Mixing Time

CS 70, Summer 2019

Bonus Lecture, 8/9/19

Disclaimer:

Much handwaving today!!

Goal is to get a high level intuition / picture for the concept of mixing time and applications

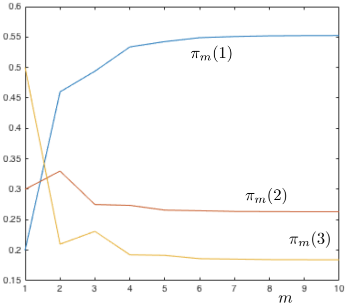
Emphasis is on **heuristics** rather than rigor

What We Know...

Every irreducible, aperiodic Markov chain has a **unique** stationary distribution.

$\lim_{n\to\infty} \mu^{(n)} = \pi$

$$\pi_m = \pi_0 P^m = \pi_0 \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0 & 0.3 & 0.7 \\ 0.6 & 0.4 & 0 \end{bmatrix}^m .$$



Q: How long does it take to get **close** to the stationary distribution?

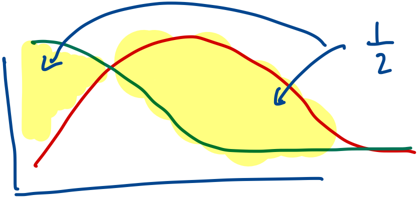
Total Variation Distance

Just one of many ways to measure how **close** two distributions are.

over states $\{1, 2, \dots, n\}$

Let P_1, P_2 be two PMFs. Their TV distance is:

$$\frac{1}{2} \sum_{\text{states } i} \left| P_1(i) - P_2(i) \right|$$



$\frac{1}{2}$ of the highlighted

Mixing Time: Definition

I have an **irreducible, aperiodic** Markov chain.
Notation: $\mu^{(n)}$ is the distribution **at time** n, and π is its **(unique) stationary distribution**.

I want to keep running my chain until:

$$d_{TV}(\mu^{(n)}, \pi) \leq \varepsilon$$

small positive number

dist at time n

stationary

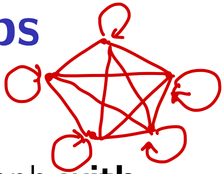
The mixing time $t_{mix}(\varepsilon)$ is the first time this happens.

worst case starting dist.

(Omitted fact: The TV distance between $\mu^{(n)}$ and π "decreases" as n increases.)

with high prob, at least

Complete Graph With Loops



Mixing time analysis sometimes direct!

Take a random walk on a complete graph **with loops**. What is the stationary distribution?

K_n : $\pi = \left[\frac{1}{n}\ \frac{1}{n}\ \cdots\ \frac{1}{n}\right]$ symmetry between vertices.

What does the transition matrix look like?

$$P = \begin{bmatrix} \frac{1}{n} & \frac{1}{n} & \cdots & \frac{1}{n} \\ \frac{1}{n} & \ddots & & \\ \vdots & & \ddots & \\ \frac{1}{n} & & & \end{bmatrix}, \quad \mu P = \text{always } \left[\frac{1}{n}\ \frac{1}{n}\ \cdots\ \frac{1}{n}\right] = \pi$$

Mixing Time? Dependence on ε?

1. for all ε.

Random-To-Top Shuffling I

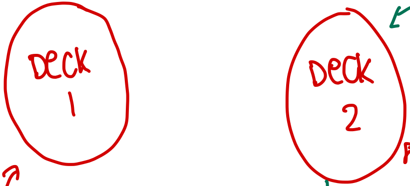
$\mathcal{S}$: All orderings of n cards in a deck. $n!$ configs.

Transitions: Choose card randomly in the deck. Move it to the top.

π here is uniform over all states.

Different strategy called **coupling:**

assume already mixed.



Start at arbitrary state.

chosen u.a.r.

If we keep applying P, dist. stays at π.

Random-To-Top Shuffling II

At each time, each card is labeled **coupled** or not.
number

Initially, all cards start **uncoupled**.
Pick a random card C.
number

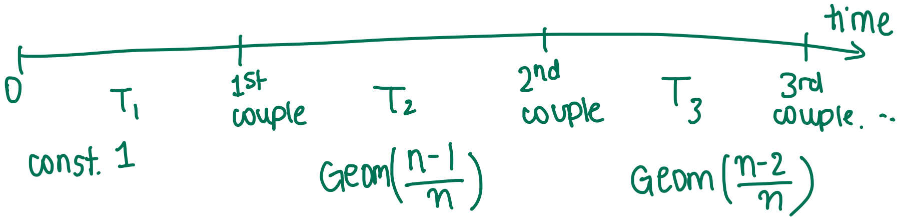
In both decks, take card C and move it to the top.
If C isn't already coupled, mark it as **coupled**.

What happens when we look at each deck individually? Looks like random-to-top move!

* once a # is labeled "coupled", it has the same position in both decks!

Random-To-Top Shuffling III

Time until all cards get coupled = time until Deck 1 is fully random.



↑ coupon collector.

$\mathbb{E}[T_1 + T_2 + \ldots + T_n] = n \log n.$

For all ϵ: $t_{mix}(\varepsilon) = O(n \log n)$ "no dep. on ε" = "cutoff"

Random Hypercube Walk I

Take an n-dimensional hypercube. vertices n-length bitstrings

Stationary distribution of hypercube walk:

$$\pi = \left[\frac{1}{2^n} \quad \frac{1}{2^n} \quad \cdots \quad \frac{1}{2^n}\right]$$

all vertices same deg.

Try coupling again:

(*) Lazy walk: Every vertex self loop wp $\frac{1}{2}$

transitions to nbr wp $\frac{1}{2n}$

arbitrary vertex: → truly random:

randomize all bits

Random Hypercube Walk II

Each **coordinate** is labeled **coupled** or not.

Initially all coordinates are **uncoupled**.
Pick a random coordinate i.
Flip a coin to set the i-th coordinate to 0 or 1.
If i-th coordinate is **uncoupled**, set it to **coupled.**

$\mathbb{E}[\text{time all coord. randomized}] = n \log n$

Random Hypercube Walk III

Use Coupon Collector again!

same as before.

For all ε, $t_{\text{mix}}(\varepsilon) =$

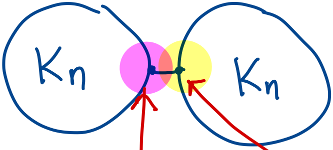
Conditions for Fast Mixing?

Complete graph K_n? ≈ 1

Path on n vertices? Start in middle.

Mixing time n^2

Dumbbell?



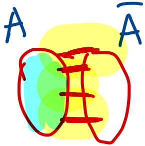
mixing time n^2

$\frac{1}{n}$ chance getting here from left side

$\frac{1}{n}$ chance getting here from pink dot.

Bottlenecks

We use a measurement called the **conductance** to quantify the notion of a bottleneck.



The conductance of a set $A \subseteq \mathcal{S}$ is:

$$\Phi(A) = \frac{\sum_{i \in A, j \in \bar{A}} \pi(i) P_{ij}}{\sum_{i \in A} \pi(i)} \leftarrow \mathrm{vol}(A)$$

The conductance of the chain M is:

$$\Phi(M) = \min_{\substack{A \text{ s.t.} \\ \mathrm{vol}(A) < \frac{1}{2}}} \Phi(A)$$

How To Measure Conductance?

Measuring conductance = looking at all subsets of states with $\mathrm{vol}(A) \leq \frac{1}{2}$.
How many subsets, potentially?

Alternative: Get **lower bound** on $\Phi(M)$ using **second largest eigenvalue** of transition matrix.

$$\Phi(M) \geq \frac{1-\lambda_2}{2}$$

Eigenvalues are much faster to compute!

Markov Chain Monte Carlo

Monte Carlo: randomized algorithm where the output is allowed to be incorrect

Use cases:

- sampling from complicated distributions
- counting combinatorial objects
- Bayesian inference
- statistical physics
- volume estimation, integration

Markov Chain Monte Carlo

Key idea: Design a Markov chain so that its stationary follows the distribution that you want to sample from. Run the chain, wait for it to mix.

Runtime depends on... $t_{mix}(\varepsilon)$!