ergodic if there is only one \( T \) invariant Borel probability measure on \( X \), i.e., \( M(X,T) \) consists of one point.

If \( T \) is uniquely ergodic and \( M(X,T) = \{ \mu \} \) then \( \mu \) is ergodic because it is an extreme point of \( M(X,T) \).
Unique ergodicity is connected to minimality by:

**Theorem 3.4.1:**

Suppose $T: X \to X$ is a homeomorphism of a compact metrisable space $X$. Suppose $T$ is uniquely ergodic and $M(X, T) = \{\mu\}$. Then $T$ is minimal iff $\mu(U) > 0$ for all non-empty open sets $U$.

**Proof:**

Suppose $T$ is minimal. If $U$ is open, $U \neq \emptyset$, then $X = \bigcup_{n=0}^{\infty} T^n(U)$, so if $\mu(U) = 0$ then $M(X) = 0$, a contradiction.

Conversely, suppose $\mu(U) > 0$ for all open non-empty $U$. Suppose also that $T$ is not minimal. There exists a closed set $K$ such that $TK = K$, $\emptyset \neq K \neq X$. The homeomorphism $T/K$ has an invariant Borel probability measure $\mu_K$ on $K$. Define $\bar{\mu}$ on $X$ by $\bar{\mu}(B) = \mu_K(K \cap B)$ for all Borel sets $B$. Then $\bar{\mu} \in M(X, T)$ and $\bar{\mu} \neq \mu$ because $\mu(X/K) > 0$, as $X/K$ is non-empty and open, while $\bar{\mu}(X/K) = 0$. This contradicts the unique ergodicity of $T$.

**Example 3.4.1:**

The map $T: K \to K$ given by

$$T(e^{2\pi i \theta}) = e^{2\pi i \theta^2}, \theta \in [0, 1],$$

is an example of a uniquely ergodic homeomorphism which is not minimal. The point $1 \in K$ is a fixed point for $T$ and $\Omega(T) = \{1\}$ so that $M(K, T) = \{\delta_1\}$.

**Remark 3.4.1:**

Results about unique ergodicity known before are given in Oxtoby. More recent results of Jewett and Kringer imply that any ergodic invertible measure-preserving transformation of a Lebesgue space is isomorphic to a minimal uniquely ergodic homeomorphism of a zero dimensional compact metrisable space. In particular there are minimal uniquely ergodic homeomorphisms with any prescribed non-negative real number for their entropy. Hahn and Katznelson had found minimal uniquely ergodic transformations with arbitrarily large measure-theoretic entropy.
Theorem 3.4.2:

Let $T: K \to K$ be a homeomorphism with no periodic points. Then $T$ is uniquely ergodic. Moreover

(a) there is a continuous surjection $\phi: K \to K$ and a minimal rotation $S: K \to K$ with $\phi T = S \phi$. The map $\phi$ has the property that for each $z \in K$, $\phi^{-1}(z)$ is either a point or a closed sub-interval of $K$.

(b) If $T$ is minimal the map $\phi$ is a homeomorphism so that every minimal homeomorphism of $K$ is topologically conjugate to a rotation.

Example 3.4.2:

H. Furstenberg constructed an example of a minimal homeomorphism $T: K^2 \to K^2$ of the two dimensional torus which is not uniquely ergodic. The example preserves Haar measure and has the form $T(z, w) = (az, \phi(z)w)$ where $\{a_n\}_{n=\infty}^0$ is dense in $K$ and $\phi: K \to K$ is a well chosen continuous map.

4. APPLICATIONS

NORMAL AND SIMPLY NORMAL NUMBERS

An important application of Birkhoff Ergodic Theorem is the study of normal numbers. A number is said to be simply normal to the base $b$ if, for $k \in \{0,1,2,3,\ldots, b-1\}$, the average frequency of occurrence of $k$ in the base $b$ expansion of $x$ is $1/b$. In other words, a number is normal to the base $b$ if, for every sequence of $m$ digits, $m \in N$, the average frequency of occurrence of that sequence in the $b$-expansion of $x$ is $1/b^m$.

Every number which is normal to base $b$ is simply normal to the base $b$ by its definition. This result and definition is due to Borel. Borel further proved in 1909 that except for a subset of measure zero, every $x \in [0,1]$ is normal. We will prove this result in this section using ergodic theory.

We formalize our definition for base $b$ series expansion of $x \in [0,1]$ by defining a transformation
$T(x) = bx \pmod{1} = \begin{cases} 
  bx & x \in \left[0, \frac{1}{b}\right] 
  bx - 1 & x \in \left[\frac{1}{b}, \frac{2}{b}\right] 
  \vdots 
  bx - (b-1) & x \in \left[\frac{b-1}{b}, 1\right] 
\end{cases}$

Our expansion of base $b$ results from iterating this map $T$. We define

$a_1(x) = [bx]$ and

$a_k(x) = [bT^{k-1}x] = a_1(T^{k-1}x)$.

From our definition, we have

$x = \frac{a_1}{b} + \frac{T(x)}{b}$

And we find that

$x = \frac{a_1}{b} + \frac{T(x)}{b}$

$= \frac{a_1}{b^2} + \frac{T^2(x)}{b^2}$

$= \frac{a_1}{b^2} + \frac{a_2}{b^3} + \ldots + \frac{a_k}{b^k} + \ldots$

$= \sum_{j=1}^{\infty} \frac{a_j(x)}{b^j}$

and this series converges to $x$.

**Definition 4.1**

A number $x \in [0, 1]$ is simply normal to base $b$ if for every $k \in \{1, 2, \ldots, b-1\}$,

$\lim_{n \to \infty} \frac{N(k,n)}{n} = \frac{1}{b}$. 
Definition 4.2

A number \( x \in [0,1] \) is normal to base \( b \) if for every \( m \)-length sequence of digits \( k_1k_2 \ldots k_m \), where \( k_1, k_2, \ldots, k_m \in \{1, 2, \ldots, b - 1\} \),

\[
\lim_{n \to \infty} \frac{1}{n} \# \{ r | 1 \leq r \leq n \text{ and } a_r(x) = k_1, \ldots, a_{r+m-1}(x) = k_m \} = \frac{1}{b^m}.
\]

We illustrate as follows.

Theorems 4.1 (Borel’s Theorem on Normal Numbers):

Almost all numbers in \([0,1]\) are normal to the base 2, i.e., for a.e. \( x \in [0,1] \) the frequency of 1’s in the binary expansion of \( x \) is \( \frac{1}{2} \).

Proof:

Let \( T: [0,1] \to [0,1] \) be defined by \( T(x) = 2x \mod 1 \). We know that \( T \) preserves Lebesgue measure \( m \) and is ergodic. Let \( Y \) denote the set of points of \([0,1]\) that have a unique binary expansion. Then \( Y \) has a countable complement so \( m(Y) = 1 \).

Suppose \( x = a_1/2 + a_2/2^2 + \ldots \) has a unique binary expansion. Then

\[
T(x) = T\left(\frac{a_1}{2} + \frac{a_2}{2^2} + \ldots\right) = \frac{a_2}{2} + \frac{a_3}{2^2} + \ldots
\]

Let \( f(x) = \chi_{\left[\frac{1}{2},1\right]}(x) \). Then

\[
f(T^i x) = f\left(\frac{a_{i+1}}{2} + \frac{a_{i+2}}{2^2} + \ldots\right) = \begin{cases} 1 & \text{iff } a_{i+1} = 1 \\ 0 & \text{iff } a_{i+1} = 0. \end{cases}
\]

Hence if \( x \in Y \) the number of 1’s in the first \( n \) digits of the dyadic expansion of \( x \) is \( \sum_{i=0}^{n-1} f(T^i x) \). Dividing both sides of the equality by \( n \) and applying the ergodic theorem we get

\[
\left(\frac{1}{n}\right) \sum_{i=0}^{n-1} f(T^i x) \to \int \chi_{\left[\frac{1}{2},1\right]} \, dm = \left(\frac{1}{2}\right) \text{ a.e.}
\]

This says that the frequency of 1’s in the binary expansion of almost all primes is \( \frac{1}{2} \).

CONCLUSION

This paper aimed to bring out the importance of ergodic theory in mathematics. Ergodic theory is one of the few branches of mathematics which has changed radically during the last two decades. Before this period, with a small number of exceptions, ergodic theory dealt primarily with averaging problems and general qualitative questions, while now it is a powerful amalgam of
methods used for the analysis of statistical properties of dynamical systems. For this reason, the
problems of ergodic theory now interest not only the mathematician, but also the research
worker in physics, biology, chemistry e.t.c.

The main principle which adhered to from the beginning was to develop the
approaches and methods of ergodic theory in the study of numerous concrete examples. Because
of this, the paper contains the description of the fundamental notions of ergodicity and mixing.

In recent years there have been some fascinating interactions of ergodic theory with
differentiable dynamics, Von Neumann algebras, probability theory, statistical mechanics, and
other topics. This paper tries to tackle many difficulties regarding ergodic theory and its
applications.

REFERENCES


