Ergodic Theory

1 Dynamical Systems

A **dynamical system**, usually written as the tuple (T, X), is described by a transformation that maps a phase space onto itself, $T : X \to X$. The set of points attained from repeated applications of the transformation from some starting point is known as its **forward orbit** or trajectory,

$$\mathcal{O}(x, T) = \{T^i(x) : i \in \mathbb{Z}^+\}.$$

If the trajectory repeats itself, the point is considered **periodic**, $T^n(x) = T^m(x), n \neq m$.

2 Measure

A measure is an extension of the concepts of length, area, or volume in Euclidean geometry. In a generic space, a measure is a way of assigning a value or weight to different parts of the space.

Let X be a set. A collection $\mathcal{F}$ of subsets of X is called a **σ-algebra** if:

1. X is in $\mathcal{F}$
2. If A is in $\mathcal{F}$, then so is the complement of A
3. if A_n is a countable sequence of sets in $\mathcal{F}$ then their union is in $\mathcal{F}$, i.e. $\displaystyle\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

This definition provides us with sets that can be considered *events* in a probability space, while elements not in $\mathcal{F}$ have no defined probability measure. σ-algebras allow us to avoid certain undefined behaviour which arises from non-measurable sets.

A function $\mu : \mathcal{F} \mapsto \mathbb{R}^+ \cup \{\infty\}$ is called a **measure** if:

1. $\mu(\emptyset) = 0$
2. if A_n is a countable collection of pairwise disjoint sets in $\mathcal{F}$ (i.e. $A_n \cap A_m = \emptyset$ for $n \neq m$) then
$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n).$$

We call $(X, \mathcal{F}, \mu)$ a measure space. If $\mu(X) = 1$ then we consider μ a probability measure similar to $0 \leq Pr(X) \leq 1$ and may refer to $(X, \mathcal{F}, \mu)$ as a probability space.

We say that T is a measure-preserving transformation or, equivalently, μ is said to be a T-invariant measure if the measure of a set is always the same as the measure of the set of points which map to it, i.e. $\mu(T^{-1}(A)) = \mu(A)$ for any measurable set $A \in \mathcal{F}$.

3 Ergodicity

Let $(X, \mathcal{F}, \mu)$ be a probability space and let $T : X \mapsto X$ be a measure-preserving transformation. We say that T is an **ergodic transformation**, or that μ is an **ergodic measure**, if whenever $A \in \mathcal{F}$ satisfies $T^{-1}(A) = A$, then $\mu(A) = 0$ or 1.

Ergodicity can be understood as a measure theoretic version of irreducibility, similar to the inability to split up markov chains into smaller sub-chains. Ergodicity can be viewed as an indecomposability condition and is concerned with how a typical orbit of a dynamical system is distributed throughout the phase space while ergodic theory studies the qualitative distributional properties of typical orbits of a dynamical system with these properties being expressed in terms of measure theory. An ergodic dynamical system is one in which, with respect to some probability distribution, all invariant sets either have measure 0 or 1.

4 Birkhoff's Ergodic Theorem

1. Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system on a σ-finite measure space $(\mu(X) < \infty)$, and let f be any integrable function.

2. We consider $\hat{f}$ the time-average of f,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=0}^{n-1} f(T^i(x)).$$

   $\hat{f}(x)$ converges for almost every $x \in X$, and is integrable i.e. $\hat{f}$ is in $L^1(\mu)$. Moreover, $\hat{f}$ is T-invariant, i.e., $\hat{f} \circ T = \hat{f}$.

3. We consider $\bar{f}$ the space-average of f,

$$\bar{f} = \frac{1}{\mu(x)}\int_X f d\mu.$$

   For a probability space, $\mu(X) = 1$.

4. If T is ergodic, then $\hat{f}$ is constant in μ almost everywhere, so we have $\hat{f} = \bar{f}$, i.e.

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

If a mapping is ergodic, as the number of finite averages taken along any of its orbit increases to infinity (the time-average), this value will converge to the continuous integral (the space average). That is, a finite average sampling of points of any orbit will be be as accurate as a continuous average integral over the entire state space.

5 Statistical Mechanics and the Law of Large Numbers

The term *ergodic* was originally introduced by Ludwig Boltzman, the founder of statistical mechanics, to describe a stronger but related property: starting from a random point in state space, orbits will typically pass through every point in state space. To understand this relation in terms of a mathematical space, consider the **indicator function** of A, $\mathbb{I}_{\mathbb{A}} : X \to \{0, 1\}$ defined by,

$$\mathbb{I}_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}.$$

The time-average of $\mathbb{I}_A$ is the fraction of time that the orbit spends in A while the space-average of I_A is the probability that a randomly picked point is in A. In an ergodic system, the two averages are almost always equal, meaning almost all trajectories end up covering the state space in the same way. The classical ergodic model is a version of the **law of large numbers** — the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed.

6 Ergodic Hierarchy

Ergodicity is at the bottom level of the ergodic hierarchy, where the higher a system is categorized the more random its behaviour. This unpredictability occurs as a result of a decay of correlations between the systems' past and present states.

$$\text{Bernoulli} \subset \text{Kolmogorov} \subset \text{Strong Mixing} \subset \text{Weak Mixing} \subset \text{Ergodic}.$$

We can gain an intuition for a mixing transformation, T, through a comparison with stirring two parts (A, B) of a fluid together, meaning that $T^n(A)$ is the region or n-orbit of the first part of fluid after n time units of mixing. If the volume of the entire fluid V is equal to C, then we consider the fluid to be thoroughly mixed if the concentration of the first part of fluid equals $\mu(A)/\mu(C)$ in both the whole volume of fluid and also in sub-region in the volume. That is, the fluid is thoroughly mixed at time n if

$$\frac{\mu(T^n A \cap V)}{\mu(V)} = \frac{\mu(A)}{\mu(C)}$$

for any volume V (of non-zero measure). We assume that the volume of the combined liquid is one unit, i.e. $\mu(C) = 1$, then for all subsets A and B of X a system is **strong mixing** (S-M) if,

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

A system is **weak mixing** (W-M) if,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} |\mu(T^k B \cap A) - \mu(B)\mu(A)| = 0.$$

A system is **K-mixing** (K-M) if for any subsets $A_0, A_1, ..., A_r$ of X, where r is an arbitrary natural number, the following condition holds:

$$\lim_{n\to\infty} \sup_{B\in\sigma(n,r)} |\mu(T^n B \cap A) - \mu(B)\mu(A)| = 0$$

A partition of X is a division of X into different mutually exclusive parts or atoms that jointly cover X. A system is a **Bernoulli system** if if there exists a partition α of X so that the images of α under its transformation T at different instants of time are independent, i.e.

$$\mu(\delta \in \beta) = \mu(\delta_i)\mu(\beta_j)$$

for all atoms δ_i of $T^k\alpha$ and all atoms β_j of $T^l\alpha$ for all $k \neq l$.

References

[1] https://personalpages.manchester.ac.uk/staff/charles.walkden/magic/

[2] https://plato.stanford.edu/entries/ergodic-hierarchy/

[3] http://bactra.org/notebooks/ergodic-theory.html

[4] https://terrytao.files.wordpress.com/2011/01/measure-book1.pdf