Lecture 12 : Convergence of DTMCs and Coupling theorem

1 Total Variation Distance

Definition 1.1. Given two probability distributions p and q defined on a countable space I, their **total variation distance** is defined as

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

Lemma 1.2. *For a countable set I, and distributions $p, q \in \Delta(I)$, we have*

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

Proof. Let $A = \{i \in I : p(i) - q(i) \geq 0\}$. 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 $S \subseteq I$, 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. □

Definition 1.3 (Convergence in total variation). Let $\{X_n : n \in \mathbb{N}_0\}$ be an I-valued stochastic process with marginal distribution $\pi(n)_i = \Pr\{X_n = i\}$ for all $i \in I$. If there exists a probability distribution $\pi \in \Delta(I)$, such that

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

Then, we say that $\lim_{n\in\mathbb{N}} \pi(n) = \pi$ in total variation distance.

Lemma 1.4. *If $X_n \to \pi$, then for all bounded functions $f : I \to \mathbb{R}$, we have*

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

Proof. Let $\sup_{i\in I} |f(i)| \leq K$ be a finite upper bound on chosen f. Then it follows from convergence in total variation, and observing that

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

□

Theorem 1.5 (Convergence in total variation of DTMC). *Let $X \in I^{\mathbb{N}_0}$ be an ergodic (irreducible, aperiodic, and positive recurrent) DTMC on countable state space I with stationary distribution $\pi \in \Delta(I)$. Then for all initial distributions on X_0, distribution of X_n converges in total variation to π.*

Proof. Let $X_0 = i$, then $\pi(n)_j = \mathbb{P}_i\{X_n = j\} = P^n_{ij}$ for all $j \in I$. We write

$$d_{TV}(\pi(n), \pi) = \frac{1}{2}\sum_{j\in I} |P^n_{ij} - \pi|.$$

This follows from ergodicity of the DTMC. □

2 The Coupling method

Definition 2.1. Consider two stochastic processes $X \in I^{\mathbb{N}}$ and $Y \in I^{\mathbb{N}}$ on state space I. Processes X and Y 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 process.

Theorem 2.2 (Coupling Inequality). *Let τ be a coupling time for coupled processes X and Y with marginal distributions $p_n, q_n \in \Delta(I)$ for X_n, Y_n respectively. Then for all $n \in \mathbb{N}$, we have*

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

Proof. Consider a finite subset $I_0 \subseteq I$ and $A = \{X_n \in I_0\}, B = \{Y_n \in I_0\}$, 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

$$p_n(I_0) - q_n(I_0) = \Pr(A \setminus C) - \Pr(B \setminus C) \leq \Pr\{X_n \in I_0, \tau > n\} \leq \Pr\{\tau > n\}.$$

□

Remark 2.3. Variation distance is bounded based on the coupling time.

Theorem 2.4 (Convergence in total variation of DTMC). *Let $X = \{X_n \in I : n \in \mathbb{N}_0\}$ be a homogenous ergodic DTMC with transition probability matrix P and stationary distribution $\pi \in \Delta(I)$. Then, for any initial distribution on X_0, distribution of X_n converges in total variation to the stationary distribution.*

Proof. We will provide an alternative proof using the coupling argument. Let X and Y be two independent ergodic DTMCs with transition matrix P, stationary distribution π, and initial states i and j respectively. We construct the product DTMC $Z_n = (X_n, Y_n)$ for all $n \in \mathbb{N}_0$. Then, $\{Z_n : n \in \mathbb{N}_0\}$ has transition probabilities,

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

We will first show that DTMC Z is irreducible, aperiodic, and positive recurrent. To this end, we notice that $\pi_z(i, j) = \pi_i\pi_j$ is a stationary distribution, since

$$\pi_z(i, j) = \pi_i\pi_j = \sum_{k\in I} \pi_k P_{ki} \sum_{l\in I} \pi_l P_{lj} = \sum_{(k,l)\in I\times I} \pi_z(k, l)P_{ki}P_{lj}.$$

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

$$\tau = \inf\{n \in \mathbb{N}_0 : X_n = Y_n\}.$$

Since τ is stopping time for ergodic DTMC Z, it follows that $\Pr\{\tau < \infty\} = 1$. Consider a process W defined as

$$W_n = X_n 1_{\{n \le \tau\}} + Y_n 1_{\{n > \tau\}} \;\forall n \in \mathbb{N}_0.$$

It turns out that W is a homogenous DTMC with transition matrix P and initial state i. That is, it inherits all the statistical properties of X. Further, τ is a coupling time for Y and W, and hence by coupling inequality, we have

$$\sum_{m \in I} |\Pr\{W_n = m\} - \Pr\{Y_n = m\}| \le 2\Pr\{\tau < n\}.$$

Since $\Pr\{\tau > n\} \to 0$ and $\Pr\{Y_n = m\} \to \pi_m$ as limit $n \in \mathbb{N}$, we see that

$$\lim_{n \in \mathbb{N}} \Pr\{W_n = i\} = \pi_i.$$

□

Remark 2.5. We can get bounds on the rate of convergence by bounding $\mathbb{P}\{\tau > n\}/$

Example 2.6. 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 $k \in I$.

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

$$W_i = U_i Z_i.$$

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

3 Mean time spent in the transient states

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

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

Remark 3.1. 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.

Definition 3.2. 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 3.3. *Fundamental matrix M for transient states of a DTMC X can be expressed in terms of its transition matrix Q as*

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

Further, we can rewrite M as

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

Since T is a set of transient states, $P_{ij} = 0$ for $i \notin T$ and $j \in T$, we get

$$M = I + QM.$$

Definition 3.4. We define expected time to visit any transient state $j \in T$, starting from initial transient state $i \in T$ as

$$f_{ij} = \mathbb{E}_i 1_{\{X_n = j \text{ for some } n\in\mathbb{N}_0\}}$$

Lemma 3.5. *For all $i, \in T$, we have $f_{ij} = \frac{M_{ij}}{M_{jj}}$.*

Proof. Let $\tau_j = \inf\{n \in \mathbb{N}_0 : X_n = j\}$. Since $j \in T$, we know $\Pr\{\tau_j < \infty\} = 1$, hence we can write

$$f_{ij} = \mathbb{P}_i\{\tau_j < \infty\} = \sum_{m\in\mathbb{N}_0} \mathbb{P}_i\{\tau_j = m\}.$$

Further, we observe

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

□