Dynamical Systems and a Brief Introduction to Ergodic Theory

Leo Baran

Spring 2014

Abstract

This paper explores dynamical systems of different types and orders, culminating in an examination of the properties of the logistic map. It also introduces Ergodic theory and important results in the field.

Contents

1 Dynamical Systems

1.1 Differential Equations

1.1.1 Flows on $\mathbb{R}^1$

1.1.2 Flows on $\mathbb{R}^2$

1.1.3 $\mathbb{R}^3$ and chaos

1.2 Maps

1.2.1 The Logistic Map

2 Ergodic Theory

2.1 Measure Theory Preliminaries

2.2 A Few Results

1 Dynamical Systems

The field of dynamics came to be in the 1600s by Newton’s developments in differential equations and their applications to the laws of gravitation and planetary motion. He realized that finding an
exact solution for some dynamical systems (such as the three-body problem, or describing the exact motion of three planetary bodies confined to the laws of gravitation) was essentially impossible. A coupled centuries later, Poincare developed a novel approach to analyzing such systems. His method was to answer qualitative questions about the system rather than develop an exact quantitative solution. Until the mid-1900s, dynamics was largely concerned with non-linear oscillating systems and their applications to physics. The development of high-speed computing led Lorenz in 1963 to discover chaotic motion in dynamics. Since then interest in dynamics and chaos has proliferated and applications to real-world systems have become exceedingly numerous.

There are two main types of dynamical systems: differential equations and iterated maps. Differential equations describe the motion of systems in continuous time, while iterated maps deal exclusively with discrete time.

1.1 Differential Equations

Consider an $n$-dimensional space whose points $(x_1(t), x_2(t), \ldots, x_n(t))$ are functions of time. In general, an autonomous dynamical system on this $n$-dimensional phase space is defined as the system

$$\frac{dx_1}{dt} = f_1(x_1, x_2, \ldots, x_n)$$

$$\frac{dx_2}{dt} = f_2(x_1, x_2, \ldots, x_n)$$

$$\vdots$$

$$\frac{dx_n}{dt} = f_n(x_1, x_2, \ldots, x_n)$$

Where autonomous refers to the fact that $f_1, \ldots, f_n$ don’t depend on $t$. In this case, the differential equations that compose the system are of first order, i.e. they only use first time derivatives of $(x_1, x_2, \ldots, x_n)$. 
1.1.1 Flows on $\mathbb{R}^1$

On the other hand, in dynamical systems a first-order system is of the form

$$\frac{d}{dt} x(t) = f(x(t))$$

Example 1.1. $x'(t) = \sin(x(t))$

In this case we can find an exact solution for the system by doing some slightly sketchy separating of variables:

$$\frac{dx}{dt} = \sin(x) \iff dt = \frac{dx}{\sin(x)} \iff \int dt = \int \frac{dx}{\sin(x)} \iff t = -\ln | \csc(x) + \cot(x) | + C.$$ 

If $x(0) = x_0$, then $0 = -\ln | \csc(x_0) + \cot(x_0) | + C \iff C = \ln | \csc(x_0) + \cot(x_0) |$. Thus

$$t = -\ln | \csc(x) + \cot(x) | + \ln | \csc(x_0) + \cot(x_0) | = \ln \left( \frac{| \csc(x_0) + \cot(x_0) |}{| \csc(x) + \cot(x) |} \right).$$

Most of the time we want to know what $x(t)$ does as $t \to \infty$ for an arbitrary initial condition $x_0$, however in this case (and in most systems) determining a formula for $x(t)$ isn’t trivial. Instead we can interpret $f(x(t))$ as a vector field: $t$ is time, $x(t)$ is the position on $\mathbb{R}$, and $\frac{dx}{dt}$ is the velocity on the line. We can graph this vector field while keeping in mind that a positive velocity means we move in the positive direction on $\mathbb{R}$ and vice versa for negative.

For every point at which $x'(t) = 0$, $x(t)$ is stationary. These points are called fixed points. In the above graph, certain fixed points are marked with solid circles and others with hollow circles. The fixed points with solid circles at $(2n+1)\pi$, $n \in \mathbb{Z}$ are points toward which $x(t)$ is attracted, while the hollow circles at $2n\pi$, $n \in \mathbb{Z}$ are points from which $x(t)$ is repelled. These points are conveniently called attracting and repelling fixed points, respectively. The fixed to which $x(t)$ tends depends on
the initial condition. If \( x(0) = \frac{\pi}{4} \), then the system will tend to \( \pi \) as \( t \to \infty \). If \( x(0) = -\frac{\pi}{4} \), then the system will tend to \(-\pi\) and if \( x(0) = 0 \), the system will stay at \( x(t) = 0 \forall t \).

We, being mathematicians, would like to know about solutions to all dynamical systems of this form. Using the fact that our dynamical system is essentially a differential equation, we can show that solutions to nonautonomous dynamical systems exist and are unique in a certain interval of time.

**Theorem 1.1. (Linear Fundamental Existence and Uniqueness Theorem)**

Let \( f \) and \( g \) be continuous functions on \((a, b) \in \mathbb{R} \) and \( t_0 \in (a, b) \) and \( x_0 \in \mathbb{R} \). Then there is a unique function \( p(t) = x \) that satisfies

\[
x' = f(t)x + g(t), \quad p(t_0) = x_0
\]

on \((a, b)\).

Namely,

\[
p(t) = x_0e^{-F(t)} + e^{-F(t)} \int_{t_0}^{t} g(\tau)e^{F(\tau)}d\tau
\]

Where \( F(t) = \int_{t_0}^{t} f(\tau)d\tau \).

**Proof.** First we prove existence. Suppose the solution \( p(t) \) takes the form above. Take a first derivative: \( p'(t) = -x_0e^{-F(t)}d \frac{d}{dt}[F(t)] + \frac{d}{dt}[e^{-F(t)}] \int_{t_0}^{t} g(\tau)e^{F(\tau)}d\tau + \frac{d}{dt}[F(t)]e^{-F(t)} = -x_0e^{-F(t)}f(t) - f(t)e^{-F(t)} \int_{t_0}^{t} g(\tau)e^{F(\tau)}d\tau + g(t)e^{-F(t)} = -f(t)[x_0e^{-F(t)} + \int_{t_0}^{t} g(\tau)e^{F(\tau)}d\tau] + g(t) = -f(t)p(t) + g(t) \iff p'(t) = -f(t)p(t) + g(t) \iff p'(t) + f(t)p(t) = g(t) \). Also \( p(t_0) = x_0e^{-F(t_0)} + e^{-F(t_0)} \int_{t_0}^{t_0} g(\tau)e^{F(\tau)}d\tau = x_0e^{\int_{t_0}^{t_0} f(\tau)d\tau} = x_0 \). Thus \( p(t) \) is a solution to the initial value problem.

Next we show that the solution is unique. Suppose \( q(t) \) is also a solution to the problem. Suppose \( r(t) = q(t)e^{F(t)} \). Then \( r(t_0) = q(t_0) = x_0 \). Also \( r'(t) = q'(t)e^{F(t)} - q(t)f(t)e^{F(t)} = e^{F(t)}[q'(t) - q(t)f(t)] \). We know by the fundamental theorem of calculus that \( \int_{t_0}^{t} r'(\tau)d\tau = r(t) - r(t_0) = r(t) + \int_{t_0}^{t} e^{F(\tau)}g(\tau)d\tau = x_0 + \int_{t_0}^{t} e^{F(\tau)}g(\tau)d\tau \).

We also know that \( r(t) = q(t)e^{F(t)} \iff q(t) = r(t)e^{-F(t)} \). Thus \( q(t) = e^{-F(t)}[x_0 + \int_{t_0}^{t} e^{F(\tau)}g(\tau)d\tau] = x_0e^{-F(t)} + e^{-F(t)} \int_{t_0}^{t} e^{F(\tau)}g(\tau)d\tau \).

We therefore know that every solution to the problem must take this form, so the solution is unique.
Dynamical systems in two dimensions become a bit more interesting.

1.1.2 Flows on $\mathbb{R}^2$

We consider the general system

$$\frac{d}{dt}x_1(t) = f_1(x_1(t), x_2(t))$$

$$\frac{d}{dt}x_2(t) = f_2(x_1(t), x_2(t))$$

Or

$$\frac{d}{dt}x = f(x)$$

If our system is linear, then

$$\frac{d}{dt}x_1(t) = ax_1(t) + bx_2(t), \quad \frac{d}{dt}x_2(t) = cx_1(t) + dx_2(t) \iff \dot{x} = Ax$$

where $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $x = \begin{pmatrix} x_1(t) \\ x_2(t) \end{pmatrix}$

We can learn a great deal about the behavior of this system by studying the matrix $A$.

We would like to find lines on the plane which are invariant under $A$. That is, $Av = \lambda v$. (Like usual in solving differential equations, we assume that the solution takes the form $x(t) = e^{\lambda t}v$.) This means we want to find the eigenvalues $\lambda_j$ and and eigenvectors $v \neq 0$ of $A$. We find them in the following way: $Av = \lambda v \iff (A - \lambda I)v = 0 \implies \det(A - \lambda I) = 0$. The last step is derived from the fact that $v$ is assumed not to be zero, thus $(A - \lambda I)$ is a non-invertible matrix, which means its determinant is zero. Next we solve $\det(A - \lambda I) = 0 \iff \det\begin{pmatrix} a-\lambda & b \\ c & d-\lambda \end{pmatrix} = 0 \iff \lambda^2 - \tau\lambda + \Delta = 0$ where $\tau = a + d = \text{Trace}(A)$ and $\Delta = ad - bc = \det(A)$. Using the quadratic equation, we see that

$$\lambda_1 = \frac{\tau + \sqrt{\tau^2 - 4\Delta}}{2}, \lambda_2 = \frac{\tau - \sqrt{\tau^2 - 4\Delta}}{2}$$

are the two eigenvalues of $A$ corresponding to eigenvectors $v_1, v_2$. The general solution to the system is a linear combination of the two solutions: $x(t) = c_1e^{\lambda_1 t}v_1 + c_2e^{\lambda_2 t}v_2$. We can use the following result in linear algebra to see where this solution exists.
Lemma 1.2. Suppose $A$ is a $2$ by $2$ nonzero matrix with eigenvalues $\lambda_1$ and $\lambda_2$ such that $\lambda_1 \neq \lambda_2$. Then the eigenvectors $v_1$ and $v_2$ that correspond to $\lambda_1$ and $\lambda_2$ are linearly independent.

Proof. We need to show that given a linear combination such that $c_1v_1 + c_2v_2 = 0$, $c_1 = c_2 = 0$.

Suppose we have such a linear combination. $0 = A[c_1v_1 + c_2v_2] = c_1\lambda_1v_1 + c_2\lambda_2v_2$

Also $0 = c_1v_1 + c_2v_2 \iff c_1\lambda_1v_1 + c_2\lambda_1v_2$

Take the difference to get $0 = c_1\lambda_1v_1 - c_1\lambda_1v_1 + c_2\lambda_2v_2 - c_2\lambda_1v_2 = c_2v_2(\lambda_2 - \lambda_1)$.

Since $v_2$ is an eigenvector, $v_2 \neq 0$. Also, since $\lambda_1$ and $\lambda_2$ are eigenvectors and not equal, $(\lambda_1 - \lambda_2) \neq 0$. Thus $c_2 = 0$. By a similar argument we know $c_1 = 0$. Thus the only linear combination of $v_1$ and $v_2$ that equals zero is one in which each linear combination coefficient is zero. This means $v_1$ and $v_2$ are linearly independent.

Thus, as long as $\lambda_1 \neq \lambda_2$, we know that $v_1$ and $v_2$ span $\mathbb{R}^2$ and we can write any initial condition in the plane as a linear combination of them:

$$x(0) = x_0 = c_1e^{0}v_1 + c_2e^{0}v_2 = c_1v_1 + c_2v_2$$

We know, then, that $x(t) = c_1e^{\lambda_1t}v_1 + c_2e^{\lambda_2t}v_2$ is a solution to

$$\dot{x} = Ax$$

and that it satisfies any initial condition $x(0) = x_0$. By the following theorem, this means that this is the only solution for the system.

Theorem 1.3. General Existence and Uniqueness Theorem

Consider a general nonlinear system: $x_1(t) = f_1(x_1(t), x_2(t), \ldots, x_n(t)), \ldots, x_n(t) = f_n(x_1(t), x_2(t), \ldots, x_n(t))$

$\dot{x} = f(x)$. Assume $f$ is continuous and all of its first partial derivatives are continuous in some open, connected set $D$ in $\mathbb{R}^n$. Then for $x(0) = x_0 \in D$, there exists a unique solution $x(t)$ on some time interval $(-t_0, t_0)$.

Proof. See [5], Picard-Lindelof Theorem. 

\[\square\]
The following example shows us that an interesting phenomenon can occur in two-dimensional systems that is unavailable to those of a dimension lower.

Example 1.2. Consider the system \( \dot{x} = -y \) \( \dot{y} = x \) \( \iff \dot{x} = (0 \; -1) x \). Since the determinant of \( A \) is 1 and the trace is 0, we know that the eigenvalues are \( \pm \sqrt{-1} = \pm i \). The solution to the system is \( x(t) = c_1 e^{it} v_1 + c_2 e^{-it} v_2 = c_1 (\cos(t) + i \sin(t)) v_1 + c_2 (\cos(t) - i \sin(t)) v_2 \). Thus systems with complex eigenvalues display oscillatory behavior. In one dimension, however, oscillations are impossible—the system can either go to \( \pm \infty \) or tend to a fixed point, but it can never come back around to a certain value without being multi-valued.

There are a variety of theorems that tell us whether or not a dynamical system has a closed orbit. The main two are the following:

Theorem 1.4. Bendixon-Dulac Theorem

Let the following system be defined on a simply connected domain \( D \): \( \dot{x} = f(x) \) with \( f \) continuously differentiable. If there exists a continuously differentiable, real-valued function \( g(x) \) such that \( \nabla (x' g(x)) \) has one sign throughout \( D \), then there are no closed orbits for the system lying entirely in \( D \).

Proof. Suppose there is a closed orbit \( C \) lying entirely in \( D \). Call the region inside \( C_A \).

Since \( g(x) \) is continuously differentiable, we can apply the divergence theorem, which says that

\[
\iint_D \nabla (x' g(x)) dA = \int_C (x' g(x)) \cdot n ds
\]

. Since we assumed \( \nabla (xg(x)) \) has one sign throughout \( D \), we know that \( \iint_D \nabla (xg(x)) dA \) is nonzero. However, since \( x' \) is tangent to the trajectory of the system (which in this case is \( C \)) and \( n \) is perpendicular to \( C \), \( x' \cdot n = 0 \implies x' g(x) \cdot n = 0 \), so \( \int_C (x' g(x)) \cdot n ds = 0 \), so we arrive at a contradiction and therefore no closed orbit \( C \) can exist in \( D \).

\[ \square \]

Theorem 1.5. Poincare-Bendixon Theorem

Suppose the following: \( R \) is a compact subset of \( \mathbb{R}^2 \), \( \dot{x} = f(x) \) where \( f \) is continuously differentiable on an open set containing \( R \), \( R \) does not contain any fixed points of the aforementioned system,
and that there exists a trajectory $C$ that starts and stays in $R$ for all time. Then $R$ contains a closed orbit.


1.1.3 $\mathbb{R}^3$ and chaos

Again, something interesting happens when we move up a dimension. The Poincare-Bendixon Theorem showed us that two-dimensional systems confined to compact sets not containing fixed points are guaranteed to approach a closed orbit as $t \to \infty$. In three dimensions, this isn’t the case.

**Example 1.3.** The Lorenz equations are defined the following:

$$\begin{align*}
\dot{x} &= \sigma (y-x), \\
\dot{y} &= rx - y - xz, \\
\dot{z} &= xy - bx
\end{align*}$$

Where $\sigma, r, b > 0$ are constants.

Lorenz discovered that this system demonstrates erratic behavior in a bounded region of space over a wide range of $\sigma, r, b$. Unlike two-dimensional systems, in three dimensions systems in a bounded region not containing a fixed point, trajectories have the possibility of never approaching a closed orbit, but wandering around the space for all time while never intersecting themselves, as is guaranteed by the existence and uniqueness theorem (a unique solution cannot intersect itself since that would mean two trajectories would come from the same point).

It turns out that the Lorenz system settles onto a set, called an attractor, which we define to be a closed set $A$ such that:

1. any trajectory of the system which starts in $A$ stays in $A$ for all time.

2. there exists an open set $U$ that contains $A$ such that if $x(0) \in U$, then as $t \to \infty$ the distance between the trajectory and $A$ approaches zero.

3. There is no proper subset of $A$ which satisfies the previous two conditions.
In the case of the Lorenz system, trajectories settle onto a *strange attractor* (see above), which is defined as an attractor which exhibits sensitive dependence on initial conditions, which is to say that two trajectories starting arbitrarily close together on the attractor rapidly diverge from each other, making long term prediction of such trajectories impossible. It turns out that strange attractors are not regular two- or three-dimensional sets, but fractals.

We define chaotic systems to be those with the following characteristics:

1. There exists an open set of initial conditions that leads to long-term aperiodic behavior, which is to say trajectories don’t settle to fixed points or closed orbits as \( t \to \infty \).

2. The system is deterministic, which is to say the system has no random parameters.

3. The system exhibits sensitive dependence on initial conditions.

### 1.2 Maps

One-dimensional maps are of the form \( x_{n+1} = f(x_n) \). Similarly to dynamical systems, we define a fixed point of a map to be a point \( x_0 \) such that \( f(x_0) = x_0 \). Maps can also have cycles—for example, a period-two cycle is one in which \( f(x_0) = x_1, f(x_1) = x_0 \).

#### 1.2.1 The Logistic Map

**Example 1.4.** The logistic map is defined as \( x_{n+1} = rx_n(1 - x_n) = f(x_n) \). We’ll consider only positive values of \( x \). If we restrict the parameter \( r \) such that \( 0 \leq r \leq 4 \), then the map will map \( \{ x : x \in [0,1] \} \) to itself.

We want to find the fixed points of the logistic map, which is to say points \( x^* \) such that \( x^* = rx^*(1 - x^*) \Leftrightarrow x^*[(r - 1) - rx^*] = 0 \). The solutions to this are \( x^*_1 = 0 \) and \( rx^*_2 = r - 1 \Leftrightarrow x^*_2 = \frac{r - 1}{r} \).
For $0 \leq r \leq 1$, $x_2^*$ is negative, so $x_1^*$ is the only fixed point we care about. We want to know the behavior of trajectories with respect to the fixed point. To do this we consider a perturbation $x^*+\delta$ and linearly approximate the map at that point. $f(x^*+\delta_n) = x^* + \delta_{n+1} = f(x^*) + f'(x^*)\delta_n + O(\delta_n^2) \iff x^* + \delta_{n+1} = x^* + f'(x^*)\delta_n + O(\delta_n^2) \iff \delta_{n+1} = f'(x^*)\delta_n$. If we can ignore the $O(\delta_n^2)$ terms, then we can determine whether the fixed point is attracting or repelling directly from the map $\delta_{n+1} = f'(x^*)\delta_n$. If $|f'(x^*)| < 1$, then $\delta_n \to 0$ as $n \to \infty$. If $|f'(x^*)| > 1$, then $\delta_n \to \infty$ as $n \to \infty$. If $|f'(x^*)| = 1$, then the $O(\delta_n^2)$ terms dictate the stability of the fixed point.

In the case of the logistic map, $|f'(0)| = |r-2(0)| = |r|$. For $r \in (0, 1)$, the origin is an attracting fixed point. For $r \in [1, 3]$, $x_2^* = \frac{r-1}{r}$ is non-negative and attracting: $f'(x_2^*) = |r-2\frac{-1}{r}| = r^2 - 2r - 2$. The origin, however, is repelling, since $|r| > 1$. Thus as $x \to \infty$, $x_n$ tends to a positive value.

Things get bizarre for $r$ greater than 3. At $r \approx 3.449$, a period-four cycle begins. At $r \approx 3.544$, a period-eight cycle begins. We call these $r$ values points of period doubling—$a_0 = 3$, $a_1 = 3.449$, etc. They eventually converge to a point $a_\infty \approx 3.5699$. Interestingly, the period doubling values converge in a geometric sense, meaning the ratio between successive period doubling points shrinks by a constant:

$$\lim_{n \to \infty} \frac{a_n - a_{n-1}}{a_{n+1} - a_n} \approx 4.669$$

Keep this number in mind.

The above (bifurcation) diagram highlights the fixed points of the logistic map. The splitting of the graph at $r = 3$ corresponds to the creation of the period-two cycle. The two branches are
the two values between which the map oscillates. As we can see, past the critical value $a_\infty$, utter pandemonium ensues. Interesting features of this bifurcation diagram include an interval of period-three behavior amid the chaos and self-similar period-doubling (each branch splitting is similar to the overall structure). The reason for the latter is obvious once we consider what successive iterates of the function look like.

Above we have a graph of $f(x, R_0)$ for $R_0 < 3$. Next to it is a graph of $f^2(x, R_1) = f(f(x, R_1))$ for $3 < R_1 < 3.449 = a_1$. The points where these graphs cross the line $y = x$ correspond to fixed points of the map. Stable fixed points of $f^2$ are points of a period-two cycle. One notices that the region inside the box in the graph on the right is an inverted copy of the graph on the left. Each time we iterate $f$ twice more, another two-cycle is created at a local maximum/minimum, as demonstrated below.

Next we consider the same system, but for the map $x_{n+1} = g(x_n) = r \sin(n)$ for $0 \leq r \leq 1$ and $0 \leq x \leq 1$. The graph of this function in the given parameters looks very similar to that of the logistic map for small $r$. In fact both are said to be unimodal, or have a single quadratic maximum in the region we’re considering. If we look at the bifurcation diagrams of these two maps, we find
that they are qualitatively exactly the same. It turns out that if we calculate $\delta = \lim_{n \to \infty} \frac{a_n - a_{n-1}}{a_{n+1} - a_n}$ (where $a_n$ is the point of the $n$th period doubling) for the sine map, we arrive at the same value for the logistic map of $\delta \approx 4.669$. In fact, this constant (dubbed the Feigenbaum constant) is the same for any unimodal map$^3$.

## 2 Ergodic Theory

Ergodic theory is the study of measurable dynamics—it uses measure theory to study the behavior of dynamical systems on abstract spaces. The original motivation for the field came from statistical mechanics, specifically the problem of modeling the evolution of $n$ particles and representing each at a certain time as a point in $\mathbb{R}^6$, with three dimensions for the position of each particle and three more for momentum vectors in each direction.

### 2.1 Measure Theory Preliminaries$^{1,2}$

**Definition 2.1.** An algebra $\Sigma$ on a nonempty set $X$ is a collection of subsets of $X$ which is closed under complements and a finite number of unions, which is to say if $A \in \Sigma$ then $X \setminus A \in \Sigma$ and if $A_1, A_2, A_3...A_n \in \Sigma$ then $A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n \in \Sigma$, respectively.

**Definition 2.2.** A $\sigma$-algebra $\Sigma$ is a collection of subsets of $X$ which is closed under complements and countably many unions.

**Definition 2.3.** A measure space is a triplet $(X, \Sigma, \mu)$ in which $X$ is a nonempty set, $\Sigma$ is a $\sigma$-algebra on $X$ and $\mu : \Sigma \to [0, \infty]$ is a function such that

1. $\mu(A) \geq 0 \ \forall A \in \Sigma$

2. $\mu(\emptyset) = 0$ where $\emptyset$ is the empty set

3. $\mu$ is $\sigma$-additive, which is to say for all countable collections $\{A_n\}$ of disjoint sets in $\Sigma$,

$$\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$$

**Definition 2.4.** If $(X, \Sigma, \mu)$ is a measure space and $\mu(X) = 1$, then $\mu$ is called a probability measure and $(X, \Sigma, \mu)$ is called a probability space.
Definition 2.5. \((X, \Sigma)\) and \((X', \Sigma')\) are measure spaces. A function \(f : X \to X'\) is measurable if for every \(A' \in \Sigma'\), \(f^{-1}(A') \in \Sigma\).

Definition 2.6. A measure preserving transformation is a quartet \((X, \Sigma, \mu, T)\) such that \((X, \Sigma, \mu)\) is a measure space and \(T : X \to X\) is a measurable function such that \(\forall A \in \Sigma, \mu(T^{-1}(A)) = \mu(A)\). T is said to preserve the measure \(\mu\).

Definition 2.7. A probability preserving transformation is a measure preserving transformation on a probability space.

2.2 A Few Results

Theorem 2.1. Monotonicity of Measure

Let \((X, \Sigma, \mu)\) be a measure space. Then \(\mu\) is monotone, which is to say if \(A, B \in \Sigma\) and \(A \subseteq B\), then \(\mu(A) \leq \mu(B)\).

Proof. Assume \(A, B \in \Sigma\) and \(A \subseteq B\). By definition \(\mu\) is an additive function for disjoint sets, so if \(A \cup B = 0\), then \(\mu(A \cup B) = \mu(A) + \mu(B)\).

We know \(B = B \cap (B \cup A) = B \cap ((B \setminus A) \cup A) = ((B \setminus A) \cup B) \cap ((B \setminus A) \cup A) = (B \setminus A) \cup (A \cap B) = (B \setminus A) \cup A\), the last equality holding since \(A \subseteq B\). Since \((B \setminus A) \cap A = 0\), \(B\) is the union of two disjoint sets. Thus \(\mu(B) = \mu(B \setminus A) + \mu(A) \geq \mu(A)\) since by definition \(\mu(C) \geq 0 \forall C \in \Sigma\). \(\square\)

Definition 2.8. A measure-preserving transformation \(T : X \to X\) on a measure space \((X, \Sigma, \mu)\) is recurrent if for every measurable set \(A \in \Sigma\) with positive measure there is a null set \(N \subset A\) (null in this context means the outer measure of \(N\) is zero) such that \(\forall x \in A \setminus N\) there is an integer \(n > 0\) such that \(T^n(x) = [T \circ T \circ \ldots \circ T](x) \in A\). Informally, \(T\) is recurrent means for any positive measure set \(A\), almost every point in \(A\) returns to \(A\) at some point in the future.

Theorem 2.2. Poincare Recurrence Theorem

Let \((X, \Sigma, \mu)\) be a measure space and \(T : X \to X\) a measure-preserving transformation. Then \(T\) is a recurrent transformation.

Proof. We want to show that for a set \(A \in \Sigma\), \(\mu(\{x \in A : \exists N such that f^n(x) \notin A \forall n > N\}) = 0\).

Define \(A_n = \bigcup_{j=n}^{\infty} f^{-j}(A)\). Thus if \(j \leq i\), \(A_i \subset A_j\). By definition, \(\mu(f^{-1}(A)) = \mu(A)\) since \(f\) is a measure preserving transformation. We can write \(A_i = f^{j-i}(A_j)\), so \(\mu(A_i) = \mu(A_j)\) since applying
$f^{-1}$ numerous times doesn’t change the measure. We know that $A \subset A_0 = A \cup f^{-1}(A) \cup \ldots$, thus $A \setminus A_n \subset A \setminus A_n$. Since we know $\mu$ has monotone measure, $\mu(A \setminus A_n) \leq \mu(A_0 \setminus A_n)$.

Notice that $A_0 = (A_0 \setminus A_n) \cup A_0 \cap A_n = (A_0 \setminus A_n) \cup A_n$ (see proof for monotonicity of measure), therefore $\mu(A_0) = \mu(A_0 \setminus A_n) + \mu(A_n)$ since $(A_0 \setminus A_n)$ and $A_n$ are disjoint. $\mu(A_0) = \mu(A_0 \setminus A_n) + \mu(A_n) \iff \mu(A_0 \setminus A_n) = \mu(A_0) - \mu(A_n) = 0$. Hence $\mu(A \setminus A_n) = 0$ for any $n > 0$. Since $(A \setminus A_n)$ is the set of $x \in A$ such that for all $j > n$, $f^j(x) \notin A$, we are done.

\[ \square \]

**Theorem 2.3. Birkhoff Ergodic Theorem**

Let $(X, \Sigma, \mu)$ be a probability space and $T$ a measure-preserving transformation on the space. If $f : X \to \mathbb{R}$ is an integrable function and $T$ is ergodic, which is to say for every $A \in \Sigma$ with positive measure, $\mu(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$. Then

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

almost everywhere.

*Proof.* See [3], page 177.

\[ \square \]

The previous theorem is the highlight of Ergodic theory. What the theorem says is that given an ergodic transformation and invariant measure, then the time average is equal to the space average. If we pick any point in the space and calculate the average of $f$ along the point’s orbit and take the limit as time goes to $\infty$, it will be equal to the average value of the function at all points in the space. Thus any orbit of the system will take up the entire space.

**References**


