Lecture-15: Distance from stationarity

1 The convergence theorem

Theorem 1.1 (Convergence Theorem). *Let $X : T \to \mathcal{X}^N$ be an irreducible and aperiodic Markov chain with transition probability matrix P and stationary distribution π. Then there exist constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x \in \mathcal{X}^N} \left\| \pi_x^t - \pi \right\|_{TV} \leqslant C\alpha^t.$$

Proof. Since the Markov chain X is irreducible and aperiodic, there exists a positive integer r such that P^r has strictly positive entries. Let Π be the matrix with $|\mathcal{X}|^N$ rows, each of which is the row vector π. We chose a $\delta > 0$ sufficiently small such that

$$\delta \leqslant \min \left\{ \frac{P^r(x,y)}{\pi(y)} : x, y \in \mathcal{X}^N \right\}.$$

Let $\bar{\delta} = 1 - \delta$, then we can define a matrix $Q = \frac{1}{\bar{\delta}}(P^r - \delta\Pi)$. We can verify that Q is a stochastic matrix by right multiplying it with vector $\mathbf{1}$. We can also verify that for any stochastic matrix M, we have $M\Pi = \Pi$ and if π is an invariant distribution of M, then $\Pi M = \Pi$. We next show by induction that

$$P^{rk} = (\delta\Pi + \bar{\delta}Q)^k = (1 - \bar{\delta}^k)\Pi + \bar{\delta}^k Q^k.$$

The base case of $k = 1$ is true by definition. We assume the inductive hypothesis holds true for $k = n$, then

$$P^{r(n+1)} = P^{rn}P^r = [(1 - \bar{\delta}^n)\Pi + \bar{\delta}^n Q^n]P^r = (1 - \bar{\delta}^n)\Pi + \bar{\delta}^n Q^n((1 - \bar{\delta})\Pi + \bar{\delta}Q).$$

The first equality follows from the fact that $\Pi P^r = \Pi$ and the second equality from the definition of stochastic matrix Q. Since $Q^n\Pi = \Pi$, we have

$$P^{r(n+1)} = (1 - \bar{\delta}^n)\Pi + \bar{\delta}^n Q^n((1 - \bar{\delta})\Pi + \bar{\delta}Q) = (1 - \bar{\delta}^{n+1})\Pi + \bar{\delta}^{n+1}Q^{n+1}.$$

Hence, the result follows from the induction. By post-multiplication with P^j, we get

$$P^{rk+j} - \Pi = \bar{\delta}^k(Q^k P^j - \Pi).$$

We can write the total-variation distance between π_x^t and π for $t = rk + j$

$$\left\| \pi_x^t - \pi \right\|_{\text{TV}} = \left\| P^{rk+j}(x,\cdot) - \pi(\cdot) \right\|_{\text{TV}} \leqslant \bar{\delta}^k \left\| Q^k P^j - \pi \right\|_{\text{TV}} \leqslant \bar{\delta}^k \leqslant C\alpha^t,$$

for $\alpha = \bar{\delta}^{1/r}$ and $C = 1/\bar{\delta}$. □

1.1 Maximal distance from stationarity

Definition 1.2. The maximal distance between t-step distribution π^t and stationary distribution π over all initial configurations $x \in \mathcal{X}^N$ is defined as

$$d(t) \triangleq \max_{x \in \mathcal{X}^N} \left\| P^t(x,\cdot), \pi(x) \right\|_{\text{TV}}.$$

The maximal distance between t-step distributions $P^t(x,\cdot)$ and $P^t(y,\cdot)$ over all initial configurations $x, y \in \mathcal{X}^N$ is defined as

$$\bar{d}(t) \triangleq \max_{x \in \mathcal{X}^N} \left\| P^t(x,\cdot), \pi(x) \right\|_{\text{TV}}.$$

Lemma 1.3. *The following relation between maximal distances is true,*

$$d(t) \leqslant \bar{d}(t) \leqslant 2d(t).$$

Proof. It is immediate from the triangle inequality for the total variation distance that $\bar{d}(t) \leqslant 2d(t)$. To show that $d(t)?\bar{d}(t)$, note first that since π is stationary, we have $\pi(A) =??\sum_{y\in\mathcal{X}^N} \pi(y)P^t(y,A)$ for any set A. Therefore,

$$P^t(x,A) - \pi(A) = \sum_{y\in\mathcal{X}^N} \pi(y)(P^t(x,A) - P^t(y,A)) \leqslant \sum_{y\in\mathcal{X}^N} \pi(y) \left\| P^t(x,\cdot) - P^t(y,\cdot) \right\|_{\mathrm{TV}} \leqslant \bar{d}(t),$$

by the triangle inequality and the definition of total variation. Maximizing the left-hand side over x and A yields $d(t) \leqslant \bar{d}(t)$. □

Exercise 1.4. Show the following.

$$d(t) = \sup_{\mu\in\mathcal{M}(\mathcal{X}^N)} \left\| \mu P^t - \pi \right\|_{\mathrm{TV}}, \qquad \sup_{\mu,\nu\in\mathcal{M}(\mathcal{X}^N)} \left\| \mu P^t - \nu P^t \right\|_{\mathrm{TV}}.$$

Lemma 1.5. *The function $\bar{d}$ is sub-multiplicative. That is, $\bar{d}(s+t) \leqslant \bar{d}(s)\bar{d}(t)$.*

Proof. Fix $x, y \in \mathcal{X}^N$ and let X_s, Y_s denote the configuration of a Markov chain with homogeneous transition probability matrix P starting from initial state x, y respectively. Let (X_s, Y_s) be the optimal coupling of $P^s(x,\cdot)$ and $P^s(y,\cdot)$. Hence

$$\|P^s(x,\cdot) - P^s(y,\cdot)\|_{\mathrm{TV}} = P\left\{X_s \neq Y_s\right\}.$$

We can write

$$P^{s+t}(x,w) = \sum_{z\in\mathcal{X}^N} P\left\{X_s = z\right\} P^t(z,w) = \mathbb{E}P^t(X_s,w).$$

Hence, we can write for a set A,

$$P^{s+t}(x,A) - P^{s+t}(y,A) = \mathbb{E}[P^t(X_s,A) - P^t(Y_s,A)] \leqslant \mathbb{E}[\bar{d}(t)\mathbb{1}_{\{X_s\neq Y_s\}}] = \bar{d}(t)P\left\{X_s \neq Y_s\right\}.$$

□