Lecture 13 : Convergence of Markov Chains

1 Mean time spent in the transient states

Consider a Markov chain X defined on a finite state space E with probability transition matrix P. Let $T \subset E$ be the set of transient states. We define a probability transition matrix Q for transient states as

$$Q_{ij} = P_{ij}, \;\; i, j \in T.$$

All row sums of Q cannot equal 1. At least one row should not sum up to 1, else it contradicts the claim that Q is a transition matrix for the set of transient states. Hence, $I - Q$ is invertible. For $i, j \in T$, we define **fundamental matrix** M such that

$$M_{ij} \triangleq \mathbb{E}_i \sum_{n \in \mathbb{N}_0} 1_{\{X_n = j\}} = \sum_{n \in \mathbb{N}_0} P^n_{ij}.$$

Lemma 1.1. *Fundamental matrix M for transient states of a Markov chain X can be expressed in terms of its transition matrix Q as*

$$M = (I - Q)^{-1}.$$

Proof. We will show that $M = I + QM$. To this end, we re-write M_{ij} as

$$M_{ij} = 1_{\{i=j\}} + \sum_{n \in \mathbb{N}} \sum_{k \in E} \mathbb{P}_i\{X_n = j, X_1 = k\} = I_{ij} + \sum_{k \in E} P_{ik} \sum_{n \in \mathbb{N}_0} P^n_{kj} = I_{ij} + \sum_{k \in T} P_{ik} M_{kj} + \sum_{k \notin T} P_{ik} \sum_{n \in \mathbb{N}_0} P^n_{kj}.$$

Since T is a set of transient states, $P_{ij} = 0$ for $i \notin T$ and $j \in T$, the result follows. □

First time to visit a transient state $j \in T$ is a stopping time

$$\tau_j = \inf\{n \in \mathbb{N}_0 : X_n = j\}.$$

The expected time to visit any transient state $j \in T$, starting from an initial transient state $i \in T$ is

$$f_{ij} \triangleq \mathbb{E}_i \sum_{m \in \mathbb{N}_0} 1_{\{\tau_j = m\}}.$$

Lemma 1.2. *For all $i, \in T$, we have $f_{ij} = M_{ij}/M_{jj}$.*

Proof. From law of total probability, we can write M_{ij} as

$$M_{ij} = \sum_{m \in \mathbb{N}_0} \sum_{n \geq m} \mathbb{P}_i\{X_n = j, \tau_j = m\} = \sum_{m \in \mathbb{N}_0} \mathbb{P}_i\{\tau_j = m\} \sum_{n \in \mathbb{N}_0} \mathbb{P}_j\{X_n = j\} = f_{ij} M_{jj}.$$

□

2 Total variation distance

Given two probability distributions p and q defined on a countable space E, their **total variation distance** is defined as

$$d_{TV}(p,q) \triangleq \frac{1}{2}\|p-q\|_1.$$

Lemma 2.1. *For a countable set E, and distributions $p,q \in \Delta(E)$, we have*

$$d_{TV}(p,q) = \sup\{p(S) - q(S) : S \subseteq E\}.$$

Proof. Let $A = \{i \in E : p(i) - q(i) \geq 0\}$. Then, we can write

$$d_{TV}(p,q) = \frac{1}{2}\left(\sum_{i\in A} p(i) - q(i) + \sum_{i\notin A} q(i) - p(i)\right) = \frac{1}{2}\left(p(A) - p(A^c) - q(A) + q(A^c)\right) = p(A) - q(A).$$

Let $S \subseteq E$, then we have

$$p(S) - q(S) \leq p(S\cap A) - q(S\cap A) \leq p(A) - q(A) = d_{TV}(p,q).$$

Hence, the result follows. □

We say that a sequence of distributions $\nu(n)$ **converges in total variation distance** to a distribution $\pi \in \Delta(E)$, if

$$\lim_{n\in\mathbb{N}} d_{TV}(\nu(n),\pi) = \lim_{n\in\mathbb{N}} \sum_{i\in E} |\nu(n)_i - \pi_i| = 0.$$

Lemma 2.2 (ergodic theorem). *Let $X = \{X_n \in E : n \in \mathbb{N}_0\}$ be a stochastic process with the marginal distribution of X_n denoted by $\nu(n)$ for each $n \in \mathbb{N}$. If $\nu(n) \to \pi$ in total variation distance, then for all bounded functions $f : E \to \mathbb{R}$, we have*

$$\lim_{n\in\mathbb{N}} \mathbb{E}[f(X_n)] = \sum_{i\in E} \pi_i f(i).$$

Proof. Let $\sup_{i\in E}|f(i)| \leq K$ be a finite upper bound on chosen f. From the finiteness of this upper bound, convergence in total variation, and triangular inequality, it follows

$$|\mathbb{E}[f(X_n)] - \sum_{i\in E} \pi_i f(i)| = |\sum_{i\in E} f(i)(\nu(n)_i - \pi_i)| \leq K d_{TV}(\nu(n),\pi).$$

□

3 The coupling method

Two stochastic processes $X \in E^{\mathbb{N}}$ and $Y \in E^{\mathbb{N}}$ defined on the common probability space are said to **coupled**, if there exists an a.s. finite random time τ such that for all $n \geq \tau$, we have $X_n = Y_n$ a.s. Moreover, τ is called a **coupling time** of the joint process (X,Y).

Theorem 3.1 (coupling inequality). *Let τ be a coupling time for the coupled processes X and Y. At each time $n \in \mathbb{N}$, let X_n, Y_n have marginal distributions $p_n, q_n \in \Delta(E)$ respectively. Then,*

$$d_{TV}(p_n,q_n) \leq \Pr\{\tau > n\}.$$

Proof. Consider a finite subset $E_0 \subseteq E$ and $A = \{X_n \in E_0\}, B = \{Y_n \in E_0\}$, and $C = \{\tau \leq n\}$. Then, from definition of coupling time, we have $X_n = Y_n$ a.s. on C. Hence, we can write

$$p_n(E_0) - q_n(E_0) = \Pr(A \setminus C) - \Pr(B \setminus C) \leq \Pr\{X_n \in E_0, \tau > n\} \leq \Pr\{\tau > n\}.$$

□

Variation distance is bounded based on the coupling time. One can bound even the convergence rate by the large deviation of coupling time.

Theorem 3.2 (convergence in total variation of Markov chains). *Let $X = \{X_n \in E : n \in \mathbb{N}_0\}$ be a homogenous ergodic Markov chain with transition probability matrix P and stationary distribution $\pi \in \Delta(E)$. Then, for any initial distribution $\nu(0)$, the distribution $\nu(n)$ at time n converges in total variation to the stationary distribution.*

Proof. Let X and Y be two independent ergodic Markov chains with the transition matrix P and stationary distribution π. We assume the initial distribution of X and Y to be δ_i and π respectively. We construct the product Markov chain $Z_n = (X_n, Y_n)$ for all $n \in \mathbb{N}_0$. Then, $\{Z_n : n \in \mathbb{N}_0\}$ has transition probabilities,

$$\Pr\{Z_n = (k, \ell) | Z_{n-1} = (i, j)\} = P_{ik} P_{j\ell}.$$

It is clear that the chain Z is irreducible, aperiodic, and positive recurrent. Further, we notice that $\pi_Z(i, j) = \pi_i \pi_j$ is a stationary distribution, since

$$\pi_Z(k, \ell) = \pi_k \pi_\ell = \sum_{i \in E} \pi_i P_{i\ell} \sum_{j \in E} \pi_j P_{j\ell} = \sum_{(i,j) \in E \times E} \pi_Z(i, j) P_{ik} P_{j\ell}.$$

Next, we define a stopping time τ for the process Z, as

$$\tau = \inf\{n \in \mathbb{N}_0 : X_n = Y_n\} = \inf\{n \in \mathbb{N}_0 : Z_n \in \{(i, i) : i \in E\}\}.$$

Since Z is an irreducible and recurrent Markov chain, the probability $f_{(k,\ell),(i,i)}$ of reaching diagonal state (i, i) in finite time from any state (k, ℓ) is unity. Hence, for the stopping time τ for ergodic Markov chain Z, we have $\Pr\{\tau < \infty\} = 1$. Consider a process W defined for each $n \in \mathbb{N}_0$ as

$$W_n = X_n 1_{\{n \leq \tau\}} + Y_n 1_{\{n > \tau\}}.$$

Clearly, W is a homogenous Markov chain with transition matrix P and initial state i. That is, it inherits all the statistical properties of chain X. In particular, the distribution of W at any time n is $\nu(n)$. Further, since τ is a coupling time for stationary process Y and its coupled process W, it follows from coupling inequality

$$\frac{1}{2} \sum_{i \in E} |\Pr\{W_n = i\} - \Pr\{Y_n = i\}| = \frac{1}{2} \sum_{i \in E} |\Pr\{X_n = i\} - \pi_i| = d_{TV}(\nu(n), \pi) \leq \Pr\{\tau > n\}.$$

□

Let X and Y be two binomial distributions with parameters (n, p) and (n, q) respectively, for $p > q$. We are interested in finding the relation between $\Pr\{X > k\}$ and $\Pr\{Y > k\}$ for all $k \in E$.

Consider n Bernoulli random variables, $Z_1, Z_2, \ldots, Z_n$ with probability $\Pr\{Z_i = 1\} = p$. Consider random variables $U_1, U_2, \ldots, U_n$ each Bernoulli with probability q/p and independent of random variables $Z_1, Z_2, \ldots, Z_n$, and defining for all $i \in [n]$

$$W_i = U_i Z_i.$$

Hence, we see that $W_i \leq Z_i$ is Bernoulli with parameter $\mathbb{E} W_i = q = \Pr\{W_i = 1\}$. Observing that $Y = \sum_i W_i \leq \sum_i Z_i = X$, it follows that $\Pr\{Y > k\} \leq \Pr\{X > k\}$.