

Aula 14

Aula passada

- Distribuição estacionária
- Tempo de chegada
- Distância de variação total
- Convergência
- Reversibilidade
- Passeios aleatórios
- Nascimento e morte

Aula de hoje

- Autovalores, autovetores, decomposição
- Convergência para estacionaridade
- Tempo de mistura
- Spectral gap
- Tempo de mistura de passeios aleatórios

Estacionaridade

- Seja P a matriz de transição de estados de uma CM
- π é uma distribuição estacionária sse

$$\pi P = \pi \quad \pi_s \geq 0 \quad \sum_{s \in S} \pi_s = 1$$

- Seja $\pi(0)$ a distribuição inicial da CM. A distribuição no tempo t é dada por

$$\pi(t) = \pi(t-1)P = \pi(0)P^t$$

- Para qualquer CM aperiódica e irredutível, temos

$$\lim_{t \rightarrow \infty} d_{TV}(\pi(t), \pi) = 0$$

- Convergência para π é única e independe de $\pi(0)$

Convergência

- Mas o quão rápido é esta convergência?
- Lembrando da distância de variação total entre dois vetores de probabilidade

$$d_{TV}(\alpha, \beta) = \frac{1}{2} \sum_k |\alpha_k - \beta_k| \quad \longleftarrow d_{TV} \text{ tem valor entre 0 e 1}$$

- Como que $d_{TV}(\pi(t), \pi)$ vai a zero com t ?

$$d_{TV}(\pi(t), \pi) = \theta(e^{-at})? \quad \longleftarrow \text{Muito rápido (exponencial)}$$

$$d_{TV}(\pi(t), \pi) = \theta(t^{-b})? \quad \longleftarrow \text{Menos rápido (lei de potência)}$$

$$d_{TV}(\pi(t), \pi) = \theta((\log t)^{-c})? \quad \longleftarrow \text{Bem menos rápido}$$

- Depende de $\pi(0)$? Depende de P ?

Autovalores e Autovetores

- Dada uma matriz P , v é chamado de autovetor associado ao autovalor λ , se

$$Pv = \lambda v \quad \longleftarrow \text{Multiplicar } v \text{ por } P \text{ é igual a reescalar } v \text{ por } \lambda$$

- P possui até n autovetores linearmente independentes, cada qual associado a um autovalor
- u é chamado de autovetor a esquerda se

$$uP = \lambda u \quad \longleftarrow \text{Multiplicar } u \text{ a esquerda de } P$$

- Se u é autovetor a esquerda, então existe autovetor v tq

$$P'v = \lambda u \quad \longleftarrow P' \text{ é a transposta da matriz } P$$

- Autovalores são os mesmos (esquerda e direita), relação entre autovetores obtida pela transposta da matriz

Autovetores e Matriz P

- π é uma distribuição estacionária da CM com matriz P sse

$$\pi P = \pi$$

- Ou seja, π é o autovetor à esquerda de P associado ao autovalor $\lambda = 1$
- Precisamos ainda $\sum_{s \in S} \pi_s = 1$
- Solução: normalizar o autovetor para garantir soma 1
- **Teorema:** Se P é uma matriz estocástica, temos $|\lambda| \leq 1$ para todo autovalor, e apenas um autovalor $\lambda = 1$
 - matriz estocástica = matriz de transição de probabilidade

Decomposição em Autovetores

- Uma matriz P pode ser escrita através de seus autovetores e autovalores

$$P = Q L Q^{-1}$$

- Onde Q é matriz com autovetores como colunas
- L é matriz diagonal, com L_{ii} autovalor associado ao autovetor i (i -ésima coluna de Q)
- Q^{-1} é a inversa da matriz Q

Exemplo

$$P = \begin{pmatrix} .3 & .2 & .5 \\ .4 & .5 & .1 \\ .7 & .2 & 1 \end{pmatrix}$$

- Autovalores: 1, 0.3, -0.4 (de acordo com nosso teorema)
- Autovetores associados: (1, 1, 1), (2, -5, 2), (-43, 13, 55)

$$P = QLQ^{-1}$$

$$Q = \begin{pmatrix} 1 & 2 & -43 \\ 1 & -5 & 13 \\ 1 & 2 & 55 \end{pmatrix}$$

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & -0.4 \end{pmatrix}$$

$$Q^{-1} = \begin{pmatrix} 43/98 & 2/7 & 27/98 \\ 3/49 & -1/7 & 4/49 \\ -1/98 & 0 & 1/98 \end{pmatrix}$$

Autovetor à esquerda associado ao $\lambda=1$

↑
Autovetores de P
(um em cada coluna)

↑
Autovalores de P
(correspondentes)

↑
Inversa de Q
 Q^{-1} : autovetores à esquerda de P , um em cada linha!

Distribuição no Tempo t

- Seja $\pi(0)$ a distribuição inicial da CM. A distribuição no tempo t é dada por

$$\pi(t) = \pi(t-1)P = \pi(0)P^t$$

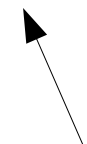
- Mas sabemos que $P = QLQ^{-1}$, então temos

$$\begin{aligned} P^t &= PP \dots P = QLQ^{-1}QLQ^{-1} \dots QLQ^{-1} \\ &= QLILIL \dots LQ^{-1} = QLL \dots LQ^{-1} = QL^t Q^{-1} \end{aligned}$$

- Logo, temos que

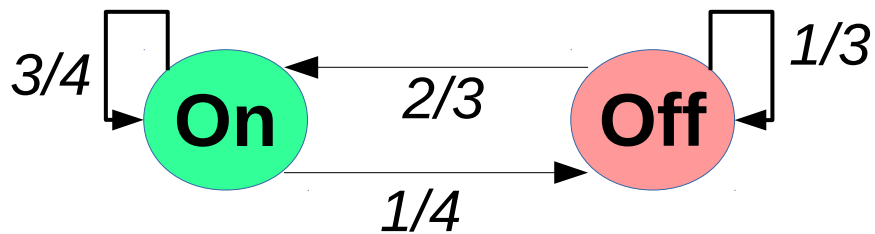
$$\pi(t) = \pi(0)QL^t Q^{-1}$$

Matriz diagonal, com cada elemento elevado a potência t



- Para onde vai L^t com t crescente ?
- $\lambda = 1$ fica no mesmo lugar, todos os outros valores vão a zero, pois $|\lambda| < 1$

Exemplo



$$P = \begin{pmatrix} .75 & .25 \\ .67 & .33 \end{pmatrix}$$

$$\pi = (8/11 \quad 3/11)$$

- Autovalores de P : $\lambda_1 = 1$, $\lambda_2 = 0.08$
- Autovetores a esquerda de P : $v_1 = (67 \ 25)$, $v_2 = (1 \ -1)$

- Temos então $\pi = \left(\frac{67}{67+25} \quad \frac{25}{67+25} \right) \approx (8/11 \quad 3/11)$

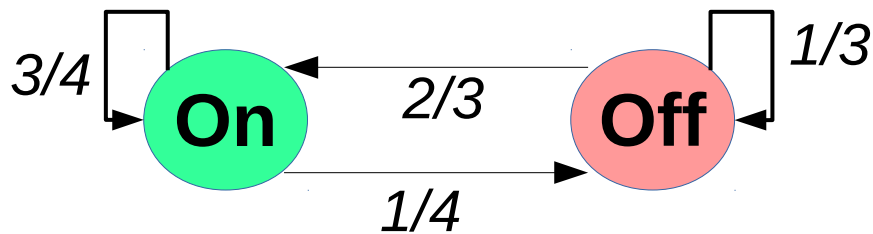
- Supor $\pi(0) = (1 \ 0)$. Podemos escrever

$$\pi(0) = (1 \ 0) = \pi + 3/11 v_2$$

- Temos então

$$\begin{aligned} \pi(t) &= \pi(0) P^t = (\pi + 3/11 v_2) P^t = \pi P^t + 3/11 v_2 P^t \\ &= \pi + 3/11 \lambda_2^t v_2 \\ &= (8/11 + 3/11 (0.08)^t \quad 3/11 - 3/11 (0.08)^t) \end{aligned}$$

Exemplo



$$P = \begin{pmatrix} .75 & .25 \\ .67 & .33 \end{pmatrix}$$

$$\pi = (8/11 \quad 3/11)$$

- Temos então

$$\pi(t) = (8/11 + 3/11(0.08)^t \quad 3/11 - 3/11(0.08)^t)$$

- Distância de variação total

$$d_{TV}(\pi(t), \pi) = \frac{1}{2} \frac{3}{11} (0.08^t + 0.08^t) = \frac{3}{11} 0.08^t = \theta (0.08^t)$$

Converge exponencialmente rápido!

- Resultado vale para qualquer CM
 - converge exponencialmente rápido em t
 - constante a depende da CM e parâmetros

Teorema da Convergência

- Considere uma CM aperiódica e irredutível com matriz de probabilidade P com distribuição estacionária π
- Existe constantes α em $(0, 1)$ e $C > 0$, tq

$$\max_{\pi(0)} d_{TV}(\pi(t), \pi) \leq C \alpha^t$$

- Distribuição transiente $\pi(t)$ converge exponencialmente rápido em t para distribuição estacionária π , independente de $\pi(0)$ ou qualquer outra coisa!

Tempo para Convergência

- Quantos passos até decidir convergência?
- **Ideia:** definir $\varepsilon > 0$ como distância até equilíbrio
 - calcular t tal $d_{TV}(\pi(t), \pi) = \varepsilon$

- Exemplo anterior

$$d_{TV}(\pi(t), \pi) = \frac{3}{11} 0.08^t = \varepsilon \longrightarrow t = \theta \left(\log \frac{1}{\varepsilon} \right)$$

- Muitos poucos passos necessários para chegar próximo do equilíbrio
 - constante em θ depende da CM e parâmetros

Tempo de Mistura

- τ_ϵ : tempo de mistura- ϵ para constante $\epsilon > 0$

$$\tau_\epsilon = \min \{ t \mid \max_{\pi(0)} d_{TV}(\pi(t), \pi) \leq \epsilon \}$$

- τ_ϵ : menor valor de t tal que para qualquer distribuição inicial, $\pi(t)$ está a distância menor que ϵ da estacionária

- Para qualquer CM aperiódica, irreduzível, temos

$$\tau_\epsilon \leq \tau_{1/4} \log 1/\epsilon$$

- Depois de estar perto o suficiente, chegar mais perto é muito fácil (fator $\log \epsilon^{-1}$)

- Tempo de mistura depende fracamente em ϵ

- ϵ vai ser tomado como constante

Spectral Gap

- Convergência depende da relação dos autovalores de P
 - segundo maior autovalor (em módulo) domina convergência
- *Spectral Gap* (δ): distância entre os dois maiores autovalores de P (maior é sempre igual a 1)

$$\delta = 1 - \max_{k > 1} \{ |\lambda_k| \} \leftarrow k\text{-ésimo autovalor de } P, \lambda_1 = 1$$

- Quanto maior for δ , mais rápido é convergência
 - base da exponencial que domina a convergência é dada por $|\lambda_2|$ (segundo maior autovalor)
 - todas as outras componentes vão a zero mais rapidamente (menor base)

Spectral Gap e Tempo de Mistura

- Relação entre δ e τ_ϵ
- Considere CM irredutível aperiódica com *spectral gap* δ e $\pi_0 = \min_i \pi_i$ (menor valor da distribuição estacionária)
- Temos a seguinte relação

$$\frac{\log 1/(2\epsilon)}{2\delta} \leq \tau_\epsilon \leq \frac{\log 1/(\pi_0 \epsilon)}{\delta}$$

↑
Limitante inferior para tempo de mistura

↑
Limitante superior para tempo de mistura

- Maior δ , menor τ_ϵ
- Maior π_0 , menor τ_ϵ

- Usar limitante superior na prática não é fácil, pois precisamos de π_0 e δ

Tempo de Mistura em N

- Resultados anteriores trazem boas notícias
 - convergência exponencial, tempo de mistura relativamente pequeno
- Mas espaço de estado da CM pode crescer com o tamanho do problema
 - convergência e tempo de mistura nestes casos?
- Seja $N = 2^n$ o número de estados da CM
 - N cresce exponencialmente em n
 - ex. estado da CM = permutações de n números
- Como τ depende de N e n ?
 - para algum ε fixo

Passeios Aleatórios

- Tempo de mistura de passeios aleatórios em grafos que podem crescer
 - modelo do grafo parametrizado por n (vértices)
 - grafo completo, grafo em anel, hipercubo, etc
- Passeio aleatório preguiçoso (*lazy random walk*)
 - permanece no vértice atual com prob $\frac{1}{2}$, caso contrário, escolhe vizinho uniformemente
 - implica CM aperiódica e irredutível
- Como que o tempo de mistura depende da estrutura do grafo?
 - τ_n é o tempo de mistura com n vértice para um ε constante

Mistura de Diferentes Grafos

- Anel (um ciclo) com n vértices

$$c n^2 \leq \tau_n \leq n^2$$

- Árvore binária cheia com n vértices

$$\tau_n \leq 16 n$$

- Grafo completo com n vértices (com n grande)

$$\tau_n = 1$$

- Hipercubo com d dimensões e $n=2^d$ vértices

$$\tau_n \leq c d \log d \quad \leftarrow \text{chamado de "fast mixing"}$$

**Tema atual de pesquisa na
matemática (e computação)**