11/17/22

Part 4: Stochastic Processes

Markov Chain

$$X_0 \rightarrow X_1 \rightarrow X_2 \rightarrow \cdots \rightarrow X_t \rightarrow X_{t+1} \rightarrow \cdots$$

notation 1 : $P(X_t)$, $P(X_{t+1}|X_t)$

notation 2 : $p^t(x)$, $K(x,y) = P(X_{t+1}=y \mid X_t = x)$

Markov kernel

notation 3 : $p^t(i)$, K_{ij}

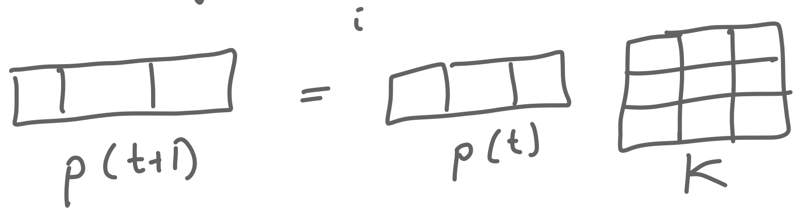
Recursive Computation:

$$P(X_{t+1}) = \sum_{x_t} P(x_t, X_{t+1})$$ marginalization

$$= \sum_{x_t} P(x_t) P(X_{t+1}|X_t)$$ factorization

translate to notation 3 :

$$p^{t+1}(j) = \sum_i p^t(i) K_{ij}$$



$$p^{(t+1)} = p^{(t)} K$$

$$K^{(n)}_{ij} = P(X_{t+n} = j \mid X_t = i)$$

$$K^{(2)}_{ij} = P(X_{t+2}=j \mid X_t = i)$$

$$= \sum_{l} P(X_{t+2}=j \;\&\; X_{t+1}=l \mid X_t=i)$$ marginalization

cross out because only need prior step

$$= \sum_{l} P(X_{t+2}=j \mid X_{t+1}=l, \cancel{X_t=i})\, P(X_{t+1}=l \mid X_t=i)$$

comma = and/intersect

↳ factorization

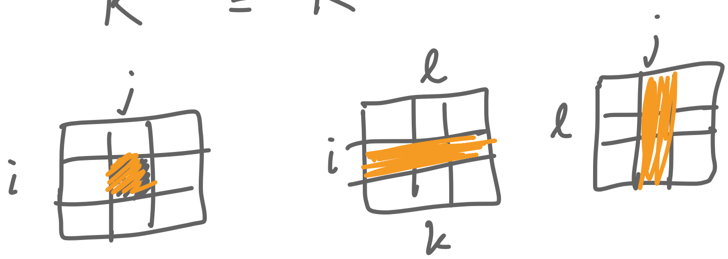
$$= \sum_{l} K_{il} K_{lj}$$

$$\tilde{P}(\cdot) = P(\cdot \mid X_t = i)$$

$$\tilde{p}(X_{t+2}=j) = \sum_{l} \tilde{p}(X_{t+2}=j \;\&\; X_{t+1}=l)$$

$$= \sum_{l} \tilde{p}(X_{t+2}=j \mid X_{t+1}=l)\, \tilde{p}(X_{t+1}=l)$$

$$K^{(2)} = K^{2}$$



$$p^{(t)} = p^{(0)} K^{t}$$

$$K^{(n)} = K^{n}$$ by induction

$$K^{(n_1+n_2)} = K^{(n_1)} K^{(n_2)}$$

perform eigen-analysis

$$K = V \Lambda U^{T} \qquad U^{T} = V^{-1}$$

$$= V \Lambda V^{-1}$$ diagonalization

↳ possible if K has linearly independent e-vectors

$$U^{T} K = \Lambda U^{T}$$

$$\begin{bmatrix} \cdots u_1^T \cdots \\ \cdots u_2^T \cdots \\ \vdots \\ \cdots u_m^T \cdots \end{bmatrix} K = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix} \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_m^T \end{bmatrix} \qquad \lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_m$$

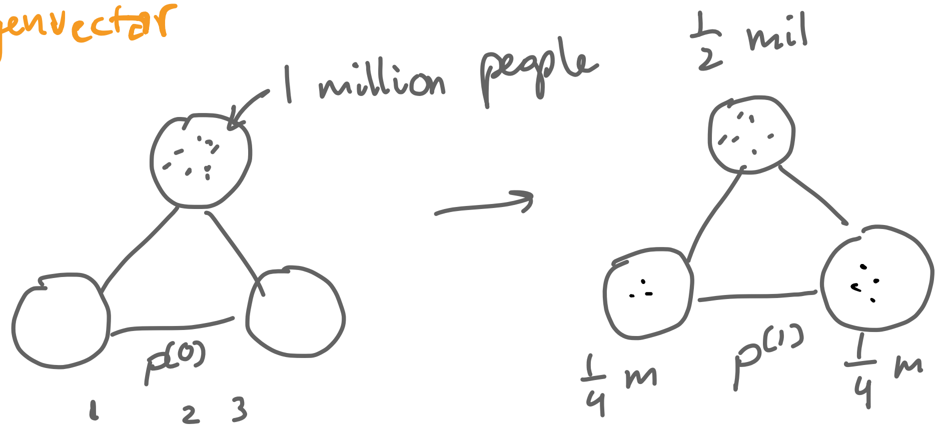
$$u_1^T K = \lambda_1 u_1^T$$

left eigen vector   left eigen value

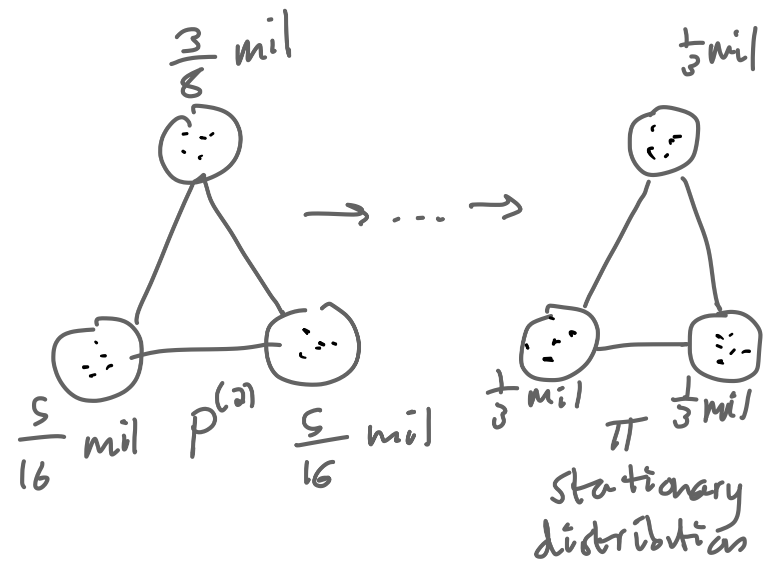
$$KV = V\Lambda$$

$$K \overset{\text{col vectors}}{\begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix}} = \begin{bmatrix} v_1 & v_2 & \cdots & v_m \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix}$$

$$K \underset{\text{right eigenvector}}{\underbrace{v_1}} = \lambda_1 v_1$$



$$K = \begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array}$$



$$p^{(t)} = p^{(0)} K^{t}$$

$$p^{(t+1)} = p^{(t)} K$$

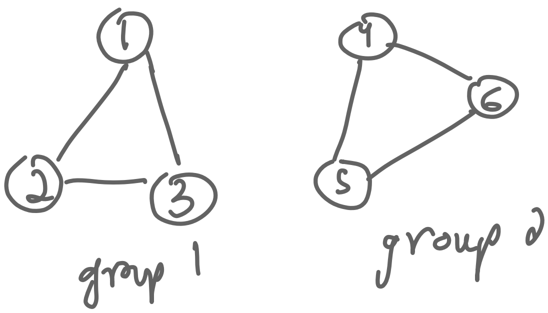
$$\pi = \pi K$$

$$K\mathbb{1} = \mathbb{1}$$

$$K^t = V \Lambda^t U^T \longrightarrow V \begin{bmatrix} 1 & & & \\ & 0 & & \\ & & 0 & \\ & & & 0 \end{bmatrix} U^T = v_1 u_1^T$$

$$= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \pi = \begin{bmatrix} \pi \\ \pi \\ \vdots \\ \pi \end{bmatrix}$$

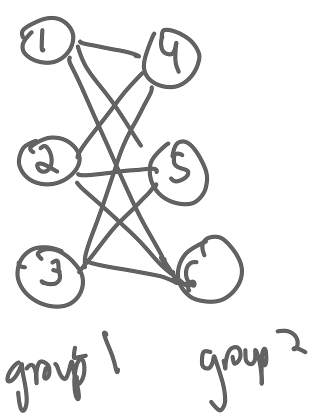
Regardless of distribution, it will converge to pi.



reducible:

$$K = \begin{bmatrix} \boxed{K_{11}} & 0 \\ 0 & \boxed{K_{22}} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}$$

$\lambda_2 = 1$



periodic

$$K = \begin{bmatrix} 0 & K_{12} \\ K_{21} & 0 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = (-1) \begin{bmatrix} -1 \\ -1 \\ -1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$\lambda_n = -1$

rate of convergence : $\max(\lambda_2, |\lambda_n|)$

arrow of time

$$D_{KL}\left(p^{(t)} \,|\, \pi\right)$$

$$p^{(t)} \longrightarrow \pi$$

$$P(x, y) = P(X_t = x,\ X_{t+1} = y)$$
$$= p^{(t)}(x)\, K(x, y)$$

$$\pi(x, y) = \pi(x)\, K(x, y)$$

$$D_{KL}(p \,|\, \pi) = E_{P(x,y)}\left[\log \frac{P(x,y)}{\pi(x,y)}\right]$$
$$= E_{P(x,y)}\left[\log \frac{p^{(t)}(x)}{\pi(x)}\right]$$
$$= E_{p^{(t)}(x)K(x,y)}\left[\log \frac{p^{(t)}(x)}{\pi(x)}\right]$$
$$= E_{p^{(t)}(x)}\left[\log \frac{p^{(t)}(x)}{\pi(x)}\right] = D_{KL}\left(p^{(t)} \,|\, \pi\right)$$

$$\begin{tikzcd} X_t \arrow[r, "1"] & X_{t+1} \arrow[l, dashed, bend left, "2"] \end{tikzcd}$$

$$p(x, y) = p^{(t+1)}(y)\, P(x|y)$$
$$\pi(x, y) = \pi(y)\, \pi(x|y)$$

$$D_{KL}(P \| \pi) = E_p\left[\log \frac{p^{(t+1)}(y)}{\pi(y)}\right] + E_p\left[\log \frac{P(x|y)}{\pi(x|y)}\right]$$

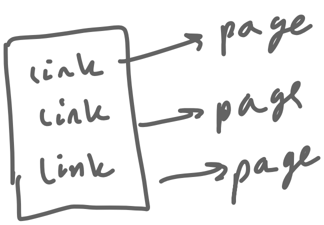
$$= D_{KL}(p^{(t+1)} \| \pi) + D_{KL}(p(x|y) \| \pi(x|y))$$

$$D_{KL}(p^{(t)} \| \pi) \geq D_{KL}(p^{(t+1)} \| \pi)$$

ex.) Google Page Rank

state space = {all webpages}

random web surfers



K_{ij}

$p^{(t)} = \pi$

proportion of population that will land on page i

↳ "popularity" of page

$$p^{(0)} \xrightarrow{K} p^{(1)} \rightarrow \cdots \xrightarrow{K} p^{(T)} \approx \pi$$

$\pi = \pi K$

Fixed point problem

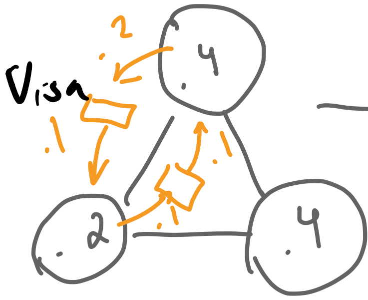
popularity

$$\pi_j = \sum_i \pi_i K_{ij}$$

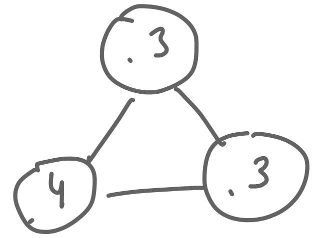
Given K, want π

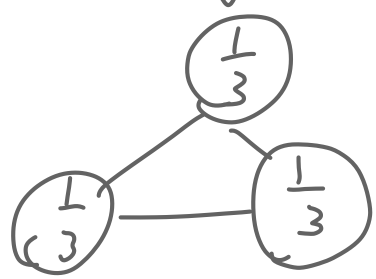
Given π, want to design K.

transition probability



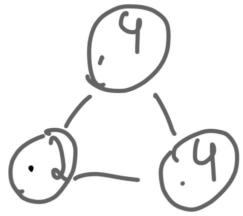
$$\longrightarrow \quad K = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix} \quad \longrightarrow$$





$\downarrow$

$M \longrightarrow$



detailed balance

Metropolis algorithm

\# people $\quad x \xrightarrow{\text{propose}} y : \quad \pi(x) K(x,y)$

$y \xrightarrow{\text{propose}} x : \quad \pi(y) K(y,x)$

After visa : $\min\left(\pi(x) K(x,y), \pi(y) K(y,x)\right)$

Probability of accepting: $x \xrightarrow{\text{propose}} y$: $\min\left(1, \dfrac{\pi(y) K(y,x)}{\pi(x) K(x,y)}\right)$

M, R^2, T^2 (1950s)

$X_t = x$

propose $X_{t+1} = y \sim K(x,y)$

if accept $X_{t+1} = y$

else $X_{t+1} = x$