Ergodicity of Markov processes: theory and computation

Yao Li

Department of Mathematics and Statistics, University of Massachusetts Amherst

September 9, 2021

ICERM, Brown University
Outline

1. Markov processes on measurable state space.
2. Coupling method and renewal theory
3. Exponential and power-law ergodicity
4. Construction of Lyapunov functions
5. Numerical computation of ergodicity
6. Numerical computation of invariant probability measures
Basic setting 1

1. $\Phi_n$ – discrete time Markov process
2. $(X, \mathcal{B}(X))$ – state space with a sigma algebra $\mathcal{B}(X)$
3. $P$ – transition probability. $P(x, A) = \mathbb{P}[\Phi_1 \in A \mid \Phi_0 = x]$.
4. $P(x, \cdot)$ is a probability measure on $(X, \mathcal{B}(X))$, $P(x, A)$ is a measurable function for any $A \in \mathcal{B}$.
5. By Markov property, this is enough to determine a Markov process
Basic setting 2

Markov property: only depends on the nearest history

\[ P[\Phi_{n+1} \in A | \Phi_0, \cdots, \Phi_n] = P[\Phi_{n+1} \in A | \Phi_n] \]

- \( P^m(x, A) = P[\Phi_{n+m} \in A | \Phi_n = x] \).

- \( P^{m+n}(x, A) = \int_X P^n(y, A) P^m(x, dy) \)

- First arrival time: \( \eta_A = \inf_{n \geq 1} \{ \Phi_n \in A \} \)

- Note that \( \eta_A \) is a stopping time (random time that only depends on historical and present states of \( \Phi_n \).)

- Hitting probability: \( L(x, A) = P[\Phi_n \in A \text{ for some } n | \Phi_0 = x] \)
Irreducibility

Main difference from discrete Markov chain: $P(x, y)$ does not make sense any more!

$\Phi_n$ is irreducible if there exists a reference measure $\psi$ on $X$ such that

1. If $\psi(A) > 0$, then $L(x, A) > 0$ for all $x \in X$
2. If $\psi(A) = 0$, then $\psi(\{y : L(y, A) > 0\}) = 0$

$\Phi_n$ can reach everywhere that could be “seen” by $\psi$. 
Example

Stochastic differential equation $X_t$. Euler-Maruyama method.

$$X_{n+1} = X_n + f(X_n)h + \sigma(X_n)\mathcal{N}(0, 1)\sqrt{h}$$

Transition kernel

$$P(x, A) = \int_A \frac{1}{\sqrt{2\pi\sigma(x)^2h}} e^{-(y-x-f(x)h)^2/2\sigma^2(x)h} dy$$

Let Lebesgue measure be the reference measure. Easy to check that $X_n$ is irreducible.
Atom and pseudo-atom

1. Discrete state space: $P(x, y) > 0$. Very useful!
2. Atom: $\alpha$ is an atom if $P(x, \cdot) = P(y, \cdot)$ for all $x, y \in \alpha$. Atom is like a discrete state.
3. Atom usually does not exist
4. Pseudo-atom: small set $C$
5. $C \in \mathcal{B}(X)$ is a small set if there exist an integer $n \in \mathbb{N}$ and a nontrivial measure $\nu$ such that

$$P^n(x, A) \geq \nu(A) \text{ for all } x \in C$$
Euler-Maruyama scheme again

\[ X_{n+1} = X_n + f(X_n)h + \sigma(X_n)\mathcal{N}(0, 1)\sqrt{h} \]

Every bounded set is a small set because the probability density of \( P \) is everywhere strictly positive.

Random walk: \( X_{n+1} = X_n + U_n, \ U_n \sim U(-1/2, 1/2) \).
\([-1/4 , 1/4]\) is a small set with \( n = 1 \) and \( \nu = \text{Lebesgue measure} \).
### Discrete space

Assume irreducibility. Define \( E = \{ n \mid P^n(x, x) > 0 \} \). Period \( d \) is the greatest common divisor of \( E \).

### General space

Assume irreducibility. \( C \) is a small set. Define

\[
E_C = \{ n \mid P^n(x, \cdot) \geq \nu(\cdot), x \in C, \nu(C) > 0 \}
\]

(positive probability that the chain will return to \( C \) after \( n \) steps.)

Period \( d \) is the greatest common divisor of \( E \).

\( \Phi_n \) is aperiodic if \( d = 1 \).
Ergodicity

From now on we assume that $\Phi_n$ is irreducible and aperiodic.

1. **Left operator:** $\mu$ – probability measure. $\mu P^n(A) = P_\mu[\Phi_n \in A]$.
2. **Right operator:** $f$ – observable (function). $P^n f(x) = E_x[f(\Phi_n)]$.
3. **Invariant probability measure.** $\pi$ is said to be invariant if $\pi P = \pi$.

Let $\mu$ and $\nu$ be two probability measures. Does

$$\|\mu P^n - \nu P^n\|_{TV}$$

converge to zero? If yes, how fast??
Main approach: Coupling

A Markov process \((Φ^1_n, Φ^2_n)\) on the state space \(X \times X\) is said to be a Markov coupling if

1. Two marginal distributions are Markov processes \(Φ_n\) with initial distribution \(µ\) and \(ν\), respectively
2. If \(Φ^1_n = Φ^2_n\), then \(Φ^1_m = Φ^2_m\) for all \(m \geq n\).

\[τ_C = \inf_{n \geq 0} \{Φ^1_n = Φ^2_n\}\] is the coupling time.
Coupling Lemma

$$\|\mu P^n - \nu P^n\|_{TV} \leq 2 \mathbb{P}[\tau_C > n].$$

(See whiteboard for the proof.)

Optimal coupling (Pitman 1970s)

There exists a coupling \( (\Phi^1_n, \Phi^2_n) \) (may not be Markov) such that

$$\|\mu P^n - \nu P^n\|_{TV} = 2 \mathbb{P}[\tau_C > n].$$

The existence of “honest” optimal coupling remains open.
Coupling at atom

1. Assume $\Phi_n$ admits an atom $\alpha$.

2. Let $(\Phi_1^n, \Phi_2^n)$ be a coupling such that $\Phi_1^n$ and $\Phi_2^n$ are independent until their first simultaneous visit to $\alpha$, and run together after that.

Easy to check: $(\Phi_1^n, \Phi_2^n)$ is a Markov coupling. Difficulty: property of $\mathbb{P}[\tau_C > n]$?

1. Exponential: $\mathbb{P}[\tau_C > n] \sim \rho^{-n}$ for $\rho > 1$

2. Power-law: $\mathbb{P}[\tau_C > n] \sim n^{-\beta}$ for $\beta > 0$
Renewal process

Let

\[ S_n = \sum_{i=0}^{n} Y_i \]

such that \( Y_1, Y_2, \ldots \) are i.i.d. random nonnegative integers. (\( Y_0 \) could be different). \( S_n \) is a renewal process. \( Y_i \) is called inter-occurrence time.

Let \( u_n = \mathbb{P}[n = S_m \text{ for some } m] \).

If \( S \) is aperiodic, \( u_n \to 1/\mathbb{E}[Y_1] \).
Renewal process from $\Phi_n$

1. $\alpha$ is the atom.
2. $Y_0 = \eta_\alpha$
3. $S_n$ is the $n$-th visit to $\alpha$
4. $S_n$ is a renewal process because $\alpha$ is an atom. $Y_i = \eta_\alpha|_{\Phi_0=\alpha}$. (Markov property: history is independent of the future.)
Simultaneous renewal

1. Now let $S_n$ and $S'_n$ be two renewal processes corresponding to $\Phi_n^1$ and $\Phi_n^2$, respectively.

2. The coupling time $\tau_C$ is the first simultaneous renewal time.

$$\tau_C = \inf \{ n = S_{k_1} = S'_{k_2} \text{ for some } k_1 \text{ and } k_2 \}$$

Three questions

1. What if there is no atom? ✓
2. First simultaneous renewal time? ✓
3. How to estimate the first visit time $\eta_\alpha$ (probably tomorrow)
How to make an atom? (1)

Atom does not exist in most scenarios

Small set is much easier to get

Simplest case. Let $C$ be a small set that satisfies

$$P(x, A) \geq \delta \mathbf{1}_C(x) \nu(A), \quad A \in \mathcal{B}(X), x \in X,$$

where $\nu$ is a probability measure with $\nu(C) = 1$.

Split $X$ into $\hat{X} = X \times \{0, 1\}$ with $X_0 = X \times \{0\}$ and $X_1 = X \times \{1\}$.

Similarly, split $A$ into $A_0$ and $A_1$.
How to make an atom? (2)

1. Let $\lambda$ be a measure on $X$. Split $\lambda$ into $\hat{\lambda}$ on $\hat{X}$ such that

$$\lambda^*(A_0) = \lambda(A \cap C)(1 - \delta) + \lambda(A \cap C^c)$$

$$\lambda^*(A_1) = \lambda(A \cap C)\delta$$

2. In other words, $\lambda^*(A_0 \cup A_1) = \lambda(A)$

3. Split transition kernel $P$ into $\hat{P}$:

$$\hat{P}(x, \cdot) = P(x, \cdot)^* \quad x \in X_0 \setminus C_0$$

$$\hat{P}(x, \cdot) = (1 - \delta)^{-1}[P(x, \cdot)^* - \delta \nu^*(\cdot)] \quad x \in C_0$$

$$\hat{P}(x, \cdot) = \nu^*(\cdot) \quad x \in C_1$$
A Markov process $\Phi_n$ is defined on $\hat{X}$ with transition probability $\hat{P}$.

$C_1$ becomes an atom.

Most result (irreducibility, aperiodicity, recurrence etc.) still holds for $\hat{\Phi}_n$.
First simultaneous renewal time?

1. \( S_n = Y_0 + Y_1 + \cdots + Y_n \), \( S'_n = Y'_0 + Y'_1 + \cdots + Y'_n \)
2. \( Y_0 = \eta_{\alpha \mid \phi_0 \sim \mu} \), \( Y'_0 = \eta_{\alpha \mid \phi_0 \sim \nu} \)
3. \( Y_1, Y'_1, Y_2, Y'_2, \cdots \) are i.i.d. with distribution \( \eta_{\alpha \parallel \phi_0 = \alpha} \)
4. Let \( T \) be the simultaneous renewal time

\[
T = \inf_n \{ n = S_{k_1} = S'_{k_2} \text{ for some } k_1, k_2 \}
\]

5. From renewal theorem: There exist \( n_0 \) and \( c \) such that

\[
P[n \text{ is a renewal time }] = P[n = S_k \text{ for some } k] \geq c
\]

for all \( n \geq n_0 \).
Theorems

Exponential tail
If $E[\rho_1^Y_0], E[\rho_1^Y_0], E[\rho_1^Y_1] < \infty$ for some $\rho_1 > 1$, then there exists $\rho_0 > 1$ such that $E[\rho_0^T] < \infty$.

Power-law tail
If $E[Y_0^\beta], E[(Y_0^\beta)]^\beta, E[Y_1^\beta] < \infty$ for some $\beta > 0$, then $E[T^\beta] < \infty$.

(Note that finite exponential/power-law moment is equivalent to exponential/power-law tail.)
Proof on whiteboard.

Ref: Lectures on the Coupling Method by Torgny Lindvall
Thank you