Notes 25 : Ergodic theory: a brief introduction

Math 733-734: Theory of Probability — *Lecturer: Sebastien Roch*

References: [Dur10, Sections 7.1-4].

We give a very brief introduction to the ergodic theorem as well as the subadditive ergodic theorem. For more, see e.g. [Dur10, Chapter 7].

1 Stationary stochastic processes

The context for ergodic theory is stationary sequences, as defined next.

1.1 Definitions and main examples

We use the notation $\sim$ to indicate identity in distribution.

DEF 25.1 (Stationary stochastic process) *A real-valued process $\{X_n\}_{n\geq 0}$ is stationary if for every k, m*

$$(X_m, \ldots, X_{m+k}) \sim (X_0, \ldots, X_k).$$

EX 25.2 *IID sequences are stationary.*

EX 25.3 *Let $\{X_n\}$ be a MC on a countable set S with transition probability p and stationary distribution $\pi > 0$. Then $\{X_n\}$ with initial distribution π is a stationary stochastic process. Indeed, by definition of π and induction*

$$X_0 \sim X_n,$$

for all $n \geq 0$. Moreover, for all m, k, by the Markov property

$$(X_0, \ldots, X_k) \sim (X_m, \ldots, X_{m+k}).$$

1.2 Abstract setting

Ergodic theory is typically developed in a more abstract setting that encompasses the above.

EX 25.4 (A canonical example) *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A map $T : \Omega \to \Omega$ is said to be measure-preserving (for $\mathbb{P}$) if for all $A \in \mathcal{F}$,*

$$\mathbb{P}[\omega : T\omega \in A] = \mathbb{P}[T^{-1}A] = \mathbb{P}[A].$$

If $X \in \mathcal{F}$ then $X_n(\omega) = X(T^n\omega)$, $n \geq 0$, defines a stationary sequence. Indeed, for all $B \in \mathcal{B}(\mathbb{R}^{k+1})$

$$\begin{aligned}\mathbb{P}[(X_0, \ldots, X_k)(\omega) \in B] &= \mathbb{P}[(X_0, \ldots, X_k)(T^m\omega) \in B] \\ &= \mathbb{P}[(X_m, \ldots, X_{m+k})(\omega) \in B].\end{aligned}$$

Kolmogorov's extension theorem indicates that all real-valued stationary stochastic processes can be realized in the framework of the previous example. Recall:

THM 25.5 (Kolmogorov Extension Theorem) *Suppose we are given probability measure μ_n on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ s.t.*

$$\mu_{n+1}((a_0, b_0] \times \cdots \times (a_n, b_n] \times \mathbb{R}) = \mu_n((a_0, b_0] \times \cdots \times (a_n, b_n]),$$

for all n and $(n+1)$-dimensional rectangles. Then there exists a unique probability measure $\mathbb{P}$ on $(\mathbb{R}^{\mathbb{Z}_+}, \mathcal{R}^{\mathbb{Z}_+})$ with marginals μ_n.

EX 25.6 (Revisiting stationary processes) *Let $\tilde{X}$ be a stationary process on $\mathbb{R}$. Then by the previous theorem, we can realize $\tilde{X}$ on $\mathbb{R}^{\mathbb{Z}_+}$ as*

$$X_n(\omega) = \omega_n.$$

The corresponding measure-preserving transformation is the shift

$$T\omega = (\omega_1, \ldots).$$

In particular, $X_n(\omega) = X_0(T^n\omega)$.

2 Ergodic theorem

Before stating the ergodic theorem, we need a few more definitions. We are interested in the convergence of empirical averages:

$$n^{-1}S_n(\omega) = n^{-1}\sum_{m=0}^{n-1} X_m(\omega) = n^{-1}\sum_{m=0}^{n-1} f(T^m\omega).$$

2.1 Invariant sets

To get some intuition in the behavior of $n^{-1}S_n$ we look at a trivial example.

EX 25.7 *Let $\Omega = \{a, b, c, d, e\}$ and $\mathcal{F} = 2^{\Omega}$. Take $f = \mathbb{1}_A$ for some set $A \in \mathcal{F}$.*

1. *Suppose $T = (a, b, c, d, e)$ (i.e. the cyclic permutation that sends a to b etc.). For T to be measure-preserving we require $\mathbb{P}[a] = \mathbb{P}[b] = \cdots$ so that $\mathbb{P}[a] = 1/5$ is the only possibility. (It is easy to see that T is indeed measure-preserving because the number of elements of Ω is invariant under T.) In that case, it is immediate that*
$$n^{-1}S_n \to \mathbb{P}[A] = \mathbb{E}[f].$$
2. *Suppose $T = (a, b, c)(d, e)$ (i.e. the permutation with the two cycles listed). Let $\Omega_1 = \{a, b, c\}$, $\mathcal{F}_1 = 2^{\Omega_1}$, $\Omega_2 = \{d, e\}$ and $\mathcal{F}_2 = 2^{\Omega_2}$. For T to be measure-preserving we need $\mathbb{P}[a] = \mathbb{P}[b] = \mathbb{P}[c] = \alpha/3$ and $\mathbb{P}[d] = \mathbb{P}[e] = \beta/2$. (Any choice of α, β with $\alpha + \beta = 1$ works because the number of elements of Ω_1 and Ω_2 is invariant under T.) Take $A = \{a, d\}$. Then $n^{-1}S_n \to 1/3$ with probability α (i.e. if $\omega \in \Omega_1$) and $n^{-1}S_n \to 1/2$ with probability β. Denoting $\hat{f}$ this limit, we note*
$$\mathbb{E}[\hat{f}] = \mathbb{P}[A] = \mathbb{E}[f],$$
but $\hat{f}$ is not constant if $\alpha, \beta \neq 0$. However, it is completely determined by whether $\omega \in \Omega_1$ or $\omega \in \Omega_2$.

DEF 25.8 *A set $A \in \mathcal{F}$ is* invariant *if*
$$(\{\omega\,:\,T\omega \in A\} =)T^{-1}A = A,$$
up to a null set. It is nontrivial *if $0 < \mathbb{P}[A] < 1$. The set of all invariant sets forms a σ-field $\mathcal{I}$ (see Exercise 7.1.1 in [Dur10]). The transformation T is said* ergodic *if $\mathcal{I}$ is trivial, that is, all invariant sets A have $\mathbb{P}[A] \in \{0, 1\}$.*

2.2 Statement of theorem

We finally state a version of the ergodic theorem without proof. (See [Dur10] for a proof.)

THM 25.9 (Ergodic theorem) *Let $f \in \mathcal{L}^1$ and assume that the measure-preserving map T is ergodic. Then*
$$n^{-1}S_n \to \mathbb{E}[f],$$
a.s and in $\mathcal{L}^1$.

EX 25.10 *Let $X_n(\omega) = \omega_n$ are IID rvs. If A is invariant then $\{\omega : \omega \in A\} = \{\omega : T\omega \in A\} \in \sigma(X_1, \ldots)$ and by induction*

$$A \in \cap_{n \geq 0} \sigma(X_n, \ldots) = \mathcal{T},$$

where $\mathcal{T}$ is the tail σ-field. Thus $\mathcal{I} \subseteq \mathcal{T}$. Since $\mathcal{T}$ is trivial by Kolmogorov's 0-1 law, so is $\mathcal{I}$. Therefore T is ergodic. Applying the ergodic theorem to $f = X_0 \in L^1$ we get

$$n^{-1} \sum_{m=0}^{n-1} X_m(\omega) \to \mathbb{E}[X_0],$$

that is, we recover the SLLN.

2.3 Back to MCs

Going back to Markov chains, recall:

DEF 25.11 *Let*

$$T_i = \inf\{n \geq 1 : X_n = i\},$$

and

$$f_{ij} = \mathbb{P}_i[T_j < +\infty].$$

A chain is irreducible *if $f_{ij} > 0$ for all $i, j \in A$. A state i is recurrent if $f_{ii} = 1$ and is positive recurrent if $\mathbb{E}_i[T_i] < +\infty$.*

THM 25.12 *If X is irreducible and finite, then every state is positive recurrent.*

THM 25.13 *Let X be an irreducible and positive recurrent MC. Then there exists a unique stationary distribution π. In fact,*

$$\pi(i) = \frac{1}{\mathbb{E}_i[T_i]} > 0.$$

EX 25.14 (MCs) *Let $\{X_n\}$ be a MC on S.*

- *ASRW on $[a, b]$: Let $\{S_n\}_{n \geq 0}$ be an asymmetric simple random walk with parameter $1/2 < p < 1$. Let $a < 0 < b$, $N = T_a \wedge T_b$. Then $\{X_n\}_{n \geq 0} = \{S_{N \wedge n}\}_{n \geq 0}$ is a Markov chain. In the ASRW on $[a, b]$, $\pi = \delta_a$ and $\pi = \delta_b$ as well as all mixtures are stationary. The invariant sets are $\{a\}$ and $\{b\}$ and therefore T is not ergodic if π has positive mass on both of them.*

- *On the other hand, assume X is irreducible and positive recurrent with stationary distribution $\pi > 0$. Let $A \in \mathcal{I}$ and note that $\mathbb{1}_A \circ T^n = \mathbb{1}_A$. Then by the Markov property,*

$$\mathbb{E}[\mathbb{1}_A \,|\, \mathcal{F}_n] = \mathbb{E}[\mathbb{1}_A \circ T^n \,|\, \mathcal{F}_n] = h(X_n),$$

where $h(x) = \mathbb{E}_x[\mathbb{1}_A]$. By Lévy's 0-1 law the LHS $\to \mathbb{1}_A$. By irreducibility and recurrence, any $y \in S$ is visited i.o. and we must have $\mathbb{E}_x[\mathbb{1}_A] \equiv h(x) \equiv 0$ or 1. Therefore $\mathbb{P}[A] \in \{0, 1\}$ and $\mathcal{I}$ is trivial. Then applying the Ergodic Theorem to $f(\omega) = g(X_0(\omega))$ where

$$\sum_y |g(y)|\pi(y) < +\infty,$$

we then have

$$n^{-1} \sum_{m=0}^{n-1} g(X_m(\omega)) \to \sum_y \pi(y) g(y).$$

3 Subadditive ergodic theorem

The ergodic theorem can also be extended to certain functionals that are not necessarily additive.

3.1 Subadditivity

Recall:

DEF 25.15 *A sequence $\{\gamma_n\}_{n \geq 0}$ is* subadditive *if for all m, n:*

$$\gamma_{m+n} \leq \gamma_n + \gamma_m.$$

THM 25.16 *If γ is subadditive then*

$$\frac{\gamma_n}{n} \to \inf_m \frac{\gamma_m}{m}.$$

Proof: Clearly

$$\liminf_n \frac{\gamma_n}{n} \geq \inf_m \frac{\gamma_m}{m}.$$

So STS

$$\limsup_n \frac{\gamma_n}{n} \leq \inf_m \frac{\gamma_m}{m}.$$

Fix m and write $n = km + \ell$ with $0 \leq \ell < m$. Applying the subadditivity repeatedly, we have

$$\gamma_n \leq k\gamma_m + \gamma_\ell,$$

so that

$$\frac{\gamma_n}{n} \leq \left(\frac{km}{km+\ell}\right)\frac{\gamma_m}{m} + \frac{\gamma_\ell}{n},$$

and the result follows by taking $n \to +\infty$.

■

EX 25.17 (Longest common subsequence) *Let $\{X_n\}$ and $\{Y_n\}$ be stationary sequences and let $L_{m,n}$ be the longest common subsequence on indices $m < k \leq n$. Clearly*

$$L_{0,m} + L_{m,n} \leq L_{0,n},$$

and $\gamma_n = -\mathbb{E}[L_{0,n}]$ is subadditive.

3.2 Statement of the theorem

The main theorem is the following.

THM 25.18 (Subadditive Ergodic Theorem) *Suppose $\{X_{m,n}\}_{0 \leq m < n}$ satisfy:*

1. $X_{0,m} + X_{m,n} \geq X_{0,n}$.
2. $\{X_{nk,(n+1)k}, n \geq 1\}$ *is a stationary sequence for each k.*
3. *The distribution of $\{X_{m,m+k}, k \geq 1\}$ does not depend on m.*
4. $\mathbb{E}X_{0,1}^+ < \infty$ *and for each n, $\mathbb{E}X_{0,n} \geq \gamma_0 n$ where $\gamma_0 > -\infty$.*

Then

- $\lim \mathbb{E}X_{0,n}/n = \inf_m \mathbb{E}X_{0,m}/m \equiv \gamma$.
- $X = \lim X_{0,n}/n$ *exists a.s. and in L^1 so $\mathbb{E}X = \gamma$.*
- *If all stationary sequences in 2. are ergodic then $X = \gamma$ a.s.*

Proof: See [Dur10]. ■

3.3 Examples

The subadditive ergodic theorem is surprisingly useful.

EX 25.19 (Age-dependent continuous-time branching process) *Start with one individual. Each individual dies independently after time $T \sim F$ and at that point produces $K \sim \{p_k\}_k$ offsprings (both with finite means). Let $X_{0,m}$ be the time of birth of the first individual from generation m and $X_{m,n}$, the time to the birth of the first descendant of that individual in generation n. We check the conditions:*

1. *Clearly*
$$X_{0,m} + X_{m,n} \geq X_{0,n}.$$
2. *$\{X_{nk,(n+1)k}\}_n$ is IID because we are looking at the descendants of a single individual (the first born) over k generations which are not overlapping.*
3. *The distribution of $\{X_{m,m+k}\}_k$ is independent of m for the same reason.*
4. *By nonnegativity and the finite mean of F, condition 4. is satisfied.*

So we can apply the thm. By the IID remark above in 2. we get that the limit is trivial. See [Dur10] for a characterization of the limit.

EX 25.20 (First-passage percolation) *Consider $\mathbb{Z}^d$ as a graph with edges connecting $x, y \in \mathbb{Z}^d$ if $\|x - y\|_1 = 1$. Assign to each edge a nonnegative random variable $\tau(e)$ corresponding to the time it takes to traverse e (in either direction). Define $t(x, y)$ (the passage time) as the minimum time to go from x to y. Let $X_{m,n} = t(mu, nu)$ where $u = (1, 0, \cdots, 0)$. We check the conditions:*

1. *Clearly*
$$X_{0,m} + X_{m,n} \geq X_{0,n}$$
2. *$\{X_{nk,(n+1)k}\}_n$ is stationary by translational symmetry.*
3. *The distribution of $\{X_{m,m+k}\}_k$ is independent of m for the same reason.*
4. *By nonnegativity and the finite mean of τ, condition 4. is satisfied.*

So we can apply the theorem. Enumerating the edges in some order, one can prove (check!) that the limit is tail-measurable and, by the IID assumption, is trivial. See [Dur10] for a characterization of the limit.

References

[Dur10] Rick Durrett. *Probability: theory and examples*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2010.