Ergodicity for dynamical systems

Anna Melnykova

29 May, 2019

Main objective of this lecture is to give an idea of ergodicity for dynamical systems. The idea is the following: first, we present the concepts of a time homogeneous Markov processes and its sets of invariant measures. We will discuss under which criteria the invariant measure exists and is unique. To put it short, to have an existence of an invariant measure, the Markov process should satisfy some compactness property, together with some regularity. In order to have uniqueness of the invariant measure, the Markov process should satisfy some irreducibility property, together with some regularity.

Proposed further reading: Hairer (2008) (from where I have taken most of the definitions and results), Stoyanov (1997) (from where I have taken some examples).

1 Basic definitions

We define a probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

Definition 1 (Markov process)**.** *A stochastic process $\{X_t\}_{t\in T}$ taking values in a state space E is called a* ***Markov process*** *if, for any $N > 0$, any ordered collection $t_{-N} < \cdots < t_0 < \cdots < t_N$ of times, and any two functions $f, g : E^N \to R$, the equality*

$$\mathbb{E}\left[f(X_{t_1}, \ldots, X_{t_N}) g(X_{t-1}, \ldots, X_{t-N}) | X_{t_0}\right] = \mathbb{E}\left[f(X_{t_1}, \ldots, X_{t_N}) | X_{t_0}\right] \mathbb{E}\left[g(X_{t-1}, \ldots, X_{t-N}) | X_{t_0}\right]$$

holds almost surely.

In most situations, Markov processes are constructed from their transition probabilities, that is the specifications that give the probability $\mathcal{P}_{s,t}(x, A)$ that the process is in the set A at time t, given that it was located at the point x at time $s < t$.

Definition 2 (Markov transition kernel)**.** *A* ***Markov transition kernel*** *over a state space E is a map $\mathcal{P} : E \times \mathcal{B}(E) \to \mathbb{R}_+$ such that:*

- *$\forall A \in \mathcal{B}(E)$, the map $x \mapsto \mathcal{P}(x, A)$ is measurable.*
- *$\forall x \in E$, the map $A \mapsto \mathcal{P}(x, A)$ is a probability measure.*

Definition 3 (Markov operator)**.** *A* ***Markov operator*** *over a space E is a bounded linear operator $\mathcal{P} : \mathcal{B}(E) \to \mathcal{B}(E)$ s.t.:*

- *$\mathcal{P}1 = 1$*
- *$\mathcal{P}\varphi$ is positive whenever φ is positive.*
- *If a sequence $\{\varphi_n\} \subset \mathcal{B}(E)$ converges pointwise to an element $\varphi \subset \mathcal{B}(E)$, then $\mathcal{P}\varphi_n$ converges pointwise to $\mathcal{P}\varphi$*

Lemma 1. *Given a measurable space E, there is a one-to-one correspondance between Markov transition kernels over E and Markov operators over E given by $\mathcal{P}(x, A) = (\mathcal{P}1_A)(x)$.*

Proof. Markov transition kernel quite straightforwardly defines a Markov operator — its last property is a consequence of Lebesgue's dominated convergence theorem. Conversely, if we define a candidate for a Markov transition kernel by $\mathcal{P}(x, A) = (\mathcal{P}\mathbb{1}_A)(x)$, we onle need to check that $A \mapsto \mathcal{P}(x, A)$ is a probability measure for every x. The only non-trivial assertion is countable additivity, which follows from the last property of a Markov operator. □

Why do we need this Lemma is to use both terms "Markov operator" and "Markov transition kernel" interchangeably, and to use the symbol $\mathcal{P}$ for both terms interchangeably.

We will make another abuse of notations and define also by the same symbol $\mathcal{P}$ the operator acting on signed measures

$$(\mathcal{P}\mu)(A) = \int_E \mathcal{P}(x, A)\mu(dx)$$

We call a family of Markov operators indexed by time, calling it a Markov semi-group if it satisfies the relation $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s \quad \forall s, t > 0$. We call a Markov process X a *time-homogeneous Markov process with semigroup* $\{\mathcal{P}_t\}$ if $\forall s < t$ we have

$$\mathbb{P}(X_t \in A | X_s) = \mathcal{P}_{t-s}(X_s, A) \quad a.s.$$

Alternative definition for the Markov transition kernel, acting on functions (see, for example, Revuz and Yor (2013)) is the following:

$$(\mathcal{P}_t\varphi)(x) = \mathbb{E}[\varphi(X_t)|X_0 = x]$$

Example 1 (Markov chain)**.** *We have a finite number of states $E = \{1, 2, \dots, n\}$. A Markov transition kernel is given by a $n \times n$ matrix P such that:*

(i) $P_{ij} > 0 \quad \forall(i, j)$

(ii) $\sum_j P_{ij} = 1 \quad \forall i$

The number P_{ij} represents the probability of jumping to the state i, given that the current state is j. The corresponding operator acting on functions is given by the transpose matrix P^T, since $\langle P\mu, \varphi\rangle = \langle \mu, P^T\varphi\rangle$.

Example 2 (Brownian motion)**.** $E = \mathbb{R}$, $\forall\sigma > 0$

$$(\mathcal{P}_t\varphi)(x) = \frac{1}{\sigma\sqrt{2\pi t}}\int_R e^{-\frac{(x-y)^2}{2\sigma^2}}\varphi(y)dy$$
$$\mathcal{P}_t(x, \cdot) = \mathcal{N}(x, \sigma^2 t)$$

Example 3 (Stochastic diffusion)**.** *Let $f : \mathbb{R}^d \to \mathbb{R}^d$, $\quad \sigma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$ be two globally Lipschitz function, and consider the SDE*

$$dX_t = f(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0,$$

where W_t is a standard Wiener process. The Markov semigroup $\mathcal{P}_t$, assosiated to solutions of a SDE is, as in the case of Brownian motion, the solution of a partial differential equation:

$$\partial_t\mathcal{P}_t\varphi = \mathcal{L}\mathcal{P}_t\varphi,$$

where the differential operator $\mathcal{L}$ is given by

$$(\mathcal{L}\varphi)(x) = \sum_i f_i(x)\frac{\partial\varphi}{\partial x_i} + \sum_{i,j,k}\sigma_{ik}(x)\sigma_{jk}(x)\frac{\partial^2\varphi}{\partial x_i\partial x_j}$$

2 Existence and uniqueness of the invariant distribution

Definition 4 (Feller and strong Feller property)**.** *Transition probability $\mathcal{P}$ is called* ***Feller****, if for any* bounded and continuous *function $f : R^d \to R$, $\mathcal{P}f$ is continuous and bounded. $\mathcal{P}$ is* ***strong Feller****, if for any* bounded and measurable *function $f : R^d \to R$, $\mathcal{P}f$ is continuous and bounded. Process X_t is called* ***Feller*** *(or* ***strong Feller****), if its semigroup $\mathcal{P}_t$ is Feller (strong Feller) respectively.*

Example 4 (Markov process which is not Feller)**.** *Let the family $X = (X_t, t \geq s)$ describe the motion for $t \geq s$ of a particle starting at time s from position $X_s = x$: if $X_s < 0$, the particle moves to the left with unit rate, if $X_s > 0$ it moves to the right with unit rate, if $X_s = 0$, the particle moves to the left or to the right with probability $\frac{1}{2}$ for each of these directions. Formally, this can be expressed by:*

$$\mathbb{P}[X_t = x + (t-s), t \geq s] = 1, \text{ if } x > 0$$
$$\mathbb{P}[X_t = x - (t-s), t \geq s] = 1, \text{ if } x < 0$$
$$\mathbb{P}[X_t = t-s, t \geq s] = \mathbb{P}[X_t = -(t-s), t \geq s] = \frac{1}{2} \text{ otherwise.}$$

It is easy to see that $X = (X_t, t \geq s)$ is a Markov family. Further, if g is a continuous and bounded function, we find explicitely that

$$(\mathcal{P}_{t-s}g)(x) = \begin{cases} g(x + (t-s)), & \text{if } x > 0 \\ g(x - (t-s)), & \text{if } x < 0 \\ \frac{1}{2}g(t-s) + \frac{1}{2}g(-t+s), & \text{if } x = 0 \end{cases}$$

Since $(\mathcal{P}_{t-s}g)(x)$ has a discontinuity at $x = 0$, it follows from this that X is not a Feller process even though it is a Markov process.

A probability measure μ on E is invariant for a Markov operator $\mathcal{P}$ if for any bounded $\varphi \in \mathcal{B}(E)$ the equality

$$\int_E (\mathcal{P}\varphi)(x)\mu(dx) = \int_E \varphi(x)\mu(dx). \tag{1}$$

Theorem 1. *1). Let X_t be a Markov process, with a semigroup $\{\mathcal{P}_t\}$. If X_t is compactly supported (E is compact) and Feller, there exist an invariant distribution. In other words, there exists an invariant measure μ for X in the sense* (1) *holds for $\mathcal{P}_t$ for any $t \in T$.*

2). In non-compact case: if $E = \mathbb{R}^d$, and $\forall t \in T, \quad \exists V > 0 : \mathbb{R}^d \to \mathbb{R}$ and

$$\lim_{\|x\| \to \infty} V(x) = \infty, \ \textit{s.t.}$$

$$\mathcal{P} \leq \lambda V + \beta, \ \textit{where } \lambda \in (0,1), \beta \in \mathbb{R}, \tag{2}$$

then there exists an invariant distribution.

Proof. Remarks: the bound (2) is not sharp, the idea is to ensure that the Markov process will always return to the same compact set. Outline of the proof

- Prokhorov theorem (prove that $\mathcal{P}(E)$ is compact set) + Kakutani's fixed point theorem (fixed point is an element of the function's domain that is mapped to itself by the function).

- We build a sequence of measures which "go to an invariant measure $\mu^\star$"

$$\mu^n(f) = \frac{1}{N} \sum_{i=0}^{N-1} \mathbb{E}[f(X_n)]$$

- Then, we prove that μ^∞ is an invariant distribution, that is $\forall f$

$$\mu^\infty(\mathcal{P}_t \varphi) = \mu^\infty(\varphi)$$

  Thanks to Feller property we know that the limit is continuous:

$$\mu^\infty(\mathcal{P}_t \varphi) = \lim_{k \to \infty} \frac{1}{Nk} \sum_{j=0}^{N-1} \mu(\mathcal{P}_t^k \varphi)$$

- Finally, our aim is to prove that

$$\mu^\infty(\varphi) - \mu^\infty(\mathcal{P}_t \varphi) = 0$$

- For the point (2), we first need to prove a tightness of $\{\mu^n\}$ (The intuitive idea is that a given collection of measures does not "escape to infinity"). Basically, what we prove is that the process always returns to some compact. Then, once we got "relative compactness", we apply Prokhorov's theorem and the point (1).

□

Uniqueness: irreducibility + strong Feller property

Intuition of irreducibility for a finite state space: We call a transition matrix P irreducible if it is possible to go from any point to any point of the associated graph. Otherwise, we call it reducible.

Definition 5. *A dynamical system on E is a collection $\{\Theta_t\}_{t\in T}$ of maps $\Theta_t : E \to E$ s.t. $\Theta_t \circ \Theta_s = \Theta_{s+t} \quad \forall s, t, \in T$ such that the map $(t, x) \mapsto \Theta_t(x)$ is jointly measurable. It is called continuous if each of the maps Θ_t is continuous.*

2.1 Ergodic measure for Markov process

Let $\Theta : E \to E$ be a measure-preserving transformation (that is, such that $\mu(A) = \mu(\Theta^{-1}A) \quad \forall A \in \mathcal{B}(E)$). We thus define a σ-algebra of all invariant subsets of I:

$$\mathcal{I} = \{A \subset E : A \text{ Borel and } \Theta_t^{-1}(A) = A \quad \forall t \in T\}$$

Sequence of such transformations $\{\Theta_t\}$ we will call *a dynamical system.*

Definition 6. *An invariant measure μ for a dynamical system $\{\Theta_t\}$ is ergodic, if $\mu(A) \in \{0, 1\} \quad \forall A \in \mathcal{I}$.*

Given a Markov semigroup $\mathcal{P}_t$ over a state space E and an invariant probability measure μ for $\mathcal{P}_t$, we associate to it a probability measure $\mathbb{P}_\mu$ on E^R in the following way. For any bounded measurable $\varphi : E^R \to \mathbb{R}$ such that there exists a $\tilde{\varphi} : E^n \to \mathbb{R}$ and an n-tuple of times $t_1 < \dots < t_n$, we write

$$(\mathbb{P}_\mu)\varphi = \int_E \cdots \int_E \tilde{\varphi}(x_1, \dots, x_n)\mathcal{P}_{t_n - t_{n-1}}(x_{n-1}, dx_n) \dots \mathcal{P}_{t_2 - t_1}(x_1, dx_2)\mu(x_1)$$

Since μ is invariant, the measure $\mathbb{P}_\mu$ is stationary, that is $\Theta_t^{-1}\mathbb{P}_\mu = \mathbb{P}_\mu$ for every $t \in \mathbb{R}$, where the shift map $\Theta_t : E^R \to E^R$ is defined by

$$(\Theta_t x)(s) = x(t + s)$$

Therefore, the measure $\mathbb{P}_\mu$ is an invariant measure for the dynamical system Θ_t over E^R.

Definition 7. *An invariant measure μ for a Markov semigroup $\mathcal{P}_t$ is ergodic, if $\mathbb{P}_\mu$ is ergodic for a shift map Θ_t.*

In essence this implies that the random process will not change its statistical properties with time and that its statistical properties (such as the theoretical mean and variance of the process) can be deduced from a single, sufficiently long sample (realization) of the process.

Theorem 2 (Ergodic theorem). *If μ is ergodic measure and Θ is a measure-preserving transformation, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \varphi(\Theta^k x) = \frac{1}{\mu(E)} \int \varphi d\mu = \mathbb{E}[\varphi(x)]$$

In other words, due to the ergodicity the time average coincides with the space average.

References

Hairer, M. (2008). Ergodic theory for stochastic pdes.

Revuz, D. and Yor, M. (2013). *Continuous martingales and Brownian motion*, volume 293. Springer Science & Business Media.

Stoyanov, J. (1997). *Counterexamples in Probability.* Wiley Series in Probability and Statistics. Wiley.