CPSC 531F, Feb 25

Last time:

$$\begin{array}{c|cc} & S_1 & S_2 \\ \hline S_1 & .99 & .01 \\ S_2 & .02 & .98 \end{array}$$

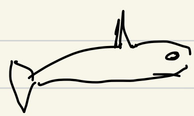
human

Dolphin - "more information"

S_1 really 2 states S_1', S_1''

In S_1' or S_1'' $\xrightarrow{.02}$ S_2

$$\begin{array}{c|cc} & S_1' & S_1'' \\ \hline S_1' & .98 & .01 \\ S_1'' & .01 & .98 \end{array} \quad \begin{array}{l} \to .99 \\ \to .99 \end{array}$$



dolphin

Dolphin:

$$
\begin{array}{cc}
 & \begin{array}{ccc} & & S_2 \end{array} \\
\begin{array}{c} S_1' \\ S_1'' \\ S_2 \end{array} &
\begin{bmatrix}
.98 & .01 & .01 \\
.01 & .98 & .01 \\
? & ? & .98
\end{bmatrix}
\end{array}
$$

Dolphin

say!

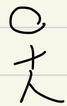
$$
\begin{bmatrix}
- & - & - \\
- & - & - \\
.01 & .01 & .98
\end{bmatrix}
$$

Dolphin sees a symmetric matrix

$$
\begin{bmatrix}
.98 & .01 & .01 \\
.01 & .98 & .01 \\
.01 & .01 & .98
\end{bmatrix}.
$$

This!  point of view

 point of view

- Dolphin refines the  POV
-  POV is obtained from projecting the dolphin POV

$\longleftrightarrow$ reversible Markov chain

Goal: How can  sees Markov matrix, P, tell if you can view P as symmetric?

When we study Markov chains we are typically interested in

- mixing times ← focus on
- other properties ...

Last time: "expander mixing lemma"

We also want a lot of examples of Markov chains & applications.

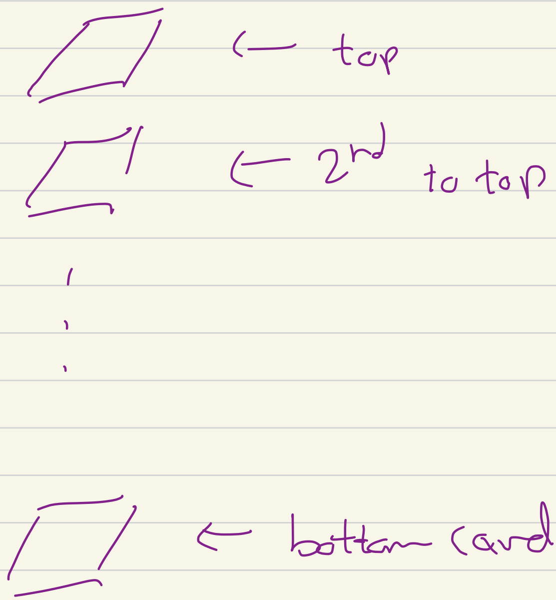
Textbook by Levin & Peres

Markov chain as $X_0, X_1, \ldots$

we'll mostly talk about Markov matrices

Markov matrix: $P \in M_n(\mathbb{R})$ with all rows being stochastic vectors.

Think about shuffling cards: 52 cards



Each card has one of 52 values

A♠ K♠ , - -   2 clubs

Each card, in one of 52 positions.

Shuffling: $52! \times 52!$ matrix 

Fix some shuffles

- take top card, put it in random position
- swap 2 cards at random
- riffle shuffle (randomness built in)
- ...

Cards : φ : Set $\xrightarrow{\text{bijection}}$ Set

A sp, K sp, ..., 2 clubs   Positions: top, ..., bottom

alternatively

φ^{-1} : Positions $\longrightarrow$ Card Labels

φ, φ^{-1} maps Set size $52! \to$ Set size $52!$

Cayley graphs:

First define mixing time:

$P \in M_n(\mathbb{R})$, (thinking of $n = 52!$)

Idea: You start entirely in some single state

$\vec{e}_1^{\,T} = [1\ 0\ \dots\ 0]$, $\vec{e}_2^{\,T} = [0\ 1\ 0\ \dots\ 0]$,

$\dots$, $\vec{e}_n^{\,T} = [0\ 0\ \dots\ 0\ 1]$

Say someone hands you 1 decks of cards! $\vec{e}_i$, $i \in [n]$ (n = 52! for cards)

start shuffling!

$\vec{e}_i^{\,T} P^t$ after applying P t times

(cards P = single shuffle)

=

Want to claim that for some t (that we can identify), $\vec{e}_i^{\,T} P^t$ puts the Markov chain into a "random order" to within ε, ($\varepsilon > 0$, real parameter)

Meaning:

— $\underbrace{\text{distrib}}_{\underbrace{\vec{e}_i^{\,T} P^t}_{\mu}}$ close to $\underbrace{\text{uniform}}_{\underbrace{\pi}_{\nu}}$:

$\pi \leftarrow$ stationary distrib of Markov chain

$\nu \leftarrow$ maybe ν is uniform

What does it mean to say that μ and ν are close to each other?

— $\underbrace{\text{dist}}_{}(\mu, \nu) \leq \varepsilon$

dist = ?

$\mu, \nu \in \mathbb{R}^n$

① $\max\limits_{i \in [n]} |\mu_i - \nu_i|$

② $\|\mu - \nu\|$, where $\| \ \| = \begin{cases} \|\cdot\|_2 \\ \\ \\ \end{cases}$

$\|\cdot\|_2 \longleftrightarrow$ well suited to eigenvectors / values

$\|\cdot\|_1 \longleftrightarrow$ essentially how mixing time is defined

$\| \ \|_\infty \longleftrightarrow$ max component

$\| \ \|_{T.V.} \longleftrightarrow$ total variation

$\| \ \|_p \longleftrightarrow$ L^p norm

$$\|\vec{x}\|_p = \left(\sum_{i \in [n]} |x_i|^p \right)^{1/p}$$

$$\|\vec{x}\|_2 = \| \ \|_2, \quad \| \ \|_\infty = \max$$

$$= \lim_{p \to \infty} \| \ \|_p$$

$$\| \ \|_1 = \|\cdot\|_p, \ p = 1$$

Total Variation:

Idea!

- Someone gives you $\vec{e}_i$, you shuffle and get $\vec{e}_i^{\,T} p^t$

Imagine that you had been given $\vec{e}_j$ instead

You allow one test to see if

$$\vec{e}_i^{\,T} P^t \quad \text{differs from} \quad \vec{e}_j^{\,T} P^t$$

tests: $A \subset [n]$ and look at

$$\underbrace{e_i^T P^t e_A}_{\sum_{a \in A} (e_i^T P^t)(a)} \quad \text{vs.} \quad \underbrace{e_j^T P^t e_A}_{\sum_{a \in A}}$$

Define:

$$\|\vec{x}\|_{TV} = \max_{A \subset [n]} \left| \overbrace{\sum_{a \in A} x_a}^{\text{not } \sum_{a \in A} |x_a|} \right|$$

If μ, ν are stochastic vectors,

$$\text{Total Var}(\mu, \nu) = \max_{A \subset [n]} \left| \sum_{a \in A} \mu_a - \sum_{a \in A} \nu_a \right|$$

$$= \max_{A \subset [n]} \sum \mu_a - \sum \nu_a$$

$$= \max_{B \subset [n]} \sum_{b \in B} \nu_b - \sum_{b \in B} \mu_b$$

$$= \frac{1}{2} \| \mu - \nu \|_1$$

$$\mu = \begin{pmatrix} 1/3 \\ 1/3 \\ 1/3 \end{pmatrix}, \quad \nu = \begin{bmatrix} 1/2 \\ 1/2 \\ 0 \end{bmatrix}$$

$\nu = 1/2$	$\cancel{\nu}\mu \neq 1/2$	
$\mu = 1/3$	$\mu \neq 1/3$	$\mu = 1/3$
		$\nu = 0$
comp 1	comp 2	comp 3
$\nu \geq \mu$		$\nu \leq \mu$
$1/2$ vs $1/3$	$1/2$ vs $1/3$	0 vs $1/3$

Diff $\frac{1}{6} + \frac{1}{6}$

Diff $-\frac{1}{3}$

Def 1: For P stochastic matrix, the ε-mixing time of P is the smallest $t = t(\varepsilon)$ s.t.

$$\max_{i,j} \| \vec{e}_i P^t - \vec{e}_j P^t \|_{TV} \leq \varepsilon.$$

Claim: If

$$\bar{d}(t) = \max_{\mu, \nu \text{ stochastic}} \| \mu P^t - \nu P^t \|_{TV}$$

$$= \max_{i,j \in [n]} \| \vec{e}_i P^t - \vec{e}_j P^t \|_{TV}$$

Then for any $s, t \in \mathbb{N}$

$$\bar{d}(s+t) \leq \bar{d}(s)\,\bar{d}(t).$$

I.e. $t = t(\varepsilon)$ is the ε mixing time of P

$t' = t(\varepsilon')$ " " ε' mixing time of P,

the $\varepsilon \cdot \varepsilon'$-mixing time of P is

at most $t \cdot t'$

Our goal: given $P \in M_n(\mathbb{R})$

Markov matrix, estimate the

ε-mixing time of P

$n = 52!$, even to get to

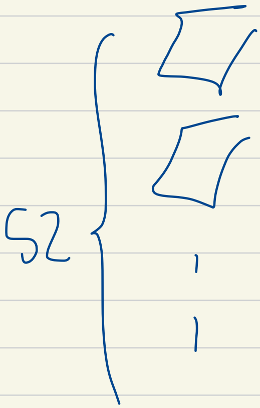
all states with prob close to

$\frac{1}{52!}$ is difficult...

Next task: define Cayley graphs...

4-min break

Cayley Graphs:



$n = 52!$

$P \in M_{52!}(\mathbb{R})$

Shuffles $\longleftrightarrow$ a single shuffle

$\longleftrightarrow$ random pick some way of operating on the order

Speak of

$$S_{52} = \{ \text{permutations on } [52] \}$$

$S_3 = \{$ permutations on $[3]\}$

Permutation on $[3] = \{1, 2, 3\}$

$f : [3] \xrightarrow{\text{bijection}} [3]$

e.g.

S_3 contains

$$\begin{matrix} 1 \mapsto 1 \\ 2 \mapsto 2 \\ 3 \mapsto 3 \end{matrix}, \quad \begin{matrix} 1 \mapsto 2 \\ 2 \mapsto 1 \\ 3 \mapsto 3 \end{matrix}, \ldots$$

=

Elt of S_{52}

$$\begin{matrix} 1 \mapsto 7 \\ 2 \mapsto 17 \\ 3 \mapsto 1 \\ \vdots \\ 52 \mapsto \cdot \end{matrix}$$

If we take a random "transposition"

on $[3] = \{1, 2, 3\}$

$$\mathrm{Tr}_{1,2} \qquad \mathrm{Tr}_{1,3} \qquad \mathrm{Tr}_{2,3}$$

$$\begin{matrix} 1 \mapsto 2 \\ 2 \mapsto 1 \\ 3 \mapsto 3 \end{matrix}, \qquad \begin{matrix} 1 \mapsto 3 \\ 2 \mapsto 2 \\ 3 \mapsto 1 \end{matrix}, \qquad \begin{matrix} 1 \mapsto 1 \\ 2 \mapsto 3 \\ 3 \mapsto 2 \end{matrix}$$

do this prob $1/3$ this prob $\frac{1}{3}$ prob $1/3$

This gives a Markov chain:

$$\cdots \left(\qquad \right) \left(\begin{matrix} T_{12} & 1/3 \\ T_{13} & 1/3 \\ T_{23} & 1/3 \end{matrix} \right) \left(\begin{matrix} \text{start} \\ \text{somewhere} \end{matrix} \right)$$

$$\underbrace{\left(Tr_{1,2}\right)\left(Tr_{1,3}\right)}_{\text{new permutation}}\left(\begin{matrix}\text{start}\\\text{here}\end{matrix}\right)$$

=

Formally, we generalize $\mathcal{S}_n$ as a "group" $g_1, g_2 \longmapsto$ "g_1 times g_2"

Formally: A group is a set $\mathcal{G}$, s.t. there is a "multiplication" defined on $\mathcal{G}$

$$\mathcal{G} \times \mathcal{G} \xrightarrow{\text{mult}} \mathcal{G}$$

$$(g_1, g_2) \longrightarrow g_1 \text{ times } g_2$$

1 $(g_1 g_2) g_3 = g_1 (g_2 g_3)$

(i.e. $g_1 g_2 g_3$ is defined independent of order)

(not generally $g_1 g_2 = g_2 g_1$)

2 there is an $e \in \mathcal{G}$ s.t.

$e g_1 = g_1 e = g_1$ for all $g_1 \in \mathcal{G}$

3 for every $g \in \mathcal{G}$ there is a

"g^{-1}" in G s.t.,

$$g^{-1} g = g g^{-1} = e$$

Section 2.6 of [LP]

<u>Rem</u>: Cyclic graphs are described by the group

$\mathbb{Z}/_{n\mathbb{Z}}$ or integers mod n

$\boxed{1/2}$ $\quad$ $\boxed{1/2}$ $\quad$ | $\quad$ $1/3$

$1/3$ $\quad$ $1/3$ $\quad$ | $\quad$ $\boxed{0}$

$1/6 \quad 1/6 \quad -1/3$

$\square - \bigcirc$

$$\| \square - \bigcirc \|_{TV} = \frac{1}{3}$$

$$\| \square - \bigcirc \|_{TV} = \frac{1}{6} + \frac{1}{6} + \frac{1}{3} = 2 \cdot \frac{1}{3}$$

$\boxed{10\left(\frac{1}{2}\right)}$ $\quad$ $\boxed{10\left(\frac{1}{2}\right)}$

$10\left(\frac{1}{3}\right)$ $\quad$ $10\left(\frac{1}{3}\right)$ $\quad$ $10\left(\frac{1}{3}\right)$

$\boxed{10(0)}$

non-neg vectors

$$\sum \square = \sum \bigcirc = 10$$

$9\left(\tfrac{1}{2}\right) + \boxed{1\left(\tfrac{1}{2}\right)}$ $\quad$ $9\left(\tfrac{1}{2}\right) + \boxed{1\left(\tfrac{1}{2}\right)}$

$\boxed{10\left(\tfrac{1}{2}\right)}$ $\quad$ $\boxed{10\left(\tfrac{1}{2}\right)}$

$1\left(\tfrac{1}{3}\right)$ $\quad$ $1\left(\tfrac{1}{3}\right)$ $\quad$ $1\left(\tfrac{1}{3}\right)$

$\boxed{10(0)}$

$$\underbrace{9\left(\tfrac{1}{2}\right) + 9\left(\tfrac{1}{2}\right) + \tfrac{1}{3}} \qquad \underbrace{1/3}$$

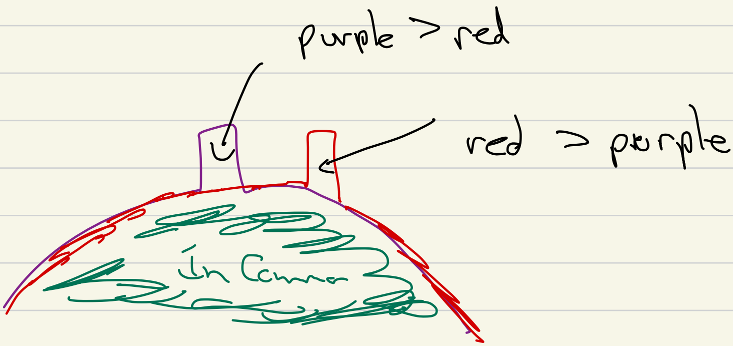
areas both equal T.V. dist.

μ $\quad$ ν

$$\|\mu - \nu\|_{L^1} = 2\,\|\mu - \nu\|_{TV}$$

L_1 norm

$$\sum_{i \in [n]} |\mu_i - \nu_i|$$



$|\mu_i - \nu_i|$ count both spikes

Claim: If $\hat{P}$ is any Markov matrix:

$$\max_{\mu,\nu \text{ stoch}} \|\vec{\mu}\hat{P} - \vec{\nu}\hat{P}\|_{TV}$$

This max is attained for

$\vec{\mu} = \vec{e}_i$, $\vec{\nu} = \vec{e}_j$ for some $i,j \in [n]$

$$\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix} = \hat{P}$$

$\|\mu\hat{P} - \nu\hat{P}\|_{TV}$ — worst case is

stochastic

$\vec{e}_1, \vec{e}_2$

$\vec{e}_1, \vec{e}_3$

$\vec{e}_1$, $\frac{1}{10}\vec{e}_2 + \frac{9}{10}\vec{e}_3$

$$\mu \tilde{P} - \nu \tilde{P}$$

$\mu_1 \ \mu_2 \ \mu_3$ $\qquad$ $\nu_1 \ \nu_2 \ \nu_3$

$$[\mu_1 \ \mu_2 \ \mu_3] \tilde{P} - [\nu_1 \ \nu_2 \ \nu_3] \tilde{P}$$

$$= [\mu_1 - \nu_1 \quad \mu_2 - \nu_2 \quad \mu_3 - \nu_3] \tilde{P}$$

$=$ $\quad$ $= (\mu_1 - \nu_1)$ 1st row of $\tilde{P}$

$+ (\mu_2 - \nu_2)$ 2nd row of $\tilde{P}$

$+ (\mu_3 - \nu_3)$ 3rd row of $\tilde{P}$

$$\left| \overset{\geq 0}{(\mu_1-\nu_1)}\ \text{row }1 + \overset{\geq 0}{(\mu_2-\nu_2)}\ \text{row }2 + \overset{\leq 0}{(\mu_3-\nu_3)}\ \text{row }3 \right|_1$$

all numbers between $-1, 1$

Sum of any subset $-1, 1$

vs. $\max_{i,j}$ row i − row j

$n \in \mathbb{N}$

sum ≤ 1

$\underbrace{\mu_1-\nu_1}_{x_1}, \underbrace{\mu_2-\nu_2}_{x_2}, \ldots, \underbrace{\mu_s-\nu_s}_{x_s} \geq 0$

$\underbrace{\mu_{s+1}-\nu_{s+1}}_{y_{s+1}}, \ldots, \underbrace{\mu_n-\nu_n}_{y_n} \leq 0$

sum ≥ -1

If $x_1, \ldots, x_s = \frac{1}{s}$, $y_{s+1}, \ldots, y_n = \frac{1}{5}$

then

$$\leq \frac{1}{S} \sum \| row_i - row_{i+S} \|$$

$S = 10^{10^{10}}$

$\| M \|_1$

have nice expressions

$\| M \|_\infty$

give HW on $\| \; \|_1$, $\| \; \|_\infty$

$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ eigenvect $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ value $2 = \lambda_1$

$\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ $0 = \lambda_2$

$$\frac{\left\| \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_2}{\left\| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\|_2} = 2 = \lambda_1$$

$$\| E \|_{L^2} = \max\left(|\lambda_2|, \dots, |\lambda_n| \right)$$