EE 512 Stochastic Processes

Summary and Review: Part 2

Yibin Xiong

March 2022

Table of Contents

Notations

- $p_{i,j}$: (one-step) transition probability from state i to state j
- $P^n_{i,j}$: the probability of starting from state i and transit to state j after n steps
$$P^n_{i,j} := \mathbb{P}\left(X_{n+m} = j \,|\, X_m = i\right)$$
- $f^n_{i,j}$: the probability that starting in state i, the first transition into state j occurs at time n
$$\begin{aligned} f^0_{i,j} &:= 0 \\ f^n_{i,j} &:= \mathbb{P}\left(X_n = j, X_k \neq j \;\; \forall k = 1, \ldots, n-1 \,|\, X_0 = i\right) \end{aligned}$$
- $f_{i,j} := \sum_{n=0}^{\infty} f^n_{i,j}$
  The probability of ever making a transition to state j given that the process starts from state i
- Hitting time $T_j := \min\{n \in \mathbb{N} : X_n = j\}$: the first time that the system is at state j
- $\mu_{i,j} = \mathbb{E}_i[T_j] := \mathbb{E}[T_j \,|\, X_0 = i]$: the expected time to transit to state j given that the initial state is i

Markov Chains: Concepts

Consider a (discrete-time) stochastic process $\{X_n, n = 0, 1, \dots\}$ that takes on a *finite* or *countable* set of values. If $X_n = i$, the process is said to be in state i at time n.

- **Markov property**: Transition probabilities only depend on current state but not the history

$$\mathbb{P}\left(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0\right) = \mathbb{P}\left(X_{n+1} = j | X_n = i\right)$$

If the given stochastic process satisfies Markov property, then it is a Markov chain.

* We denote the transition probability as $p_{i,j} := \mathbb{P}\left(X_{n+1} = j | X_n = i\right)$. It satisfies $\forall\, i : \sum_{j=0}^{\infty} p_{i,j} = 1$ (must transit into some state).

Markov Chains: Concepts

- Transition matrix

$$P := \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots \\ p_{1,0} & p_{1,1} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

- Distribution
  ◦ At each time, we have a probability distribution $\mathbb{P}(X_t) := \pi^{(t)}$ of which state the random variable is at. It is represented by a row vector

$$\begin{bmatrix} \pi_0^{(t)} & \pi_1^{(t)} & \dots \end{bmatrix}$$

  ◦ We perform a 1-timestep update of the state by multiplying the transition matrix to the vector of current distribution.

- Stationary distribution and Equilibrium
  If the distribution satisfies $\pi P = \pi$, it is the **stationary distribution** and the Markov chain reaches equilibrium. Essentially, π is the eigenvector of P associated with eigenvalue 1.

Stationary Distribution

- Does the stationary distribution always exist? When is it unique?
  – No, see *irreducible, aperiodic* Markov chains later.
  – No, transition matrix can have multiple eigenvectors associated with eigenvalue 1. Note that if π_1 and π_2 are two stationary distributions, then $\forall\, \alpha \in [0, 1] : \alpha\pi_1 + (1 - \alpha)\pi_2$ is also a stationary distribution.

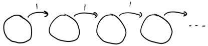
Stationary Distribution

- Does the stationary distribution always exist? When is it unique?
  – No, see *irreducible, aperiodic* Markov chains later.
  – No, transition matrix can have multiple eigenvectors associated with eigenvalue 1. Note that if π_1 and π_2 are two stationary distributions, then $\forall\, \alpha \in [0, 1] : \alpha\pi_1 + (1-\alpha)\pi_2$ is also a stationary distribution.
- How to find the stationary distribution
  – Apply "cuts" to the graph, i.e. the outflow at the cut is the same of the inflow, to get a set of equations that often have some recursive structure. Then use the fact that the total probability mass of a distribution is 1 (normalization condition) to solve for the stationary distribution.

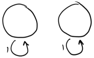
Stationary Distribution

- Does the stationary distribution always exist? When is it unique?
  – No, see *irreducible, aperiodic* Markov chains later.
  – No, transition matrix can have multiple eigenvectors associated with eigenvalue 1. Note that if π_1 and π_2 are two stationary distributions, then $\forall\, \alpha \in [0,1] : \alpha\pi_1 + (1-\alpha)\pi_2$ is also a stationary distribution.
- How to find the stationary distribution
  – Apply "cuts" to the graph, i.e. the outflow at the cut is the same of the inflow, to get a set of equations that often have some recursive structure. Then use the fact that the total probability mass of a distribution is 1 (normalization condition) to solve for the stationary distribution.
- Does a Markov chain necessarily converge to the stationary distribution?
  – No, see an example later

Bad examples:

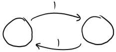
1) 

No stationary distribution: state diverges to ∞
"transient"

2) 

$P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

$(p, 1-p)$ is stationary for any
- the graph is disconnected

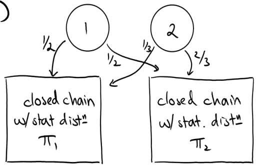
3) 

$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

$(\frac{1}{2}, \frac{1}{2})$ is the unique stationary distribution,
but the system does not converge

It just oscillates with period 2:

$$X_n = \begin{cases} X_0 & \text{, if } n \text{ is even} \\ 1 - X_0 & \text{, if } n \text{ is odd} \end{cases}$$

4) 

- multiple stationary distributions
- state converges to stationarity, but doesn't forget where it started

$X_0 \in$ chain 1 $\Rightarrow (\pi_1, 0)$

$X_0 \in$ chain 2 $\Rightarrow (0, \pi_2)$

$X_0 = 1 \Rightarrow (\frac{1}{2}\pi_1, \frac{1}{2}\pi_2)$

$X_0 = 2 \Rightarrow (\frac{1}{3}\pi_1, \frac{2}{3}\pi_2)$

Irreducibility and Periodicity

- Accessibility: State i is accessible to j if $\exists\, n \in \mathbb{N} : P^n_{i,j} > 0$ (The probability of getting from i to j in n steps in not zero). Denote this as $i \to j$. Additionally, if j is also accessible to i, we say i and j **communicate**, denoted as $i \leftrightarrow j$.

  * Communication is an equivalence relation. It defines equivalent classes.

- Irredicibility: A Markov chain is **irreducible** if $\forall\, i, j : i \leftrightarrow j$.

Irreducibility and Periodicity

- Accessibility: State i is accessible to j if $\exists\, n \in \mathbb{N} : P^n_{i,j} > 0$ (The probability of getting from i to j in n steps in not zero). Denote this as $i \to j$. Additionally, if j is also accessible to i, we say i and j **communicate**, denoted as $i \leftrightarrow j$.

  * Communication is an equivalence relation. It defines equivalent classes.

- Irredicibility: A Markov chain is **irreducible** if $\forall\, i, j : i \leftrightarrow j$.
- Recurrence: state j is **recurrent** if $f_{j,j} = \sum_{n=0}^{\infty} f^n_{j,j} = 1$, i.e. given that the system starts on state j, it will come back with probability 1.

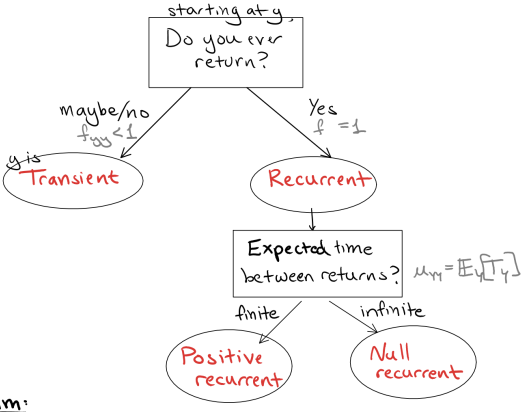
  ◦ If a state is NOT recurrent, then it is **transient**

  ◦ a recurrent state i is **null recurrent** if there exists some j that communicates with i and

$$\lim_{n\to\infty} P^n_{i,j} = 0$$

  Otherwise, it is **positive recurrent**.

Transient and recurrent states



Claim:

- If y is transient, then $\forall x$,

$$\mathbb{P}_x[N_y < \infty] = 1$$

$$\mathbb{E}_x[N_y] = \frac{f_{xy}}{1 - f_{yy}}$$

- If y is recurrent, then

$$\mathbb{P}_y[N_y = \infty] = 1$$

you return to y infinitely many times!

Also, $\mathbb{P}_x[N_y = \infty] = \mathbb{P}_x[T_y < \infty] = f_{xy}$.

Irreducibility and Periodicity

- Periodicity: The **period** of state i is $d(i) := \gcd(\{n \mid P^n_{i,i} > 0\})$.
  * **Aperiodic** means the period is 1.
  * Any self-loop makes the state aperiodic.
  * If $i \leftrightarrow j$, then $d(i) = d(j)$.
  * A Markov chain is called **aperiodic** if all its states have period 1.
  * If it is impossible to get back to state i, the period is defined as ∞.

For Chinese readers, I recommend this post on the concepts we introduced and the limit theorems [1]

[1] https://zhuanlan.zhihu.com/p/389201529

Irreducibility and Periodicity

Irreducible Aperiodic Markov Chain

An irreducible aperiodic Markov chain satisfies either one of the following:

1) states are all transient or all null recurrent, i.e. $\forall j : \mu_{j,j} = \infty$ or equivalently

$$\forall i,j : \lim_{n\to\infty} P^n_{i,j} = \frac{1}{\mu_{j,j}} = 0$$

In this case, there is no stationary distribution.

2) states are all positive recurrent, i.e. $\forall j : \mu_{j,j} < \infty$ or equivalently

$$\forall i,j : \lim_{n\to\infty} P^n_{i,j} = \frac{1}{\mu_{j,j}} > 0$$

In this case, $\left(\frac{1}{\mu_{1,1}}, \frac{1}{\mu_{2,2}}, \dots\right)$ is the unique stationary distribution (**ergodic theorem**).

* If a irreducible, aperiodic Markov chain has *finitely* many states, then it is ergodic and falls into the second case.

Renewal Theories Applied to Markov Chains

Consider a delayed renewal process that starts from state i and the event is defined as arriving at state j.

Elementary Renewal Theorem and Blackwell's Theorem (lattice)

If $i \leftrightarrow j$, then

i) $\mathbb{P}\left(\lim_{n\to\infty} \frac{N_j(n)}{n} = \frac{1}{\mu_{j,j}} \,|\, X_0 = i\right) = 1$

ii) $\lim_{n\to\infty} \frac{\mathbb{E}\left[N_j(n) \,|\, X_0 = i\right]}{n} = \frac{1}{\mu_{j,j}}$

iii) If j is aperiodic, $\lim_{n\to\infty} P^n_{i,j} = \frac{1}{\mu_{j,j}}$

iv) If period$(j) = d$, $\lim_{n\to\infty} P^{nd}_{j,j} = \frac{d}{\mu_{j,j}}$

Reversible Markov Chains

Consider an irreducible Markov chain $\{X_n, n \geq 0\}$ on the finite state space $\mathcal{X}$. We may wonder when it is the case that the Markov chain "looks the same" regardless of whether we run it forwards or backwards in time.

Reversible Markov Chains

Consider an irreducible Markov chain $\{X_n, n \geq 0\}$ on the finite state space $\mathcal{X}$. We may wonder when it is the case that the Markov chain "looks the same" regardless of whether we run it forwards or backwards in time.

Given $N \in \mathbb{N}$, we define the **reversed chain** $Y_n := X_{N-n}$ for $n = 0, 1, \ldots, N$ with transition probability matrix P.

Reversible Markov Chains

Consider an irreducible Markov chain $\{X_n, n \geq 0\}$ on the finite state space $\mathcal{X}$. We may wonder when it is the case that the Markov chain "looks the same" regardless of whether we run it forwards or backwards in time.

Given $N \in \mathbb{N}$, we define the **reversed chain** $Y_n := X_{N-n}$ for $n = 0, 1, \ldots, N$ with transition probability matrix P.

Reversed Chain

If the irreducible Markov chain $\{X_n, n \geq 0\}$ is started from the stationary distribution π, then the reversed chain $\{Y_n, n = 0, \ldots, N\}$ is an irreducible Markov chain with transition probabilities $\hat{P}(x, y) = \frac{\pi(y)P(y,x)}{\pi(x)}$ for all $x, y \in \mathcal{X}$. The stationary distribution for the reversed chain is also π.

Reversible Markov Chains: Detailed Balance

Now we can answer the question when the reversed chain looks the same as the original chain, i.e. $\hat{P}(x, y) = P(x, y) \, \forall \, x, y \in \mathcal{X}$.

Reversible Markov chains

A Markov chain is **reversible** if $\forall \, x, y \in \mathcal{X} : \pi(x)P(x, y) = \pi(y)P(y, x)$. This condition is called **detailed balance**.

Detailed balance condition implies that the current distribution is stationary. In fact, it is stronger than the conditions required for a stationary distribution [2].

[2] https://en.wikipedia.org/wiki/Detailed_balance

Application: Metropolis Algorithm

Suppose we want to sample from a complicated (discrete) probability distribution p. We may not know the probabilities p_j exactly but we know $a_j = c \cdot p_j \; \forall j$. For instance, there are so many states that it is difficult to calculate the normalization factor $A = \sum_{j=1}^{m} a_j$.

We can still sample a sequence of i.i.d. random variable that follows the desired distribution by making it the stationary distribution of a certain Markov chain.

Metropolis Algorithm

Let Q be an irreducible transition matrix on states $1, \ldots, m$ such that $p_{i,j} = p_{j,i} \, \forall \, i, j$. Now define a Markov chain $\{X_n, n \geq 0\}$.
1) If $X_n = i$, then propose a random variable with distribution q_i (the i-th row of Q).
2) If the random variable takes on value j, then we set $X_{n+1} = j$ with probability $\min\{1, \frac{a_j}{a_i}\}$ and set $X_{n+1} = i$ otherwise.

Mixing Time for Finite Chains

We wonder how fast the distribution of a Markov chain converges to the stationary distribution.

Mixing Time

For a Markov chain with transition matrix P and stationary distribution π, the **mixing time** is the smallest t such that

$$\forall x : d_{TV}(\pi, xP^t) < \frac{1}{2e}$$

where the **total variation difference** between two probability distributions is defined as $d_{TV}(\pi_1, \pi_2) := \frac{1}{2} \int_{-\infty}^{\infty} |\pi_1(x) - \pi_2(x)| \, dx$

Mixing Time for Finite Chains

We wonder how fast the distribution of a Markov chain converges to the stationary distribution.

Mixing Time

For a Markov chain with transition matrix P and stationary distribution π, the **mixing time** is the smallest t such that

$$\forall\, x : d_{TV}(\pi, xP^t) < \frac{1}{2e}$$

where the **total variation difference** between two probability distributions is defined as $d_{TV}(\pi_1, \pi_2) := \frac{1}{2} \int_{-\infty}^{\infty} |\pi_1(x) - \pi_2(x)|\ dx$

Example: sample a random graph coloring

Let the maximum degree be Δ. If the number of available colors $q > 4\Delta$, then the mixing time is $\mathcal{O}(n \log n)$, where n is the number of vertices.

Approach: Construct a Markov chain on colorings, with uniform stationary distribution. We run the chain until it converges.

Table of Contents

Definition

If we generalize from discrete time to continuous time, the Markov property becomes

$$\mathbb{P}\left(X(t+s)=j | X(s)=i, X(u)=x(u), 0 \leq u < s\right) = \mathbb{P}\left(X(t+s)=j | X(s)=i\right)$$

Definition

If we generalize from discrete time to continuous time, the Markov property becomes

$$\mathbb{P}\left(X(t+s)=j|X(s)=i,X(u)=x(u),0\leq u<s\right)=\mathbb{P}\left(X(t+s)=j|X(s)=i\right)$$

By this property, we can find that the time that the system stays in state i, denoted as τ_i, follows an exponential distribution since

$$\mathbb{P}\left(\tau_i>s+t|\tau_i>s\right)=\mathbb{P}\left(\tau_i>t\right)$$

Definition

If we generalize from discrete time to continuous time, the Markov property becomes

$$\mathbb{P}\left(X(t+s)=j|X(s)=i, X(u)=x(u), 0\leq u<s\right)=\mathbb{P}\left(X(t+s)=j|X(s)=i\right)$$

By this property, we can find that the time that the system stays in state i, denoted as τ_i, follows an exponential distribution since

$$\mathbb{P}\left(\tau_i>s+t|\tau_i>s\right)=\mathbb{P}\left(\tau_i>t\right)$$

Now we can define a **continuous time Markov chain**. It is a stochastic process having the properties that each time it enters state i:
i) the amount of time it spends in the state before making a transition into a different state follows $\mathrm{Exp}(v_i)$; and
ii) when the process leaves state i, it will enter the next state j with some probability $p_{i,j}$, where $\sum_{j\neq i} p_{i,j}=1$

Transition Rate

$\triangleright$ Idea: We regard the 2 phases of Poisson "clock" and transition as the *combined* process of the independent Poisson processes *split* from the Poisson "clock".

Namely, waiting at state i for time that is exponentially distributed with rate v_i and transit to state j with probability $p_{i,j}$ is EQUIVALENT to in the next δ time, transition to state j with probability $\delta v_i p_{i,j}$ for all $j \neq i$ or stay in state i with probability $1 - \delta v_i$

Time Evolution of Continuous-time Markov Chains

Chapman-Kolmogorov Equations

Let $P(t) = (P_{i,j}(t))$ be the transition matrix where $P_{i,j}(t) := \mathbb{P}\left(X(t) = j \,|\, X(0) = i\right)$ Then

$$P(s+t) = P(s)P(t)$$

Examples: Queues

Examples: Birth-Death Process in Continuous Time

Examples: Tandem Queues

References

- Professor Ben Reichardt's Lecture Notes. USC Spring 2022.
- Ross, Sheldon M., et al. *Stochastic Processes*. Vol. 2. New York: Wiley, 1996.