Chapter 25

Ergodicity

This lecture explains what it means for a process to be ergodic or metrically transitive, gives a few characterizes of these properties (especially for AMS processes), and deduces some consequences. The most important one is that sample averages have deterministic limits.

25.1 Ergodicity and Metric Transitivity

Definition 300 *A dynamical system* $\Xi, \mathcal{X}, \mu, T$ *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 301 *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 there periodic orbits do not exist, and *a fortiori* are not dense.

Proposition 302 *A dynamical system is ergodic if it is metrically transitive.*

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. □

There is a partial converse.

Proposition 303 *If a dynamical systems is ergodic and stationary, then it is metrically transitive.*

PROOF: Take any $\mu(A), \mu(B) > 0$. Let $A_{\mathrm{ever}} \equiv \bigcup_{n=0}^{\infty} T^{-n}A$ — the union of A with all its pre-images. This set contains its pre-images, $T^{-1}A_{\mathrm{ever}} \subseteq A_{\mathrm{ever}}$, 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, $\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} T^{-k}A = A_{\mathrm{i.o.}}$, the set of points which not only visit A eventually, but visit A infinitely often. This is an invariant set (Lemma 264), so by ergodicity it has either measure 0 or measure 1. By the Poincaré recurrence theorem (Corollaries 66 and 67), since $\mu(A) > 0$, $\mu(A_{\mathrm{i.o.}}) = 1$. Hence, for any B, $\mu(A_{\mathrm{i.o.}} \cap B) = \mu(B)$. But this means that, for some n, $\mu(T^{-n}A \cap B) > 0$, and the process is metrically transitive. □

Theorem 304 *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 262), 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. □

Lemma 305 *If μ is T-ergodic, and μ is AMS with stationary mean m, then*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} \mu(B \cap T^{-n}C) = \mu(B)m(C) \tag{25.1}$$

for any measurable events B, C.

PROOF: Exercise. □

Theorem 306 *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 $F \in \mathcal{F}$,*

$$\lim_{t\to\infty} \frac{1}{t} \sum_{n=0}^{t-1} \mu(F \cap T^{-n}F) = \mu(F)m(F) \tag{25.2}$$

i.e., iff Eq. 25.1 holds, taking $B = C = F \in \mathcal{F}$.

PROOF: "Only if": Lemma 305. "If": Exercise. □

25.1.1 Examples of Ergodicity

Example 307 (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!) 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 308 (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 $S_1, S_2, \ldots$ 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!), and so, by a generating-class argument, to all measurable sets.*

Example 309 (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 310 (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 internals 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 311 *Ergodicity does not ensure a uni-directional evolution of the density. (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 stae 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 approch 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.*

25.1.2 Consequences of Ergodicity

The most basic consequence of ergodicity is that time-averages converge to deterministic, rather than random, limits.

Theorem 312 *Suppose μ is AMS, with stationary mean m, and T-ergodic. Then, almost surely,*

$$\lim_{t\rightarrow\infty} A_t f(x) = \mathbf{E}_m\left[f\right] \tag{25.3}$$

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 287). Hence, $\mathbf{E}_m\left[f\right]$ is a version of $\mathbf{E}_m\left[f|\mathcal{I}\right]$. Since $A_t f$ is also a version of $\mathbf{E}_m\left[f|\mathcal{I}\right]$ (Corollary 299), they are equal almost surely. □

An important consequence is the following. Suppose S_t is a strictly stationary random sequence. Let $\Phi_t(S) = f(S_{t+\tau_1}, S_{t+\tau_2}, \ldots S_{t+\tau_n})$ 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 $S_1, S_2, \ldots$, 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 $\mathbf{E}\left[\Phi\right]$ 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, $t^{-1}\sum_{t=1}^{\infty} S_t S_{t+\tau} \rightarrow \mathbf{E}\left[S_t S_{t+\tau}\right]$ 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.

25.2 Preliminaries to Ergodic Decompositions

It is always the case, with a dynamical system, that if x lies within some invariant set A, then all its future iterates stay within A as well. In general, therefore, one might expect to be able to make some predictions about the future trajectory by knowing which invariant sets the initial condition lies within. An ergodic process is one where this is actually not possible. Because all invariants sets have probability 0 or 1, they are all independent of each other, and indeed of every other set. Therefore, knowing which invariant sets x falls into is *completely uninformative* about its future behavior. In the more general non-ergodic case, a limited amount of prediction is however possible on this basis, the limitations being set by the way the state space breaks up into invariant sets of points with the same long-run average behavior — the ergodic components. Put slightly differently, the long-run behavior of an AMS system can be represented as a mixture of stationary, ergodic distributions, and the ergodic components are, in a sense, a minimal parametrically sufficient statistic for this distribution. (They are not in generally *predictively* sufficient.)

The idea of an ergodic decomposition goes back to von Neumann, but was considerably refined subsequently, especially by the Soviet school, who seem to have introduced most of the talk of predictions, and all of the talk of ergodic components as minimal sufficient statistics. Our treatment will follow Gray (1988, ch. 7), and Dynkin (1978). The rest of this lecture will handle some preliminary propositions about combinations of stationary measures.

Proposition 313 *Any convex combination of invariant probability measures is an invariant probability measure.*

PROOF: Let μ_1 and μ_2 be two invariant probability measures. It is elementary that for every $0 \leq a \leq 1$, $\nu \equiv a\mu_1 + (1-a)\mu_2$ is a probability measure. Now consider the measure under ν of the pre-image of an arbitrary measurable set $B \in \mathcal{X}$:

$$\begin{aligned} \nu(T-1B) &= a\mu_1(T^{-1}B) + (1-a)\mu_2(T^{-1}B) && (25.4) \\ &= a\mu_1(B) + (1-a)\mu_2(B) && (25.5) \\ &= \nu(B) && (25.6) \end{aligned}$$

so ν is also invariant. □

Proposition 314 *If μ_1 and μ_2 are invariant ergodic measures, then either $\mu_1 = \mu_2$, or they are singular, meaning that there is a set B on which $\mu_1(B) = 0$, $\mu_2(B) = 1$.*

PROOF: Suppose $\mu_1 \neq \mu_2$. Then there is at least one set C where $\mu_1(C) \neq \mu_2(C)$. Because both μ_i are stationary and ergodic, $A_t\mathbf{1}_C(x)$ converges to $\mu_i(C)$ for μ_i-almost-all x. So the set

$$\left\{x | \lim_t A_t \mathbf{1}_C(x) = \mu_2(C)\right\}$$

has a μ_2 measure of 1, and a μ_1 measure of 0 (since, by hypothesis, $\mu_1(C) \neq \mu_2(C)$. □

Proposition 315 *Ergodic invariant measures are extremal points of the convex set of invariant measures, i.e., they cannot be written as combinations of other invariant measures.*

PROOF: By contradiction. That is, suppose μ is ergodic and invariant, and that there were invariant measures ν and λ, and an $a \in (0,1)$, such that $\mu = a\nu + (1-a)\lambda$. Let C be any invariant set; then $\mu(C) = 0$ or $\mu(C) = 1$. Suppose $\mu(C) = 0$. Then, because a is strictly positive, it must be the case that $\nu(C) = \lambda(C) = 0$. If $\mu(C) = 1$, then C^c is also invariant and has μ-measure 0, so $\nu(C^c) = \lambda(C^c) = 0$, i.e., $\nu(C) = \lambda(C) = 1$. So ν and λ would both have to be ergodic, with the same support as μ. But then (Proposition 314 preceeding) $\lambda = \nu = \mu$. □

Remark: The converse is left as an exercise (25.2).

25.3 Exercises

Exercise 25.1 *Prove Lemma 305.*

Exercise 25.2 *Prove the converse to Proposition 315: every extermal point of the convex set of invariant measures is an ergodic measure.*