STAT 31512 Spring 2024 — Scribe: Jiaheng Chen
March 27, 2024 — Lecturer: Frederic Koehler

These notes have not received the scrutiny of publication. They could be missing important references, etc.

Lecture 4 – Intro to Markov Chain

1 Intro to Markov Chain

What is Markov Chain?

Definition 1 (Markov Chain)**.** A discrete-time *Markov chain* is a sequence of random variables $X_1, X_2, X_3, \ldots$, that satisfies for any t, $\Pr[X_{t+1} = \cdot | X_1, \ldots, X_t] = \Pr[X_{t+1} = \cdot | X_t]$.

A related but different object is *martingale*:

Definition 2 (Martingale)**.** A discrete-time *martingale* is a discrete-time stochastic process $X_1, X_2, X_3 \ldots$, that satisfies for any t, $\mathbb{E}[|X_t|] \le \infty$ and $\mathbb{E}[X_{t+1} | X_1, \ldots, X_t] = X_t$.

Definition 3 (Time-Homogeneous Markov Chain)**.** The probability of the transition is independent of t, $\Pr[X_{t+1} = x \mid X_t = y] = \Pr[X_t = x \mid X_{t-1} = y]$ for all t.

Suppose $X_1, X_2, \ldots, X_T \in \Omega$ where Ω is a finite set and $|\Omega| = S$. Let $P \in \mathbb{R}^{S \times S}$ be the transition matrix

$$P_{ij} = \Pr[X_{t+1} = j | X_t = i], \quad 1 \le i, j \le S.$$

Properties of P (row stochastic matrix):

(1) For any i, j, $P_{ij} \ge 0$.

(2) For any i, $\sum_j P_{ij} = 1$.

Example 1. Consider the transition matrix

$$P = \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0 & 0.6 & 0.4 \\ 0 & 0.4 & 0.6 \end{bmatrix}.$$

The Markov Chain is

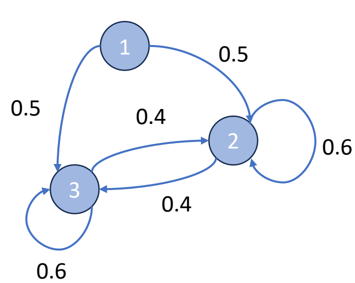

Figure 1: Markov Chain

Remark 1. S is usually very big, e.g. Ising model, $\Omega = \{\pm 1\}^n, |\Omega| = S = 2^n$.

Fact 1. *For any i, $(P\mathbf{1})_i = \sum_j P_{ij} = 1$, so $P\mathbf{1} = \mathbf{1}$.* $\operatorname{spec}(P^\top) = \operatorname{spec}(P)$*, so there exists $\pi \in \mathbb{C}^S$ such that $\pi P = \pi$.*

Fact 2. *All eigenvalues $\lambda \in \operatorname{spec}(P)$ satisfy $|\lambda| \le 1$.*

Proof. Gershgorin disk theorem says that every eigenvalue of $M \in \mathbb{C}^{n\times n}$ lies in

$$\bigcup_{i=1}^{n} \left\{ z : |z - |M_{ii}|| \le \sum_{j\ne i} |M_{ij}| \right\}.$$

This implies for all $\lambda \in \operatorname{spec}(P)$, $|\lambda| \le \sum_j P_{ij} = 1$. □

Definition 4. P is *irreducible* and *aperiodic* if there exists $t \ge 1$ such that P^t has all positive entries.

Theorem 1 (Perron-Frobenius theorem, Fundamental theorem of Markov Chains)**.** *If P is irreducible and aperiodic, then there exists π such that*

(1) $\pi P = \pi$.

(2) $\pi_i > 0$ *for all* $1 \le i \le S$.

(3) $|\lambda_i| < 1$ *for all* $i \ne 1$.

We first do some preparations before proving Theorem 1.

Fact 3. $\Pr(X_t = j | X_0 = i) = (P^t)_{ij}$.

Lemma 1. *Suppose $u, v \in \mathbb{R}^S_{\ge 0}$ are nonnegative vectors and $\sum_i u_i = \sum_i v_i = 1$. Then*

(1) $u^\top P, v^\top P$ *are also probability distributions.*

(2) (strong data processing inequality) $|(u-v)^\top P|_1 \le (1 - 2\min_{i,j} P_{ij})|u-v|_1$.

Proof of Lemma 1. (1) $\sum_j (u^\top P)_j = \sum_j \sum_i u_i P_{ij} = \sum_i u_i = 1$; $\sum_j (v^\top P)_j = \sum_j \sum_i v_i P_{ij} = \sum_i v_i = 1$.

(2) Notice that

$$|(u-v)^\top P|_1 = \sum_j |\sum_i (u_i - v_i) P_{ij}| = \sup_{\sigma \in \{\pm 1\}^n} (u-v)^\top P \sigma.$$

If $\sigma = \mathbf{1}$ or $-\mathbf{1}$, then $(u-v)^\top P\sigma = 0$ following from (1); if σ has at least one positive entry and one negative entry, then for all i,

$$(P\sigma)_i = \sum_j P_{ij}\sigma_j \in [-1 + 2\min_{ij} P_{ij}, 1 - 2\min_{ij} P_{ij}],$$

thus

$$(u-v)^\top P\sigma \le |u-v|_1 |P\sigma|_\infty \le (1 - 2\min_{i,j} P_{ij})|u-v|_1.$$

Therefore, $|(u-v)^\top P|_1 \le (1 - 2\min_{i,j} P_{ij})|u-v|_1$. □

Now we give the proof of Theorem 1.

Proof of Theorem 1. Without loss of generality, we assume $P_{ij} > 0$ for all i, j. Recall that

Theorem 2 (Banach fixed point theorem)**.** *Let $(\mathcal{X}, d)$ be a non-empty complete metric space, and $f : \mathcal{X} \to \mathcal{X}$ satisfies*

$$d(f(x), f(y)) < (1-\varepsilon) d(x,y), \quad \forall\, x, y \in \mathcal{X}$$

for some $\varepsilon > 0$. Then there is a unique x^* such that $f(x^*) = x^*$, and

$$\lim_{t\to\infty} (\underbrace{f \circ \cdots \circ f}_{t \text{ times}})(x) = x^*, \quad \forall\, x \in \mathcal{X}.$$

Consider $\mathcal{X} = \left\{u \in \mathbb{R}^S : u \geq 0, \sum_i u_i = 1\right\}$ and $d(u,v) = |u - v|_1$. By Lemma 1, $P^\top : \mathcal{X} \to \mathcal{X}$ is a contraction mapping, applying Banach fixed point theorem gives that there exists $\pi^* \in \mathcal{X}$ such that $P^\top \pi^* = \pi^*$, and for any $u \in \mathcal{X}$, $\lim_{t\to\infty}(P^\top)^t u = \pi^*$. For every i, $(\pi^*)_i = (P^\top \pi^*)_i = \sum_j P^\top_{ij} \pi^*_j > 0$ since $P^\top_{ij} > 0$ and $\sum_j \pi^*_j = 1$. This proves (1) and (2) in Theorem 1.

For the third part (3), we shall prove all other eigenvalues λ_i satisfy $|\lambda_i| < 1$. Let f be the right eigenvector of P corresponding to eigenvalue λ, $Pf = \lambda f$, thus $P^t f = \lambda^t f$ for $t \geq 1$. For any j, it holds that

$$f^\top \pi^* = \lim_{t\to\infty} f^\top (P^\top)^t e_j = \lim_{t\to\infty} e_j^\top P^t f = \lim_{t\to\infty} \lambda^t e_j^\top f = \lim_{t\to\infty} \lambda^t f_j. \tag{1}$$

If $|\lambda| > 1$ or $|\lambda| = 1$ but $\lambda \neq 1$, $\lim_{t\to\infty} \lambda^t$ does not exist, which is a contradiction. Therefore, $\lambda = 1$ or $|\lambda| < 1$. It remains to show $\mathbf{1}$ is the only right eigenvector of P corresponding to eigenvalue 1. Suppose $Pf = f$, then (1) gives that

$$f^\top \pi^* = f_j, \quad 1 \leq j \leq S,$$

which implies $f_1 = f_2 = \cdots = f_S$ and so f is a multiple of $\mathbf{1}$. This completes the proof. □