Markov Chains

$X_0 \quad X_1 \quad X_2 \quad \ldots\ldots$

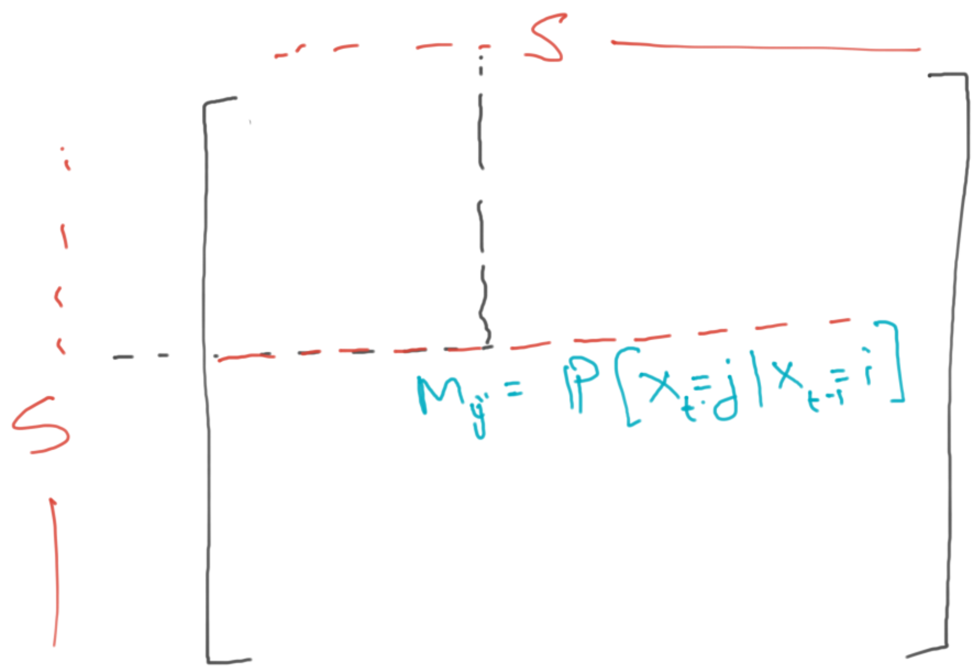
$$P[X_t = a_t \mid X_{t-1} = a_{t-1}, \ldots, X_0 = a_0] = P[X_t = a_t \mid X_{t-1} = a_{t-1}]$$

Random processes with limited history

$$X_0, X_1, \ldots \in S \Big\} \text{ (finite) state space}$$

$|S| = n$

$$\underbrace{\mathbb{P}[X_t = j \mid X_{t-1} = i] = \mathbb{P}[X_{t'} = j \mid X_{t'-1} = i]}_{\text{time homogeneous}}$$



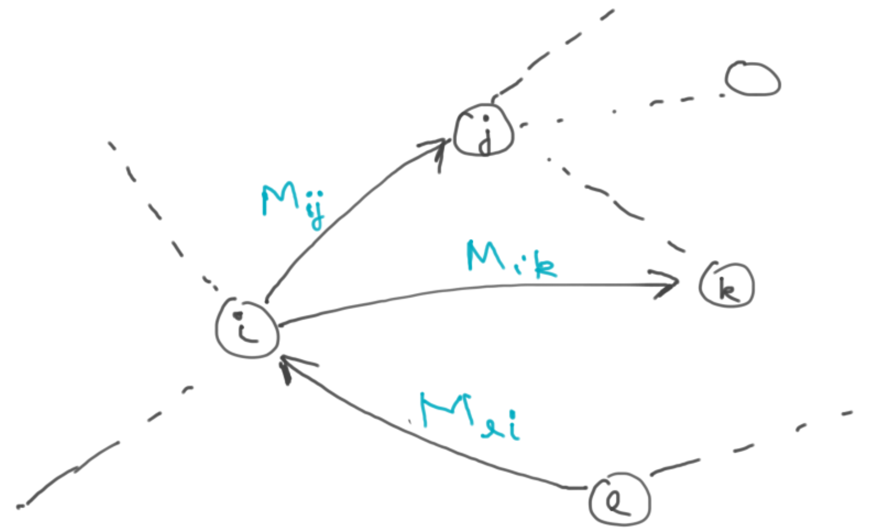
Transition Matrix M

- $M_{ij} \geq 0 \quad \forall i, j$
- $\sum_j M_{ij} = 1 \quad \forall i$

stochastic matrix

Markov Chains as directed graphs

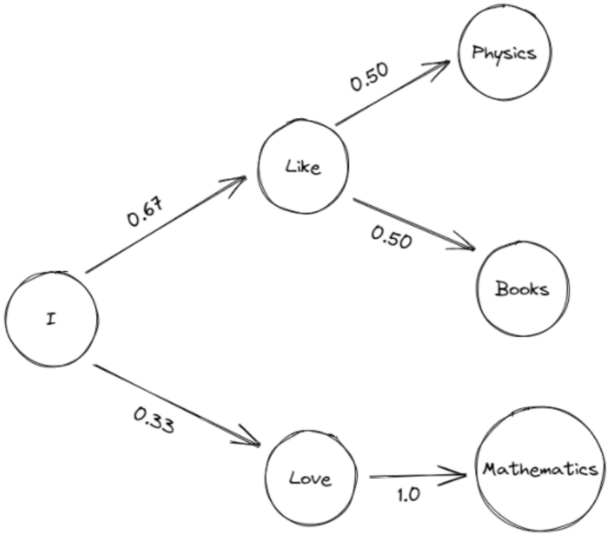
$$M_{ij} = \mathbb{P}[X_t = j \mid X_{t-1} = i] \quad \equiv$$

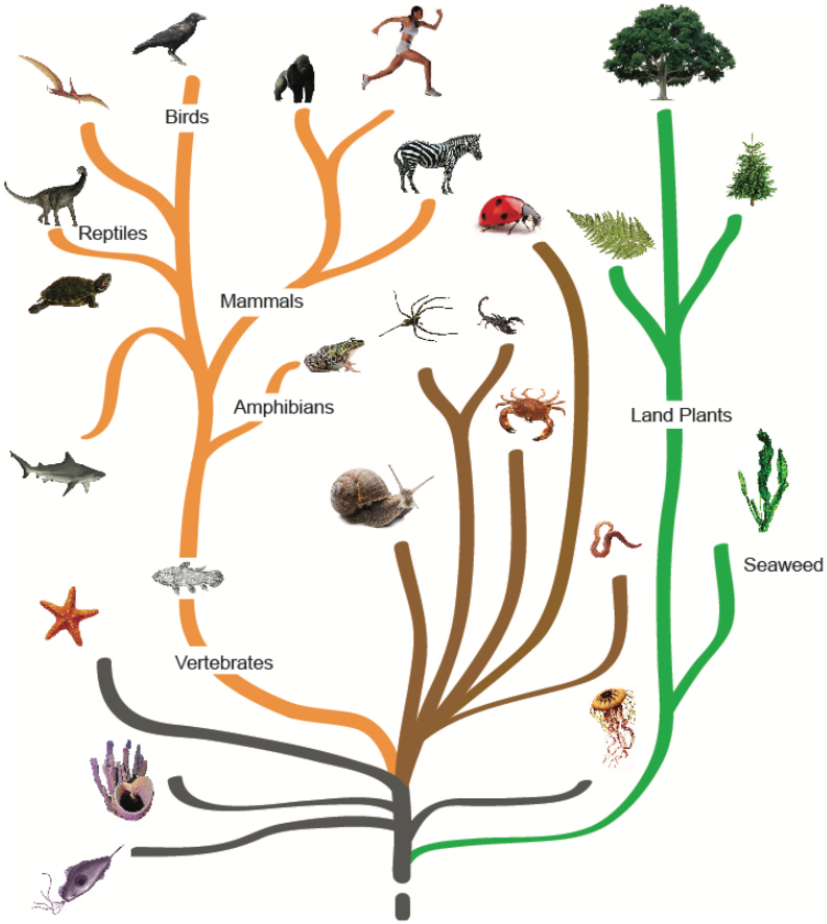


Vertex Set = S

<u>Ex:</u> $\mathbb{P}[X_t = j \mid X_0 = i] = (M^t)_{ij}$

Examples (maybe)





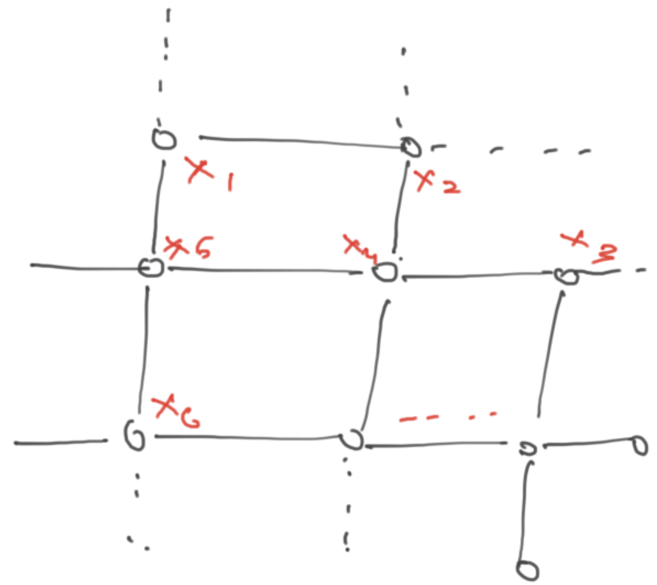
Examples

Card Shuffling

$$x_{t+1} = \sqrt{1-\beta}\cdot x_t + \sqrt{\beta}\, g$$

$$g \sim N(0,1)$$

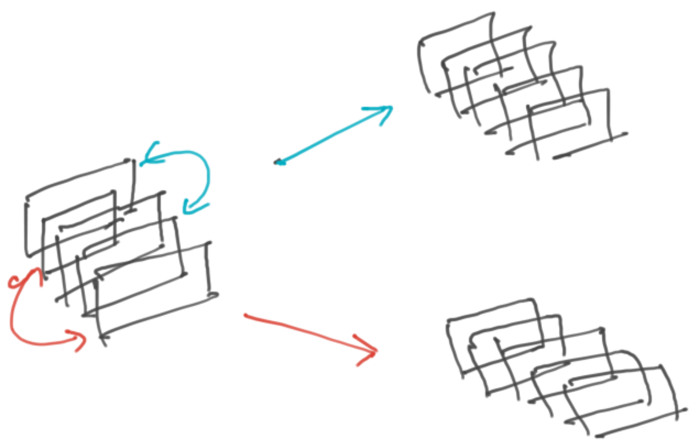
Gaussian noise

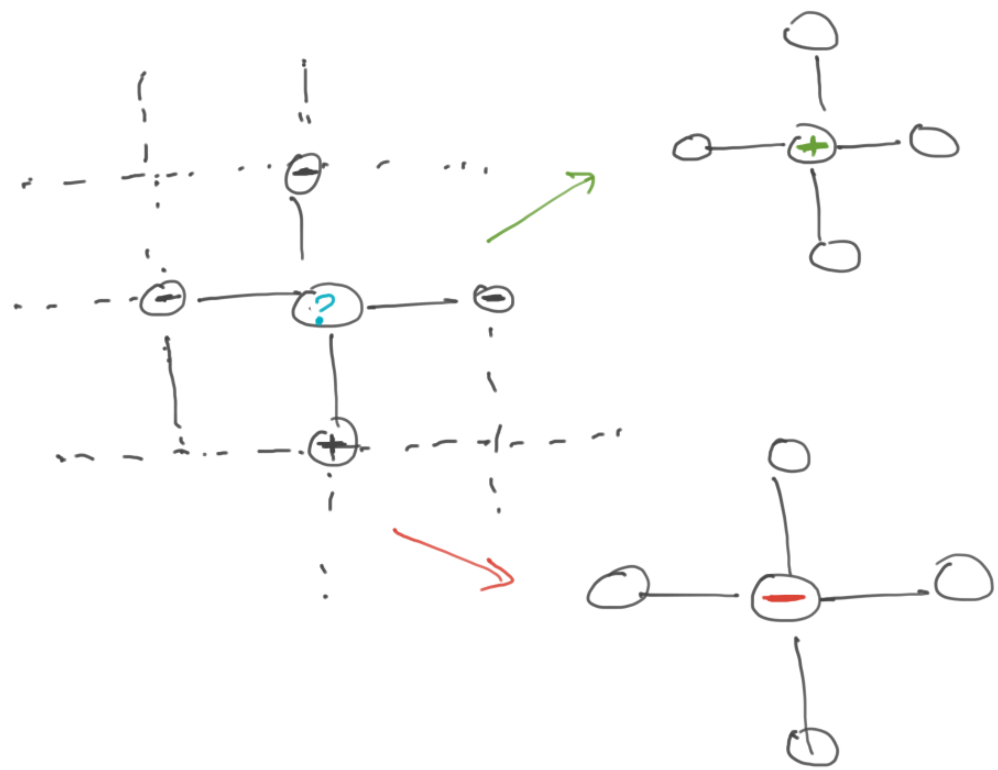


$$\mathbb{P}[x_1 x_2 \cdots] = \frac{e^{-\beta H(x)}}{Z}$$

Ising Model

Studying distributions via "local moves"

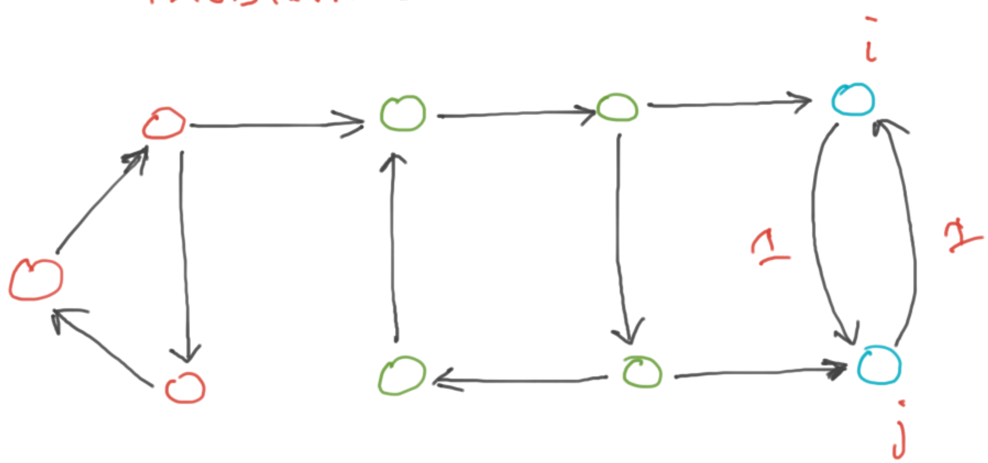




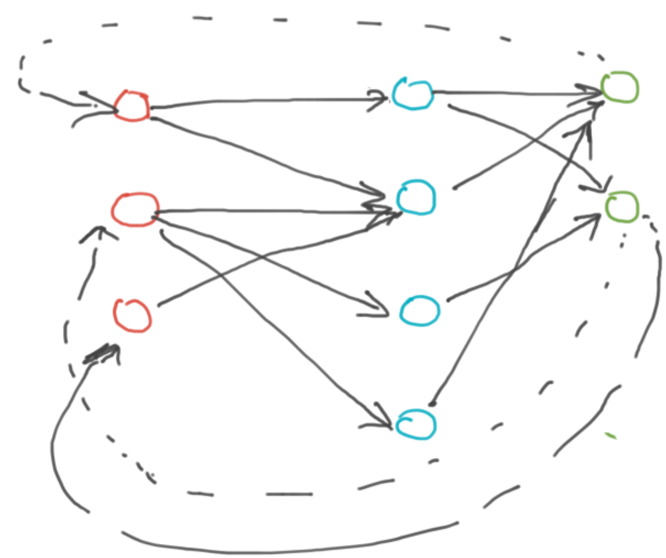
- Often impossible to write full state space.

Behavior of Markov Chains over time

Problem 1



Problem 2



Chain is <u>irreducible</u> if graph is is strongly connected

Chain is <u>aperiodic</u> if $\forall i \; \exists t_0$ s.t. $(M^t)_{ii} > 0$ $\forall t \geq t_0$

Fundamental Thm of Markov Chains

- A finite state irreducible, aperiodic Markov chain admits a unique stationary distribution

$\pi = (\pi_1, \ldots \ldots, \pi_n)$ $(|S| = n)$ s.t.

- $\sum \pi_i = 1$
- $\pi M = \pi$
- $\forall i \quad (M^t)_{ij} \longrightarrow \pi_j$ as $t \to \infty$

A nice special case

Defn

- An ~~irreducible, aperiodic~~ ergodic Markov chain is **reversible** w.r.t. π if $\pi_i M_{ij} = \pi_j M_{ji}$

- If chain is reversible w.r.t. π, then π is stationary distribution.

Speed of Convergence : Linear Algebra

- For $x, y \in \mathbb{R}^n$ $(|S| = n)$ define inner product

$$\langle x, y \rangle_\pi = \sum_{i \in S} x_i y_i \cdot \pi_i = \mathbb{E}_{i \sim \pi}[x_i y_i]$$

- M is reversible $\iff$ M is self-adjoint w.r.t. inner-product $\langle \cdot , \cdot \rangle_\pi$

M is self-adjoint $\Rightarrow$ real eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$

orthonormal $w_1, w_2, \ldots, w_n$

controls convergence

Ex: $|\lambda_i| \leq 1 \ \forall i$ (Gershgorin) and $\lambda_1 = 1$ ($M\mathbb{1} = \mathbb{1}$)

$$(M^t)_{ij} = \big(\underbrace{M^t e_j}_{j^{th}\text{ col}}\big)_i$$

$$(M^t e_j) = \sum_{k=1}^{n} \lambda_k^t \cdot \langle e_j, w_k \rangle_\pi \, w_k = \sum_{k=1}^{n} \lambda_k^t \cdot (w_k)_j \, \pi_j \cdot w_k$$

$$= \pi_j \cdot \mathbb{1} + \sum_{k=2}^{n} \lambda_k^t \cdot (w_k)_j \cdot \pi_j \cdot w_k$$

$$(M^t)_{ij} - \pi_j = \pi_j \cdot \sum_{k=2}^{n} \lambda_k^t \cdot (w_k)_j \cdot (w_k)_i$$

$$\left|(M^t)_{ij} - \pi_j\right| \leq \pi_j \cdot \sum_{k \geq 2} |\lambda_k|^t \cdot |(w_k)_i| \cdot |(w_k)_j|$$

$$\sum_j \left|(M^t)_{ij} - \pi_j\right| \leq \sum_{k \geq 2} |(w_k)_i| \cdot |\lambda_k|^t \cdot \underbrace{\sum_j \pi_j \cdot |(w_k)_j|}_{\langle \mathbb{1}, |w_k| \rangle_\pi \leq \|\mathbb{1}\|_\pi \cdot \|w_k\|_\pi \leq 1}$$

$$\leq \sum_{k \geq 2} \underbrace{|(w_k)_i|}_{\leq \frac{1}{\pi_i} \text{ (why?)}} \cdot |\lambda_k|^t$$

$$\leq \frac{n}{\pi_i} \cdot |\lambda_2|^t$$

$$\therefore \quad \sum_j |(M^t)_{ij} - \pi_j| \leq \frac{n}{\pi_i} \cdot |\lambda_2|^t$$

$$\forall i \quad \sum_j |(M^t)_{ij} - \pi_j| \leq \frac{n}{\pi_{min}} |\lambda_2|^t = \frac{n}{\pi_{min}} \cdot (1-\delta)^t$$

if $|\lambda_2| = 1-\delta$

small after $\underbrace{t \geq \frac{c}{\delta} \cdot \log\left(\frac{n}{\pi_{min}}\right)}_{\text{Mixing time}}$

- Several mathematical techniques to bound $\delta = 1 - |\lambda_2|$
- Other parameters to control mixing times

References

- Diaconis : The Markov Chain Monte Carlo Revolution
- Levin, Peres, Wilmer : Markov Chains & Mixing Times
- Blum, Hopcroft, Kannan : Foundations of Data Science (Ch. 4)