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

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

- 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

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

- 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

$$d_{TV}(\alpha,\beta)=\frac{1}{2}\sum_k |\alpha_k-\beta_k|$$

$\longleftarrow$ d_{TV} tem valor entre 0 e 1

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

$$d_{TV}(\pi(t),\pi)=\theta(e^{-at})?$$

$\longleftarrow$ Muito rápido (exponencial)

$$d_{TV}(\pi(t),\pi)=\theta(t^{-b})?$$

$\longleftarrow$ Menos rápido (lei de potência)

$$d_{TV}(\pi(t),\pi)=\theta\left((\log t)^{-c}\right)?$$

$\longleftarrow$ 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

$$Pv = \lambda v$$

Multiplicar *v* por *P* é igual a reescalar *v* por λ

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

$$uP = \lambda u$$

Multiplicar *u* a esquerda de *P*

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

autovetor normal

$$P'v = \lambda u$$

P' é a transposta da matriz *P*

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

Figueiredo 2018

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 λ = 1
- Precisamos ainda $\sum_{s \in S} \pi_s = 1$ impomos que π_s seja um vetor de probabilidade
- Solução: normalizar o autovetor para garantir soma 1
- **Teorema:** Se *P* é uma matriz estocástica, temos |λ| <= 1 para todo autovalor, e apenas um autovalor λ = 1
  - matriz estocástica = matriz de transição de probabilidade

↳ matriz cujos elementos estão entre 0 e 1 e cuja soma da linha seja 1.

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)

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

A matriz Q^{-1} nos dá os autovetores à esquerda de P!

Autovetor à esquerda associado ao λ=1

$$Q = \begin{pmatrix} 1 & 2 & -43 \\ 1 & -5 & 13 \\ 1 & 2 & 55 \end{pmatrix} \quad L = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.3 & 0 \\ 0 & 0 & -0.4 \end{pmatrix} \quad Q^{-1} = \begin{pmatrix} 43/98 & 2/7 & 27/98 \\ 3/49 & -1/7 & 4/49 \\ -1/98 & 0 & 1/98 \end{pmatrix}$$

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!

Figueiredo 2018

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

$$P^t = PP\ldots P = QLQ^{-1}QLQ^{-1}\ldots QLQ^{-1}$$
$$= QLILI\ldots LQ^{-1} = QLL\ldots LQ^{-1} = QL^tQ^{-1}$$

Matriz diagonal, com cada elemento elevado a potência *t*

- Logo, temos que

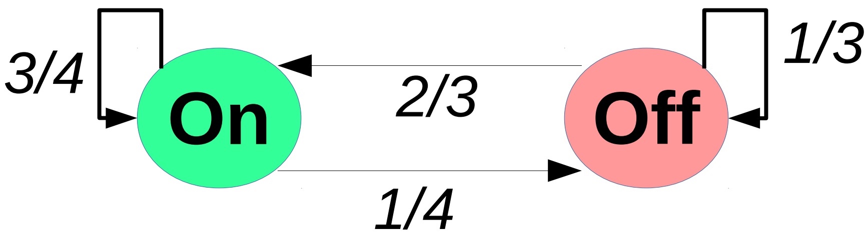
$$\pi(t) = \pi(0)QL^tQ^{-1}$$

- Para onde vai L^t com *t* crescente ?

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

- $\lambda = 1$ 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

$$\pi(0)=(1\ 0)=\underbrace{\pi+\overset{\lambda_2}{3/11}\,v_2}$$

→ π(0) é a soma de 2 vetores

- Temos então

Forço aparecer π

π é a própria distribuição estacionária

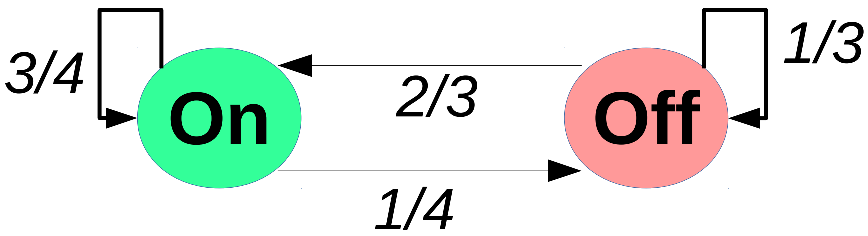
$$\pi(t)=\pi(0)P^t=(\pi+3/11\,v_2)P^t=\pi P^t+3/11\,\underbrace{v_2 P^t}$$

$v_2 P^t = \lambda_2^t v_2$

$$=\pi+3/11\,\lambda_2^t v_2$$

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

Exemplo



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

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

- Temos então

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

- Distância de variação total

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

Converge exponencialmente rápido!

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

O que depende da CM
é a base da
exponencial

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

$$\underbrace{\max_{\pi(0)} d_{TV}\left(\pi(t), \pi\right)}_{\text{Pior } \pi(0)} \le C\,\alpha^{t}$$

Pra cada $\pi(0)$ tenho uma convergência diferente. Para o pior $\pi(0)$ [ou seja, aquele que MAXIMIZA $d_{TV}(\pi(t),\pi)$. mesmo com este $\pi(0)$, o tempo de convergência decresce exponencialmente

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

α representa todos os autovalores menores que 1. O que estiver mais próximo de 1 demora mais para ir a 0. Assim, o 2º maior autovalor é o nosso "freio-de-mão" na convergência

Figueiredo 2018

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

$$t = \frac{1}{0.08} \log\left(\frac{11}{3}\epsilon\right) = \frac{1}{0.08} \log \varepsilon + \frac{\log 11/3}{0.08}$$

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

- 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 $\varepsilon > 0$

mínimo tempo que faça com que minha d_{TV} seja menor que ε

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

mistura → evolução das probabilidades

- τ_ε : menor valor de *t* tal que para qualquer distribuição inicial, π(t) está a distância menor que ε da estacionária
- Para qualquer CM aperiódica, irredutível, temos

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

Após chegar a uma distância de 1/4 de τ_1, a convergência é bem rápida! Isso significa que a maior parte dos passos é gasta no início

- Depois de estar perto o suficiente, chegar mais perto é muito fácil (fator log ε^{-1})
- Tempo de mistura depende fracamente em ε
  - ε vai ser tomado como constante

Spectral Gap

"VÃO ESPECTRAL"

- 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)

2º maior autovalor

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

Distância entre os 2 maiores autovalores

- Quanto maior for δ, mas rápido é convergência → faço com que a base que domina a exponencial seja pequena.
  - 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)

VÃO ESPECTRAL GRANDE → MAIS RÁPIDA A CONVERGÊNCIA

Spectral Gap e Tempo de Mistura

- Relação entre δ e τ_ε

CM convencional

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

$\pi_o \rightarrow \min_i \pi_i \Rightarrow$ menor componente de π

precisamos descobrir

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

- Maior δ, menor τ_ε
- Maior π_o, menor τ_ε

Nos dá um limitante inferior e superior

Limitante inferior para tempo de mistura

Limitante superior para tempo de mistura

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

Se eu tenho A, eu tenho π.

Se eu tenho π, eu já tenho τ_ε

Figueiredo 2018

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 ½, 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

$\tau_n \leq c\,d\log d$ ← chamado de "*fax mixing*"

Tema atual de pesquisa na matemática (e computação)