

# Markov Chains

Shangtong Zhang

University of Virginia

# Discrete-time Markov chains

A sequence of random variables  $S_0, S_1, S_2, \dots$  taking values in a state space  $\mathcal{S}$  that satisfy the Markov property:

$$\Pr(S_{t+1} = s | S_0 = s_0, S_1 = s_1, \dots, S_t = s_t) = \Pr(S_{t+1} = s | S_t = s_t)$$

Example: random walk on the number line

# Finite Markov chains

The state space  $\mathcal{S}$  is finite.

# Time-homogeneous Markov chains

The transition probability does not depend on time.

For any  $s, s' \in \mathcal{S}$  and any time step  $t, \tau$

$$\Pr(S_{t+1} = s' | S_t = s) = \Pr(S_{\tau+1} = s' | S_{\tau} = s)$$

# Transition matrix

$$P \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{S}|}$$

$$P(s, s') = \Pr(S_{t+1} = s' | S_t = s)$$

When  $\mathcal{S} \doteq \{1, 2, \dots, N\}$ ,

$$P \in \mathbb{R}^{N \times N},$$

$$P(i, j) = \Pr(S_{t+1} = j | S_t = i)$$

## Multi-step transition matrix

$$\begin{aligned} & P^2(i, j) \\ &= \sum_k P(i, k) P(k, j) \\ &= \sum_k \Pr(S_{t+1} = k | S_t = i) \Pr(S_{t+1} = j | S_t = k) \\ &= \sum_k \Pr(S_{t+1} = k | S_t = i) \Pr(S_{t+2} = j | S_{t+1} = k) \\ &= \sum_k \Pr(S_{t+1} = k | S_t = i) \Pr(S_{t+2} = j | S_{t+1} = k, S_t = i) \\ &= \sum_k \Pr(S_{t+2} = j, S_{t+1} = k | S_t = i) \\ &= \Pr(S_{t+2} = j | S_t = i) \end{aligned}$$

# Multi-step transition matrix

$$P^k(i, j) = \Pr(S_{t+k} = j | S_t = i)$$

Chapman-Kolmogorov equations

$$\begin{aligned} P^{n+m}(i, j) &= \sum_k P^n(i, k) P^m(k, j) \\ &\geq P^n(i, k_0) P^m(k_0, j) \end{aligned}$$

# Communicating states

The states  $i, j$  communicate with each other if there exist some  $k_1, k_2$  such that

$$P^{k_1}(i, j) > 0, P^{k_2}(j, i) > 0.$$



# Irreducible Markov chains

For any  $i, j$ , there is a  $k$  such that

$$P^k(i, j) > 0.$$

All states communicate with each other.

# Return time

Let  $\tau_i$  be the return time of  $i$ , i.e.,

$$\tau_i \doteq \inf \{t > 0 \mid S_t = i, S_0 = i\},$$
$$\inf \emptyset \doteq \infty$$

$\tau_i$  is a random variable

# Transient and recurrent states

A state  $i$  is said to be transient iff

$$\Pr(\tau_i < \infty) < 1.$$

A state  $i$  is said to be recurrent iff

$$\Pr(\tau_i < \infty) = 1.$$

A state  $i$  is positive recurrent iff

$$\mathbb{E}[\tau_i] < \infty.$$

A state  $i$  is null recurrent iff

$$\mathbb{E}[\tau_i] = \infty.$$

# Number of visits

Let  $N_i$  denote the number of visits to the state  $i$  starting from  $i$ , i.e,

$$N_i \doteq \sum_{t=0}^{\infty} \mathbb{I}\{S_t = i | S_0 = i\}.$$

$$\begin{aligned}\Pr(N_i = n) &= \Pr(\text{return to } i)^{n-1} \Pr(\text{not return to } i) \\ &= \Pr(\tau_i < \infty)^{n-1} (1 - \Pr(\tau_i < \infty))\end{aligned}$$

$$\mathbb{E}[N_i] \doteq \frac{1}{1 - \Pr(\tau_i < \infty)}$$

## Transient and recurrent states

A state  $i$  is said to be transient iff

$$\mathbb{E}[N_i] < \infty.$$

A state  $i$  is said to be recurrent iff

$$\mathbb{E}[N_i] = \infty.$$

$$\begin{aligned}\mathbb{E}[N_i] &\doteq \sum_{t=0}^{\infty} \mathbb{E}[\mathbb{I}\{S_t = i | S_0 = i\}] \\ &= \sum_{t=0}^{\infty} 1 \times \Pr(S_t = i | S_0 = i) \\ &= \sum_{t=0}^{\infty} P^t(i, i)\end{aligned}$$

# Communication and recurrence

If  $i$  and  $j$  communicate and  $i$  is recurrent, then  $j$  is recurrent.

If

$$P^{k_1}(j, i) > 0 \quad P^{k_2}(i, j) > 0$$

$$\sum_{t=0}^{\infty} P^t(i, i) = \infty$$

Then

$$\begin{aligned} P^t(j, j) &\geq P^{k_1}(j, i) P^{t-k_1-k_2}(i, i) P^{k_2}(i, j), \\ \Rightarrow \sum_t P^t(j, j) &\geq P^{k_1}(j, i) \left( \sum_t P^{t-k_1-k_2}(i, i) \right) P^{k_2}(i, j) \end{aligned}$$

# Communication and recurrence

If  $i$  and  $j$  communicate and  $i$  is recurrent/positive recurrent/transient, then  $j$  is recurrent/positive recurrent/transient.

# Recurrent Markov chains

If all states of a Markov chain are recurrent/positive recurrent/transient, then the chain is said to be recurrent/positive recurrent/transient.



# Stationary distribution

The portion of time that the chain spends in a state  $i$  starting from  $j$ :

$$\pi_i \doteq \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \mathbb{I}\{S_\tau = i | S_0 = j\}$$

$$\mathbb{E}[\pi_i] = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^\tau(j, i)$$

# Stationary distribution

Consider an irreducible Markov chain, if the chain is

- positive recurrent, then  $\pi_i$  exists,  $\pi_i > 0$ ,  $\sum_i \pi_i = 1$ , and  $\pi_i$  does not depend on  $j$ . We call

$$\pi \doteq [\mathbb{E}[\pi_1], \dots, \mathbb{E}[\pi_N]]^\top$$

the stationary distribution.

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^\tau = \begin{bmatrix} \pi^\top \\ \dots \\ \pi^\top \end{bmatrix}$$

- otherwise,  $\pi_i = 0$

# Properties of Stationary distribution

Let  $S_0 = i$ . Let  $t_n$  be the time of the  $n$ -th return to  $i$ .

Let  $Y_n = t_n - t_{n-1}$ . Then

$$\begin{aligned}\pi_i &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \mathbb{I}\{S_\tau = i | S_0 = i\} \\ &= \lim_{n \rightarrow \infty} \frac{1}{\sum_{k=1}^n Y_n} * n \\ &= \frac{1}{\mathbb{E}[Y_n]} \quad \text{w.p. 1} \\ &= \frac{1}{\mathbb{E}[\tau_i]} \quad \text{w.p. 1} \\ \implies \mathbb{E}[\pi_i] &= \frac{1}{\mathbb{E}[\tau_i]}\end{aligned}$$

# Properties of Stationary distribution

$$\begin{aligned}\begin{bmatrix} \pi^\top \\ \dots \\ \pi^\top \end{bmatrix} P &= \left( \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^\tau \right) P \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^{\tau+1} \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \left( \sum_{\tau=1}^{t+1} P^\tau - P \right) \\ &= \lim_{t \rightarrow \infty} \frac{t+1}{t} \frac{1}{t+1} \sum_{\tau=1}^{t+1} P^\tau - \frac{1}{t} P \\ &= \begin{bmatrix} \pi^\top \\ \dots \\ \pi^\top \end{bmatrix}\end{aligned}$$

# Properties of Stationary distribution

$$\pi^\top P = \pi^\top,$$
$$\sum_i \pi_i P(i, j) = \pi_j$$

# A finite irreducible Markov chain is always recurrent

If otherwise, a state  $i$  is transient, then

$$\sum_{t=0}^{\infty} P^t(i, i) < \infty.$$

Let  $\epsilon \doteq \min_{x,y} P^{k_{x,y}}(x, y) > 0$ , then

$$P^t(i, i) = \sum_j P^{t-k_{j,i}}(i, j) P^{k_{j,i}}(j, i) \geq \epsilon \sum_j P^{t-k_{j,i}}(i, j) = \epsilon$$

## A finite irreducible Markov chain is always positive recurrent

If, otherwise, null recurrent,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^{\tau}(j, i) = \frac{1}{\mathbb{E}[\tau_i]} = 0$$
$$\Rightarrow \sum_i \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^{\tau}(j, i) = 0$$

However,

$$\begin{aligned} & \sum_i \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^{\tau}(j, i) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t \sum_i P^{\tau}(j, i) \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} t = 1 \end{aligned}$$

# Periodicity

The period of a state  $i$  is

$$q_i = \gcd\{t \geq 1 \mid P^t(i, i) > 0\}.$$

If  $q_i \geq 2$ , then  $i$  is said to have a period of  $d$ .

If  $q_i = 1$ , then  $i$  is said to be aperiodic.

If  $i$  and  $j$  communicate, then they share the same period.



## Stronger convergence under aperiodicity

If a Markov chain is irreducible and positive recurrent,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=1}^t P^\tau = \begin{bmatrix} \pi^\top \\ \dots \\ \pi^\top \end{bmatrix}.$$

If it is further aperiodic,

$$\lim_{t \rightarrow \infty} P^t = \begin{bmatrix} \pi^\top \\ \dots \\ \pi^\top \end{bmatrix}.$$

# Ergodic chains

A Markov chain is said to be ergodic if it is irreducible, positive recurrent, and aperiodic.

If a finite Markov chain is irreducible and aperiodic, then there exists an integer  $t_0$  such that  $\forall t \geq t_0, i, j$ ,

$$P^t(i, j) > 0$$

# Convergence theorem

If a finite Markov chain is irreducible and aperiodic, then there exist constants  $\alpha \in (0, 1)$  and  $C > 0$  such that  $\forall t$ ,

$$\max_i \|P^t(i, \cdot) - \pi\|_1 \leq C\alpha^t$$

An ergodic chain converges geometrically

# Ergodic theorem

If a finite Markov chain is irreducible and aperiodic, for any  $f : \mathcal{S} \rightarrow \mathbb{R}$ , we have

$$\Pr\left(\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} f(S_\tau) = \mathbb{E}_{s \sim \pi} [f(s)]\right) = 1$$

Time averages equal space averages

# Spectral radius

Let  $X$  be a square matrix, the spectral radius of  $X$  is

$$\rho(X) \doteq \max_i \{|\lambda_i|\}$$

# Stochastic matrix

A stochastic matrix is a nonnegative matrix with each row summing to 1

$$\rho(P) = 1$$

$$\|Px\|_1 \leq \|x\|_1,$$

$$\|Px\|_1 = \|\lambda x\|_1 = |\lambda| \|x\|_1$$

## (Parts of) Perron–Frobenius theorem

If a finite Markov chain is irreducible and aperiodic, then

- Both left and right eigenspaces of  $P$  associated with 1 is 1-dimensional

$$\begin{aligned}\left\{x \mid x^{\top} P = x^{\top}\right\} &= \{\alpha \pi \mid \alpha \in \mathbb{R}\}, \\ \{x \mid P x = x\} &= \{\alpha \mathbf{1} \mid \alpha \in \mathbb{R}\}\end{aligned}$$

- If  $x$  is an eigenvector of  $P$  and  $x_i > 0$ , then the corresponding eigenvalue is 1.

# References

- Lecture notes by Karl Sigman
- Markov Chains and Mixing Times by David Asher Levin, Elizabeth Wilmer, and Yuval Peres
- Wikipedia