An introduction to ergodic theory

A state of an n-particle system M depends on $6n$ parameters. Each of the particles must be assigned three position and three momentum coordinates. When the parameters $p_1,\ldots, p_{3n}, q_1,\ldots, q_{3n}$ are assigned, the state of the system is fixed. Hence it is possible to represent each state as a point in a $6n$-dimensional space that is isomorphic to a subspace of the Euclidean E^{6n}. This space is the phase space of M. One of the way of describing the evolution of M is by stating its Hamiltonian $H(p_1,\ldots, p_{3n}, q_1,\ldots, q_{3n})$ using the equations

$$\partial H/\partial p_i = dq_i/dt, \quad \partial H/\partial q_i = -dp_i/dt.$$

The Hamiltonian can be described as a transformation from E^{6n} to itself, which determines for every state its dynamic evolution after an infinitesimal time interval. Every state lies on a unique trajectory that is determined by its Hamiltonian. And because the solutions are required to be reversible in time, every state has a unique past trajectory (backward motion) as well. Every possible state of M lies on exactly one such trajectory. The exclusiveness is the result of the uniqueness of the solutions of H, and the exhaustiveness is the result of the existence of solutions for H.

One of the results concerning Hamiltonian systems is Liouville's theorem. Let A_i be a set of states, and let $A_{t+\Delta t}$ be the set of the future evolutions of the members of A_t after Δt. Liouville's theorem says that the volumes of A_t and $A_{t+\Delta t}$ are the same. This property of Hamiltonian systems is known as *stationary* or *incompressible*.

Lets abstract from Hamiltonian only its trajectory that it assigns to every state. First, we look at the Hamiltonian as a topological group of operations $\{U_t\}_{t\in R}$; for every $t\in R$, U_t: $E^{6n}\rightarrow E^{6n}$ assigns to every state $s\in E^{6n}$ its future at t. The next step is to replace the continuous structure of $\{U_t\}_{t\in R}$ with a discrete $\{U_t\}_{t\in Z}$ (Z is the set of integers). We may do so by defining a transformation T: $E^{6n}\rightarrow E^{6n}$ such that, for all $s\in E^{6n}$, $T(s)$ is the future of s after a short interval of time; we then identify $\{U_t\}_{t\in Z}$ with the group of iteration T^i.

Let $s\in E^{6n}$ be an arbitrary state. The trajectory of s under T is $O_{T(s)} = \{s, T(s), T^2(s),\ldots\}$. Let A be an arbitrary set of states. The time average that M stays in A given its present state s is $O_T(s, A) = \lim_{n\to\infty} 1/n \sum_{i=1} X_A(T^i(s))$. ($X_A$ is the characteristic function of A.) $O_T(s, A)$ can be thought of as the frequency in which $O_T(s)$ assumes states that are located in A. Let A_E be the set of states that are compatible with macroscopic equilibrium. We may ask the following questions:

1. Given that M is in the state s now, how often will M assume states that are compatible with macroscopic equilibrium, or what is $O_T(s, A_E)$?
2. Under which circumstances will $O_T(s, A_E)$ be close to 1?
3. Under which circumstances will $O_T(s, A_E)$ be independent of the choice of s?

Recall that: When M is made of many particles, the volume of A_E is overwhelmingly large. In this case we are well on our way to justifying our expectations that M will reach a persisting equilibrium independent of its initial conditions. Ergodic systems are those for which the above-mentioned conditions are satisfied.

In ergodic theory, the basic mathematical structure is a quadruple $<S, B, m, T>$, which is called a *dynamic system*. S is a set of possible states of M, B is a σ-algebra of subsets of S, m: $B\rightarrow R$ is a measure, and T: $S\rightarrow S$ is the evolution transformation.

Definition. T is ergodic iff, for all $A\in B$, $T^{-1}(A) = A$ implies that $m(A) = 0$ or $m(A) = 1$.

Theorem 1 (Birkhoff 1931) If T is ergodic, then for all $f\in L^2$, and for m-almost every $s\in S$,

$$lim_{n\rightarrow\infty} 1/n \sum f(T^{1}(s)) = f^{*},$$

where $f^{*} \in L^{2}$. (L^{2} is the space of the square-integrable functions. # A random variable $f{:}\Omega \rightarrow R$ is square-integrable if $\int |f|^{2}\, dP < \infty$.)

Corollary 2. If T is ergodic, for every $A \in B$ and for m-almost every $s \in S$,

$$lim_{n\rightarrow\infty} 1/n \sum X_{A}(T^{i}(s)) = m(A).$$

If T is ergodic, the time averages $O_{T}(s, A)$ are the same as the measure of A for every $A \in B$ and for almost every $s \in S$. In particular, if $m(A_{E}) = 1$ and T is ergodic, $O_{T}(s, A) = 1$ almost always.

Theorem 2. If T is ergodic, for every $A, C \in B$,

$$lim_{n\rightarrow\infty} 1/n \sum f(T^{1}(C) \cap A) = m(A)m(C).$$

Corollary 1.

$$1/n \sum X_{A}(T^{i}(S))X_{C} \rightarrow m(A)X_{C} \text{ for almost every } s \in S.$$

Theorem 2 states that the ergodicity of T implies that, when we calculate $O_{T}(s, C)$ using only s that are members of A, the result will not be biased, on the average, because the C are distributed among the A in the same proportion that they are distributed among the general population. This is precisely why we can discard additional information when we predict the future state of an ergodic system. In particular, this explains why any extra knowledge concerning M will not affect our predictions concerning whether and how often M will assume an equilibrium state.

One example of a stochastic process is a coin that is tossed repeatedly. Such a process can be described as a dynamic system in the following way:

1 As the space of states S we shall take all of the infinite sequences of heads and tails. Each sequence $\ldots w_{-1}, w_{0}, w_{1}, \ldots$ is an exhaustive description of one of the possible outcomes of the infinite coin-tossing process.
2 We shall take the σ-algebra generated by the cylindrical sets. Cylindrical sets define sets of sequences by stipulating that their ith coordinate should have a particular value. For example, "the set of all sequences that have heads in their 103rd place" is a cylindrical set.
3 As T we may take the shift transformation. The shift is defined as $T(\ldots w_{-1}, w_{0}, w_{1}, \ldots) = \ldots w_{0}, w_{1}, w_{2}, \ldots$ The shift assigns to every sequence a different sequence that has the same elements ordered in the same way but parametrized differently. T can be thought of as the "unfolding in time" of a particular sequence. The shift reveals the identity of the result of the next toss at each moment.
4 Let $A \in B$. For example, take

$$A_{103} = \{\ldots w_{-1}, w_{0}, w_{1}, \ldots | w_{103} = \text{heads}\}.$$

The quantity $m(A_{103})$ measures the "relative proportion" of sequences that have heads in the 103th coordinate. We say that m is stationary if it id shift-invariant, that is, if m assigns the same probability to the event that heads appears in the 103rd coordinate as it assigns to the appearance of heads in any other coordinate. (This rationale for stationarity is unacceptable for subjectivists. For them, the fact that the tossing mechanism remains the same is besides the point.)

5 The conditional measure $m(A|A')$ is interpreted as the "relative proportion" of the sequences that have the property A among those that have the property A'. Measures for which for all $A, A' \in B$, $m(A|A') = m(A)$ are called Bernoulli measures.

6 For the definition of an arbitrary stochastic process, replace the set {heads, tails} with a fixed partition $p_1, \ldots, p_n$. The assumption is that at each moment exactly one of the p_i will happen.

7 The trajectory $O_T(s) = ..T^{-1}(s), s, T(s), \ldots$ of an element of a stochastic process that is shifted is the infinite sequence infinite sequences $(\ldots w_{-1}, w_0, w_1, \ldots), (\ldots w_0, w_1, w_2, \ldots), \ldots, (\ldots w_n, w_{n+1}, w_{n+2}, \ldots), \ldots$ The time average $O_T(s, A_{103})$ is the measure of the relative frequency of the sequences in $O_T(s)$ that have heads in the 103th coordinate. Note that $O_T(s, A_{103}) = O_T(s, A_{102})$ because $O_T(s)$ has the same set of sequences but whose coordinates were shifted to the left.

Ergodicity in the shift space means that the "future" coordinates of a sequence are asymptotic independent of the present coordinates. Therefore, Bernoulli measures are clearly ergodic, because independence implies asymptotic independence.

Ergodic Theory

Ergodic Theory is the study of dynamical systems with an invariant probability measure. Imagine we start out with some dynamical system $T:X \rightarrow X$. This means that X is a statespace, and T is a map from X to itself. Each step of time we move forward corresponds to an application of T to X, transforming the state of the system. Now imagine that we endow X with a probability distribution, so that the probability of some event occurring is unchanging over time. That is:

For any event A, the probability of A occurring at time t is the same as the probability of A occurring at time t+1

Formally, the "event" A corresponds to some subset of the space X. Suppose our probability measure was called P. The occurrence of the event A "right now" means that the state of the system (a point in X) currently resides inside A. The occurrence of the event A during the next time iteration means that the future state of the system, $T(x)$, resides in A; or, conversely, the current state resides in $T^{-1}(A)$ The probability of event A happening "right now" is thus $P(A)$. The probability of event A happening during the next time iteration is then $P(T^{-1}(A))$. So, formally, we can rewrite the previous hypothesis:

$$\text{For any measurable subset } A \text{ in } X, \quad P(A) = P(T^{-1}(A))$$

Such a probability measure is called an *invariant* measure, and a dynamical system equipped with an invariant measure is called a measure-preserving dynamical system. These are what ergodic theorists study.

Intuitively, the probability measure on the dynamical system reflects the amount of information which you, the observer, have, a priori concerning the current "state" of the system. Ergodic theorists are particularly interested in questions concerning how rapidly this information is destroyed over time. The kinds of systems ergodic theorists study could be called "chaotic" - they are systems where the information you have about the system's current state rapidly becomes obsolete. The statespace gets "mixed up" so quickly that your current information rapidly becomes useless in predicting the future state of the system.

For example, suppose that you know that the system, right now, is in state A. Suppose you want to answer the question: "After N time steps (i.e. after N iterations of the transformation T), how likely is it that the system will be in state B?"

We can pose this question formally:

Given that we know the system is currently in state A, what is the probability of $T^N(B)$? In other words, what value will the conditional probability $P(T^N(B)\,|\,A)$ take?

If this conditional probability is near 1 or 0, then knowledge of event A gives you a lot of predictive power about the future occurrence of event B. On the other hand, if $P(T^{-N}(B)\,|\,A)$ is close to $P(B)$, then knowledge of A tells you nothing about the possible future occurrence of B.

We say that a measure-preserving dynamical system is strongly mixing if, for any pair of events A and B (of nonzero probability),

$$lim_{N \to \infty} P(T\text{-}N(B)\,|\,A) = P(B)$$

What this means is that predicting the occurrence of any event B totally impossible, as we move far into the future. If I give you information about the current state of the system, (in the form of specifying some subset A of nonzero probability), then, no matter how accurate that information is (in other words, no matter how small the probability of A is), over time, the amount of information A tells you about the future occurrence of B dwindles to nothing.

This is a probabilistic way of formulating the idea of chaos. The term "chaos" actually refers to a property of topological dynamical systems, and what it means is that two points in the statespace, no matter how close together they start out, will eventually be "peeled apart", and separated by arbitrarily large distances. This image inspires the catchy slogan, "Sensitive dependence upon initial conditions."

In the field of topological dynamics, there are many concrete instantiations of this vision; topological mixing, nonzero topological entropy, expansive dynamics, etc. Ergodic theorists work in a more abstract setting, forgetting about the actual specific topology, and concentrating on questions of information; for them, the question is "How rapidly is information about the system destroyed?"

The concept of strong mixing is one way of measuring this. A weaker notion of mixing exists, called (predictably) weak mixing. Weak mixing has many formulations, but one of the simplest is:

A measure-preserving dynamical system is weakly mixing if it has no nontrivial eigenvalues

Peeling back the layers of jargon, what this basically means is that the dynamical system possesses no periodic behavior. Many dynamical systems are periodic in at least some "aspects". That is, even though the system as a whole is not periodic, certain measurements made of the system will vary in a periodic fashion over time, thus demonstrating that, somehow, trapped within the system, are periodic "components". A weakly mixing system has no such periodic components; its behavior is entirely nonperiodic.

The different mixing concepts measure the degree to which currently valid information about the system is basically completely destroyed as time progresses asymptotically into the far future, but they deal only with this asymptotic "far future" picture (i.e. the limit as time "goes to infinity"). An entirely different notion of chaos deals with the rate at which information is being destroyed right now. This is called entropy, and is closely related to the famous measure of information invented by Claude Shannon, progenitor of information theory.