

Chapter 25

Ergodicity and Metric Transitivity

Section 25.1 explains the ideas of ergodicity (roughly, there is only one invariant set of positive measure) and metric transitivity (roughly, the system has a positive probability of going from anywhere to anywhere), and why they are (almost) the same.

Section 25.2 gives some examples of ergodic systems.

Section 25.3 deduces some consequences of ergodicity, most importantly that time averages have deterministic limits (§25.3.1), and an asymptotic approach to independence between events at widely separated times (§25.3.2), admittedly in a very weak sense.

25.1 Metric Transitivity

Definition 341 (Ergodic Systems, Processes, Measures and Transformations) *A dynamical system Ξ, \mathcal{X}, μ, T is ergodic, or an ergodic system or an ergodic process when $\mu(C) = 0$ or $\mu(C) = 1$ for every T -invariant set C . μ is called a T -ergodic measure, and T is called a μ -ergodic transformation, or just an ergodic measure and ergodic transformation, respectively.*

Remark: Most authorities require a μ -ergodic transformation to also be measure-preserving for μ . But (Corollary 54) measure-preserving transformations are necessarily stationary, and we want to minimize our stationarity assumptions. So what most books call “ergodic”, we have to qualify as “stationary and ergodic”. (Conversely, when other people talk about processes being “stationary and ergodic”, they mean “stationary with only one ergodic component”; but of that, more later.

Definition 342 (Metric Transitivity) *A dynamical system is metrically transitive, metrically indecomposable, or irreducible when, for any two sets $A, B \in \mathcal{X}$, if $\mu(A), \mu(B) > 0$, there exists an n such that $\mu(T^{-n}A \cap B) > 0$.*

Remark: In dynamical systems theory, metric transitivity is contrasted with *topological* transitivity: T is topologically transitive on a domain D if for any two open sets $U, V \subseteq D$, the images of U and V remain in D , and there is an n such that $T^n U \cap V \neq \emptyset$. (See, e.g., Devaney (1992).) The “metric” in “metric transitivity” refers not to a distance function, but to the fact that a measure is involved. Under certain conditions, metric transitivity in fact implies topological transitivity: e.g., if D is a subset of a Euclidean space and μ has a positive density with respect to Lebesgue measure. The converse is not generally true, however: there are systems which are transitive topologically but not metrically.

A dynamical system is *chaotic* if it is topologically transitive, and it contains dense periodic orbits (Banks *et al.*, 1992). The two facts together imply that a trajectory can start out arbitrarily close to a periodic orbit, and so remain near it for some time, only to eventually find itself arbitrarily close to a *different* periodic orbit. This is the source of the fabled “sensitive dependence on initial conditions”, which paradoxically manifests itself in the fact that all typical trajectories look pretty much the same, at least in the long run. Since metric transitivity generally implies topological transitivity, there is a close connection between ergodicity and chaos; in fact, most of the well-studied chaotic systems are also ergodic (Eckmann and Ruelle, 1985), including the logistic map. However, it is possible to be ergodic without being chaotic: the one-dimensional rotations with irrational shifts are, because their periodic orbits do not exist, and *a fortiori* are not dense.

Lemma 343 (Metric transitivity implies ergodicity) *If a dynamical system is metrically transitive, then it is ergodic.*

PROOF: By contradiction. Suppose there was an invariant set A whose μ -measure was neither 0 nor 1; then A^c is also invariant, and has strictly positive measure. By metric transitivity, for some n , $\mu(T^{-n}A \cap A^c) > 0$. But $T^{-n}A = A$, and $\mu(A \cap A^c) = 0$. So metrically transitive systems are ergodic. \square

There is a partial converse.

Lemma 344 (Stationary Ergodic Systems are Metrically Transitive) *If a dynamical system is ergodic and stationary, then it is metrically transitive.*

PROOF: Take any $\mu(A), \mu(B) > 0$. Let $A_{\text{ever}} \equiv \bigcup_{n=0}^{\infty} T^{-n}A$ — the union of A with all its pre-images. This set contains its pre-images, $T^{-1}A_{\text{ever}} \subseteq A_{\text{ever}}$, since if $x \in T^{-n}A$, $T^{-1}x \in T^{-(n+1)}A$. The sequence of pre-images is thus non-increasing, and so tends to a limiting set, $\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} T^{-k}A = A_{\text{i.o.}}$, the set of points which not only visit A eventually, but visit A infinitely often. This is an invariant set (Lemma 306), so by ergodicity it has either measure 0 or measure 1. By the Poincaré recurrence theorem (Corollaries 67 and 68), since $\mu(A) > 0$, $\mu(A_{\text{i.o.}}) = 1$. Hence, for any B , $\mu(A_{\text{i.o.}} \cap B) = \mu(B)$. But this means that, for some n , $\mu(T^{-n}A \cap B) > 0$, and the process is metrically transitive. \square

25.2 Examples of Ergodicity

Example 345 (IID Sequences, Strong Law of Large Numbers) *Every IID sequence is ergodic. This is because the Kolmogorov 0-1 law states that every tail event has either probability 0 or 1, and (Exercise 25.3) every invariant event is a tail event. The strong law of large numbers is thus a two-line corollary of the Birkhoff ergodic theorem.*

Example 346 (Markov Chains) *In the elementary theory of Markov chains, an ergodic chain is one which is irreducible, aperiodic and positive recurrent. To see that such a chain corresponds to an ergodic process in the present sense, look at the shift operator on the sequence space. For consistency of notation, let S_1, S_2, \dots be the values of the Markov chain in Σ , and X be the semi-infinite sequence in sequence space Ξ , with shift operator T , and distribution μ over sequences. μ is the product of an initial distribution $\nu \sim S_1$ and the Markov-family kernel. Now, “irreducible” means that one goes from every state to every other state with positive probability at some lag, i.e., for every $s_1, s_2 \in \Sigma$, there is an n such that $\mathbb{P}(S_n = s_2 | S_1 = s_1) > 0$. But, writing $[s]$ for the cylinder set in Ξ with base s , this means that, for every $[s_1], [s_2]$, $\mu(T^{-n}[s_2] \cap [s_1]) > 0$, provided $\mu([s_1]) > 0$. The Markov property of the S chain, along with positive recurrence, can be used to extend this to all finite-dimensional cylinder sets (Exercise 25.4), and so, by a generating-class argument, to all measurable sets.*

Example 347 (Deterministic Ergodicity: The Logistic Map) *We have seen that the logistic map, $Tx = 4x(1-x)$, has an invariant density (with respect to Lebesgue measure). It has an infinite collection of invariant sets, but the only invariant interval is the whole state space $[0, 1]$ — any smaller interval is not invariant. From this, it is easy to show that all the invariant sets either have measure 0 or measure 1 — they differ from \emptyset or from $[0, 1]$ by only a countable collection of points. Hence, the invariant measure is ergodic. Notice, too, that the Lebesgue measure on $[0, 1]$ is ergodic, but not invariant.*

Example 348 (Invertible Ergodicity: Rotations) *Let $\Xi = [0, 1)$, $Tx = x + \phi \bmod 1$, and let μ be the Lebesgue measure on Ξ . (This corresponds to a rotation, where the angle advances by $2\pi\phi$ radians per unit time.) Clearly, T preserve μ . If ϕ is rational, then, for any x , the sequence of iterates will visit only finitely many points, and the process is not ergodic, because one can construct invariant sets whose measure is neither 0 nor 1. (You may construct such a set by taking any one of the periodic orbits, and surrounding its points by intervals of sufficiently small, yet positive, width.) If, on the other hand, ϕ is irrational, then $T^n x$ never repeats, and it is easy to show that the process is ergodic, because it is metrically transitive. Nonetheless, T is invertible.*

This example (suitably generalized to multiple coordinates) is very important in physics, because many mechanical systems can be represented in terms of

“action-angle” variables, the speed of rotation of the angular variables being set by the actions, which are conserved, energy-like quantities. See Mackey (1992); Arnol’d and Avez (1968) for the ergodicity of rotations and its limitations, and Arnol’d (1978) for action-angle variables. Astonishingly, the result for the one-dimensional case was proved by Nicholas Oresme in the 14th century (von Plato, 1994).

Example 349 (Ergodicity when the Distribution Does Not Converge)

Ergodicity does not ensure a uni-directional evolution of the distribution. (Some people (Mackey, 1992) believe this has great bearing on the foundations of thermodynamics.) For a particularly extreme example, which also illustrates why elementary Markov chain theory insists on aperiodicity, consider the period-two deterministic chain, where state A goes to state B with probability 1, and vice versa. Every sample path spends just much time in state A as in state B , so every time average will converge on $\mathbf{E}_m[f]$, where m puts equal probability on both states. It doesn’t matter what initial distribution we use, because they are all ergodic (the only invariant sets are the whole space and the empty set, and every distribution gives them probability 1 and 0, respectively). The uniform distribution is the unique stationary distribution, but other distributions do not approach it, since $U^{2n}\nu = \nu$ for every integer n . So, $A_t f \rightarrow \mathbf{E}_m[f]$ a.s., but $\mathcal{L}(X_n) \not\rightarrow m$. We will see later that aperiodicity of Markov chains connects to “mixing” properties, which do guarantee stronger forms of distributional convergence.

25.3 Consequences of Ergodicity

The most basic consequence of ergodicity is that all invariant functions are constant almost everywhere; this in fact characterizes ergodicity. This in turn implies that time-averages converge to deterministic, rather than random, limits. Another important consequence is that events widely separated in time become nearly independent, in a somewhat funny-looking sense.

Theorem 350 (Ergodicity and the Triviality of Invariant Functions)

A T transformation is μ -ergodic if and only if all T -invariant observables are constant μ -almost-everywhere.

PROOF: “Only if”: Because invariant observables are \mathcal{I} -measurable (Lemma 304), the pre-image under an invariant observable f of any Borel set B is an invariant set. Since every invariant set has μ -probability 0 or 1, the probability that $f(x) \in B$ is either 0 or 1, hence f is constant with probability 1. “If”: The indicator function of an invariant set is an invariant function. If all invariant functions are constant μ -a.s., then for any $A \in \mathcal{I}$, either $\mathbf{1}_A(x) = 0$ or $\mathbf{1}_A(x) = 1$ for μ -almost all x , which is the same as saying that either $\mu(A) = 0$ or $\mu(A) = 1$, as required. \square

25.3.1 Deterministic Limits for Time Averages

Theorem 351 (The Ergodic Theorem for Ergodic Processes) *Suppose μ is AMS, with stationary mean m , and T -ergodic. Then, almost surely,*

$$\lim_{t \rightarrow \infty} A_t f(x) = \mathbf{E}_m[f] \quad (25.1)$$

for μ - and m - almost all x , for any $L_1(m)$ observable f .

PROOF: Because every invariant set has μ -probability 0 or 1, it likewise has m -probability 0 or 1 (Lemma 329). Hence, $\mathbf{E}_m[f]$ is a version of $\mathbf{E}_m[f|\mathcal{I}]$. Since $A_t f$ is also a version of $\mathbf{E}_m[f|\mathcal{I}]$ (Corollary 340), they are equal almost surely. \square

An important consequence is the following. Suppose S_t is a strictly stationary random sequence. Let $\Phi_t(S) = f(S_{t+\tau_1}, S_{t+\tau_2}, \dots, S_{t+\tau_n})$ for some fixed collection of shifts τ_n . Then Φ_t is another strictly stationary random sequence. Every strictly stationary random sequence can be represented by a measure-preserving transformation (Theorem 52), where X is the sequence S_1, S_2, \dots , the mapping T is just the shift, and the measure μ is the infinite-dimensional measure of the original stochastic process. Thus $\Phi_t = \phi(X_t)$, for some measurable function ϕ . If the measure is ergodic, and $\mathbf{E}[\Phi]$ is finite, then the time-average of Φ converges almost surely to its expectation. In particular, let $\Phi_t = S_t S_{t+\tau}$. Then, assuming the mixed moments are finite, $t^{-1} \sum_{t=1}^{\infty} S_t S_{t+\tau} \rightarrow \mathbf{E}[S_t S_{t+\tau}]$ almost surely, and so the sample covariance converges on the true covariance. More generally, for a stationary ergodic process, if the n -point correlation functions exist, the sample correlation functions converge a.s. on the true correlation functions.

25.3.2 Ergodicity and the approach to independence

Lemma 352 (Ergodicity Implies Approach to Independence) *If μ is T -ergodic, and μ is AMS with stationary mean m , then*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{n=0}^{t-1} \mu(B \cap T^{-n}C) = \mu(B)m(C) \quad (25.2)$$

for any measurable events B, C .

PROOF: Exercise 25.1. \square

Theorem 353 (Approach to Independence Implies Ergodicity) *Suppose \mathcal{X} is generated by a field \mathcal{F} . Then an AMS measure μ , with stationary mean m , is ergodic if and only if, for all $F \in \mathcal{F}$,*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{n=0}^{t-1} \mu(F \cap T^{-n}F) = \mu(F)m(F) \quad (25.3)$$

i.e., iff Eq. 25.2 holds, taking $B = C = F \in \mathcal{F}$.

PROOF: “Only if”: Lemma 352. “If”: Exercise 25.2. \square

25.4 Exercises

Exercise 25.1 (Ergodicity implies an approach to independence) *Prove Lemma 352.*

Exercise 25.2 (Approach to independence implies ergodicity) *Prove the “if” part of Theorem 353.*

Exercise 25.3 (Invariant events and tail events) *Prove that every invariant event is a tail event. Does the converse hold?*

Exercise 25.4 (Ergodicity of ergodic Markov chains) *Complete the argument in Example 346, proving that ergodic Markov chains are ergodic processes (in the sense of Definition 341).*