Notes to AI2613 Stochastic Processes

YBiuR

A long long time ago in a far far away SJTU

Preface

Learning Convex Optimization is non-convex.

Yet learning Stochastic Porcesses is indeed stochastic.

Chapter 1

Discrete Markov Chains

“和女朋友在商场走散了，是在原地等碰到的概率大还是随机走碰到的概率大？急，在线等。”

1.1 Finite Space Markov Chain

1.1.1 Basic Definitions

Definition 1.1.1 (State Space)**.** A **state space** $\mathcal{S}$ is a finite or countable set of states, i.e. the values that random variables X_i may take on.

Definition 1.1.2 (Initial Distribution)**.** The **initial distribution** π_0 is the probability distribution of the Markov Chain at time 0. Denote $\mathbb{P}[X_0 = i]$ by $\pi_0(i)$.

Remark. Formally, π_0 is a function from $\mathcal{S}$ to $[0, 1]$ s.t.

$$\pi_0(i) \geq 0 \text{ for all } i \in \mathcal{S}$$

$$\sum_{i \in \mathcal{S}} \pi_0(i) = 1$$

Definition 1.1.3 (Probability Transition Matrix)**.** The **Transition Matrix** is a matrix $P = (p_{ij})$, where

$$p_{ij} = \mathbb{P}[X_{n+1} = j | X_n = i]$$

i.e. the probability given that the chain is at state i at $T = n$ that jumps to j at $T = n + 1$

Remark.

- The *rows* of P sum up to 1.
- The entries in P are all non-negative.

1.1.2 The Markov Property

Definition 1.1.4 (Markov Property)**.** We say a stochastic process $X_1, X_2, \ldots$ satisfies the **Markov property** if

$$\mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n, \ldots, X_0 = i_0] = \mathbb{P}[X_{n+1} = i_{n+1} | X_n = i_n]$$

That is, the *next* state X_{n+1} depends only on the *current* state X_n.

Definition 1.1.5 (Time-homogeneous Markov Chain)**.** A Markov Chain is said to be **time homogeneous** if

$$\forall t \quad \mathbb{P}[X_{t+1} = j | X_t = i] = P(i, j)$$

For now, we only consider *time homogeneous* Markov Chains.

1.1.3 Matrix Interpretation of Markov Chains

We now computes the probability distribution at $T = n+1$, denoted by π_{n+1}.

$$\pi_{n+1}(j) = \mathbb{P}[X_{n+1} = j] = \sum_{i=1}^{N} \mathbb{P}[X_n = i]\mathbb{P}[X_{n+1} = j|X_n = i] = \sum_{i=i}^{N} \pi_n(i)P(i,j)$$

Therefore

$$\pi_{n+1}^T = \pi_n^T P$$

$$\pi_n^T = \pi_0^T P^n$$

We use $P(i,j)$ to denote the element (i,j) of P, we use P^n to denote the n-th power of P, and we assume that π_i's are column vectors.

Theorem 1.1.1 (Chapman-Kolmogorov Equality)**.**

$$P^{m+n}(i,j) = \sum_k P^m(i,k)P^n(k,j)$$

1.2 Stationary Distribution

Definition 1.2.1 (Stationary Distribution)**.** π is called a **stationary distribution** of a Markov Chain if

$$\pi^T P = \pi$$

Remark. A Markov Chain may have 0, 1 or infinitely many stationary distribution.

Here comes the question: When does a stationary distribution exist? If it exists, is it unique? If it is unique, does the chain converges to it?

1.3 Irreducibility, Aperiodicity and Recurrence

For convenience, we will use $\mathbb{P}_i[A]$ to denote $\mathbb{P}[A|X_0 = i]$, and use $\mathbb{E}_i$ to denote expectation in an analogous way.

1.3.1 Irreducibility

Definition 1.3.1 (Accessibility)**.** Let i, j be two states, we say j is **accessible from** i if it is possible (with positive probability) for the chain to ever visit j if the chain starts from i.

$$\mathbb{P}_i[\bigcup_{n=0}^{\infty}\{X_n = j\}] > 0$$

or equivalently

$$\sum_{n=0}^{\infty} P^n(i,j) = \sum_{n=0}^{\infty} \mathbb{P}_i[X_n = j] > 0$$

Definition 1.3.2 (Communication)**.** We say i **communicates with** j if j is accessible from i and i is accessible from j.

Definition 1.3.3 (Irreducibility)**.** We say a Markov Chain is **irreducible** if all pairs of states communicate. And it is reducible otherwise.

The relation *communicate with* is an equivalent relation, and irreducible simply means the number of equivalent classes is 1.

1.3.2 Aperiodicity

Definition 1.3.4 (Period)**.** Given a Markov Chain, its **period** of state i is defined to be the greatest common divisor d_i of the lengths of loops starting from i.

$$d_i = \gcd\{n|P^n(i,i) > 0\}$$

Theorem 1.3.1. If states i and j communicate, then $d_i = d_j$

Sketch of Proof.

- $P^{n_1}(i,j) > 0$ and $P^{n_2}(j,i) > 0$.
- $P^{n_1+n_2}(i,i) > 0 \Rightarrow d_i|n_1 + n_2$.
- Suppose $P^n j, j > 0$, then $P^{n+n_1+n_2}(i,i) > 0 \Rightarrow d_i|n + n_1 + n_2$.
- $d_i|n \Rightarrow d_j \geq d_i$.
- Similarly $d_i \geq d_j$. We are done.

Remark. Therefore all states in a communicating class have the same period, and all states in an irreducible Markov chain have the same period.

Definition 1.3.5 (Aperiodicity)**.** An irreducible Markov chain is said to be **aperiodic** if its period is 1, and periodic otherwise.

Lemma 1.3.1. If P(i,i) > 0, then the Markov Chain is aperiodic.

Remark. This is a sufficient but not necessary condition.

1.3.3 Recurrence

We define the **First Hitting Time** T_i of the state i by

$$T_i = \inf\{n > 0 | X_n = i\}$$

and we can define recurrence as follows

Definition 1.3.6 (Recurrence)**.** The state i is **recurrent** if $\mathbb{P}_i[T_i < \infty] = 1$, and is transient if it is not recurrent.

Remark. Recurrence means that starting from state i at $T = 0$, the chain *is sure to* return to i eventually.

Theorem 1.3.2. Let i be a recurrent state, and let j be accessible from i, then all of the following hold:

1. $\mathbb{P}_i[T_j < \infty] = 1$.
2. $\mathbb{P}_j[T_i < \infty] = 1$.
3. The state j is recurrent.

Sketch of Proof.

- The paths starting from i can be seen as infinitely many *cycles*.
- Whether the chain visits p in the cycles can be seen as a Bernoulli distribution with probability $p > 0$.
- The probability of not visiting j in the first n cycles is $(i - p)^n$, which goes to 0 as $n \to \infty$. So (1)hold.

- (2) can be proved by contradiction. $\mathbb{P}_j[T_i < \infty] < 1$ will lead to contradiction against the fact that i is recurrent.
- (1)(2) implies (3).

Theorem 1.3.3 (Equivalent Statement of Recurrence)**.** The state i is recurrent if and only if $\mathbb{E}_i[N_i] = \infty$, where $N_i = \sum_{i=0}^{\infty} \mathbb{I}\{X_n = i\}$.

Sketch of Proof.

- Recurrence $\Rightarrow \mathbb{P}_i[N_i = \infty] = 1 \Rightarrow \mathbb{E}_i[N_i] = \infty$.
- The converse is proved by contradiction.
- If i is transient, there is a chance of p that the chain never return to i.
- So N_i is distributed geometrically, and $\mathbb{E}_i[N_i]$ will be finite. Contradiction.

Remark. By taking expectation on N_i, we have:

$$\mathbb{E}_i[N_j] = \sum_{i=0}^{\infty} P^n(i,j)$$

Corollary 1.3.1. If j is transient, then $\lim_{n\to\infty} P^n(i,j) = 0$ for all states i.

Sketch of Proof.

- $\mathbb{E}_j[N_j] < \infty$.
- $\mathbb{E}_i[N_j] = \mathbb{P}_i[T_j < \infty]\mathbb{E}_i[N_j|T_j < \infty]$.
- $\mathbb{E}_i[N_j] \leq \mathbb{E}_i[N_j|T_j] = \mathbb{E}_j[N_j] < \infty$ since the probability *restarts* once the chain visits j again.
- $\mathbb{E}_i[N_j] = \sum_{i=0}^{\infty} P^n(i,j) < \infty$ and this implies our conclusion.

Proposition 1.3.1. Suppose a Markov Chain has a stationary distribution π, if the state j is transient, then $\pi(j) = 0$.

This follows from Corollary 1.3.1.

Corollary 1.3.2. If an irreducible Markov Chain has a stationary distribution, then the chain is recurrent.

Sketch of Proof. The chain cannot be transient, or otherwise all $\pi(j)$ would be 0 and the sum of π does not equal to 1.

Remark. The converse is not true!

Proposition 1.3.2. A drunk man will find his way home, but a drunk bird may get lost forever.

This is because the random walk on $\mathbb{Z}$ and $\mathbb{Z}^2$ is recurrent, while the walks on higher dimensions are transient.

1.3.4 More on Recurrence

Definition 1.3.7 (Null Recurrence)**.** The state i is **null recurrent** if it is recurrent and $\mathbb{E}_i[T_i] = \infty$.

Definition 1.3.8 (Positive Recurrence)**.** The state i is **positive reccurent** it is recurrent and $\mathbb{E}_i[T_i] < \infty$.

Proposition 1.3.3. Given an irreducible Markov Chain, it is either transient, null recurrent or positive recurrent.

1.4 Strong Law of Large Numbers of Markov Chains

Theorem 1.4.1 (SLLN of Markov Chains)**.** Let $X_0, X_1, \dots$ be a Markov Chain starting in the state $X_0 = i$. Suppose state i communicates with state j. The limiting fraction of time that the chain spends in j is $\frac{1}{\mathbb{E}_{\jmath}[T_j]}$.

$$\mathbb{P}_i[\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}\{X_t = j\} = \frac{1}{\mathbb{E}_{\jmath}[T_j]}] = 1$$

Sketch of Proof. Read the book. I'm not going to write this because I am lazy.

1.5 Basic Limit Theorem

aka. Foundamental Theorem of Markov Chains

1.5.1 Finite State Space

Definition 1.5.1 (Spectral Radius)**.** Given a non-negative matrix A, the spectral radius $\rho(A)$ is the maximum norm of its eigenvalues

$$\rho(A) = \max\{\lambda(A)\}$$

Proposition 1.5.1. Let A be a non-negative matrix, then

$$\min_{1\le i\le n} \sum_{j=1}^{N} a_{i,j} \le \rho(A) \le \max_{1\le i\le n} \sum_{j=1}^{N} a_{i,j}$$

Lemma 1.5.1 (Perron-Frobenius Theorem)**.** Let A be a non-negative matrix with spectral radius $\rho(A) = \alpha$, then α is an eigenvalue of A, and has both left and right non-negative eigenvectors.

Remark. Lemma 1.5.1 implies that for a finite probability transition matrix P, it always has at least one stationary distribution, because it always has eigenvalue 1 ($P\mathbf{1} = \mathbf{1}$) and a corresponding eigenvector π s.t. $\pi^T P = \pi$.

Lemma 1.5.2. Suppose a $k \times k$ matrix P is irreducible. Then there exists a unique solution to $\pi P = \pi$.

Theorem 1.5.1 (Foundamental Theorem, Finite Case)**.** If a finite Markov Chain is *irreducible* and *Aperiodic*, then it has a *unique stationary distribution* π any initial distribution *converges* to it:

$$\forall \pi_0, \quad \lim_{t\to\infty} \pi_0^T P^t = \pi^T$$

1.5.2 Countable Case

Theorem 1.5.2 (Basic Limit Theorem)**.** Let $X_1, X_2, \dots$ be an *irreducible aperiodic* Markov Chain that has a *stationary distribution* π. Then

$$\lim_{n\to\infty} \pi_n(i) = \pi(i)$$

for any state i and for any initial distribution π_0.

Lemma 1.5.3 (Bounded Convergence Theorem)**.** If $X_n \to X$ with probability 1 and there is a finite number b such that $|X_n| < b$ for all n, then

$$\mathbb{E}[X_n] \to \mathbb{E}[X]$$

Recall SLLN 1.4.1, and the following is a corollary.

Corollary 1.5.1. For an *irreducible* Markov Chain, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} P^t(i,j) = \frac{1}{\mathbb{E}_j[T_j]}$$

Sketch of Proof. Take expectations on both sides of Theorem 1.4.1 yields the conclusion.

Proposition 1.5.2 (Cesaro Average)**.** If a sequence of numbers a_n converges to a value a, then the **Cesaro Average** $(1/n)\sum_{t=1}^{n} a_t$ also converges to it.

This corollary follows immediately from the proposition,

Corollary 1.5.2. For an *irreducible aperiodic* Markov Chain with a *stationary distribution*,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} P^t(i,j) \to \pi(j)$$

Theorem 1.5.3. An *irreducible, aperiodic* Markov Chain with a *stationary distribution* has a stationary distribution given by

$$\pi(j) = \frac{1}{\mathbb{E}_j[T_j]}$$

Sketch of Proof. The proof of the theorem is trivial, and is done by comparing Corollary 1.5.1 and 1.5.2.

Theorem 1.5.4 (Foundamental Theorem, Countable Case)**.** An *irreducible* Markov Chain has a *unique stationary distribution* given by

$$\pi(j) = \frac{1}{\mathbb{E}_j[T_j]}$$

if and only if it is *positive recurrent.*

1.5.3 Other Conclusions

Since they are not covered in class, no proof will be provided. I'm lazy you know.

Definition 1.5.2 (Doubly Stochastic Chains)**.** A transition matrix P is said to be **doubly stochastic** if all of its columns sum up to 1.

Theorem. For a doubly stochastic Markov Chain with N states, the uniform distribution $\pi(i) = 1/N$ is a stationary distribution.

Theorem. For a Markov Chain with symmetric P and N states, the uniform distribution $\pi(i) = 1/N$ is a stationary distribution.