Lecture 8: Convergence and Mixing Time

Nicolás Rivera John Sylvester Luca Zanetti Thomas Sauerwald

Lent 2019

UNIVERSITY OF CAMBRIDGE

Outline

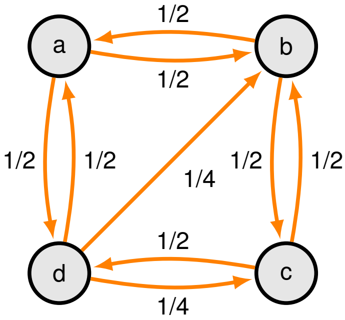
Periodicity and Convergence

Total Variation Distance

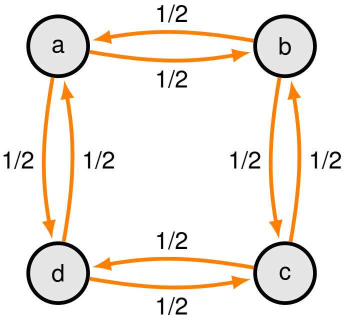
Mixing Times

Periodicity

- A Markov chain is *Aperiodic* if for all $x, y \in \mathcal{I}$, $\gcd\{t : P^t_{x,y} > 0\} = 1$.
- Otherwise we say it is *Periodic*.



✓ Aperiodic

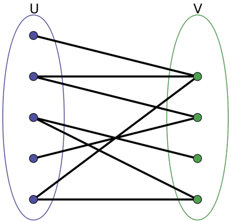


× Periodic

Random Walks and Bipartiteness

Bipartite Graph

A graph is *bipartite* if its vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V.



Theorem

Let G be an undirected connected graph. Then the Simple random walk on G is aperiodic if and only if G is non-bipartite

Proof on the Visualiser

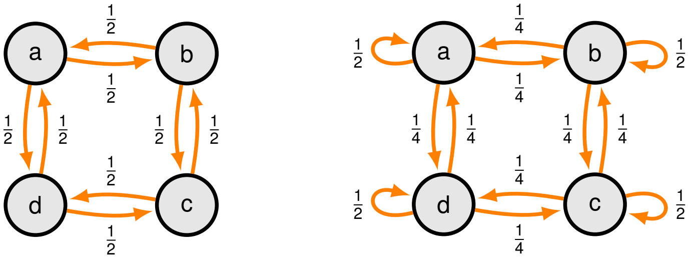
Lazy Random Walks and Periodicity

The *Lazy Random Walk (LRW)* on G given by $\widetilde{P} = (P + I)/2$,

$$\widetilde{P}_{i,j} = \begin{cases} \frac{1}{2d(i)} & \text{if } ij \in E \\ \frac{1}{2} & \text{if } i = j \\ 0 & \text{Otherwise} \end{cases}.$$

P - SRW matrix
I - Identity matrix.

Fact: for any graph G the LRW on G is Aperiodic.



SRW on C_4, *Periodic*

LRW on C_4, *Aperiodic*

Convergence

Convergence Theorem

Let P be any finite, aperiodic, irreducible Markov chain with stationary distribution π. Then for any $i, j \in \mathcal{I}$

$$\lim_{t \to \infty} P^t_{j,i} = \pi_i.$$

- Proved : For finite irreducible Markov chains π exists, is unique and

$$\pi_x = \frac{1}{\mathbf{E}_x\left[\tau_x^+\right]} > 0.$$

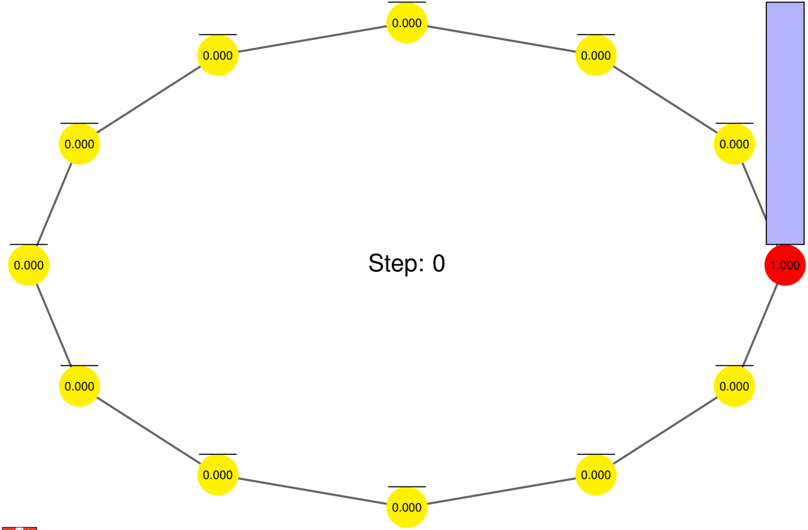
- Luca will prove the Convergence Theorem assuming Reversibility .
- If $P^t_{j,i}$ converges for all i, j we say the chain *Converges to Stationarity*.

Corollary

The Lazy random walk on any finite connected graph converges to stationarity.

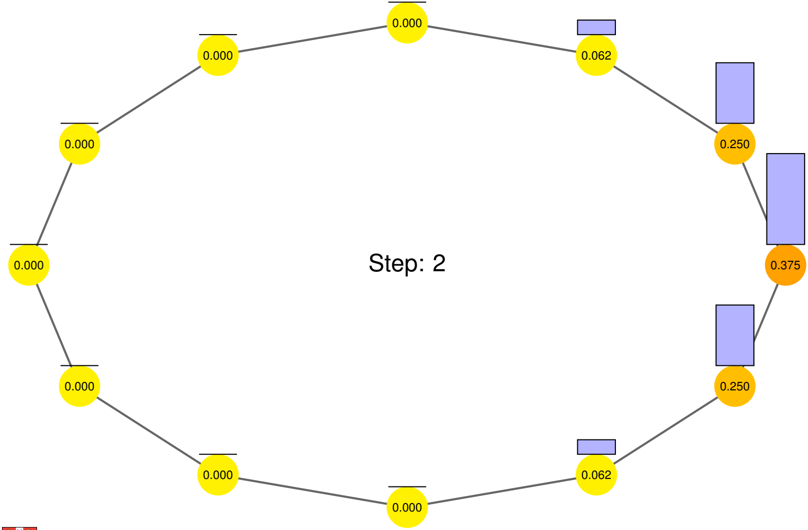
Convergence to Stationarity for the LRW on C_{12} from 0

At step t the value at vertex x is $P^t_{0,x}$.



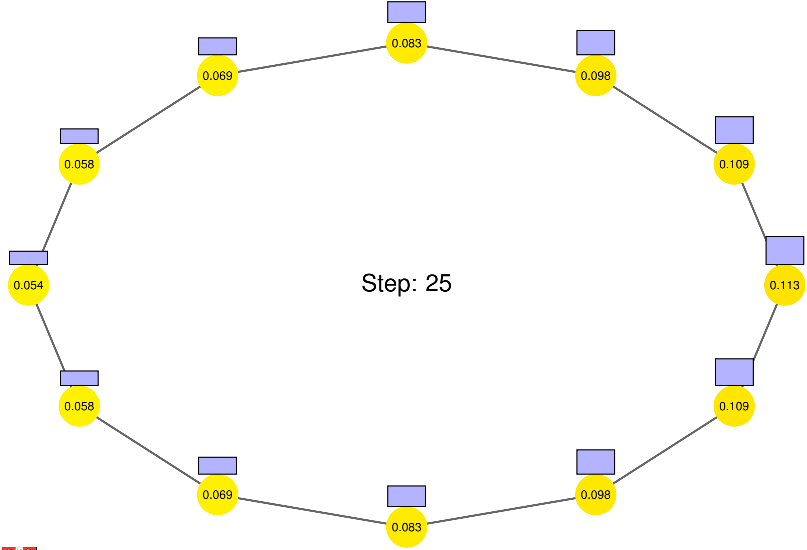
Convergence to Stationarity for the LRW on C_{12} from 0

At step t the value at vertex x is $P^t_{0,x}$.



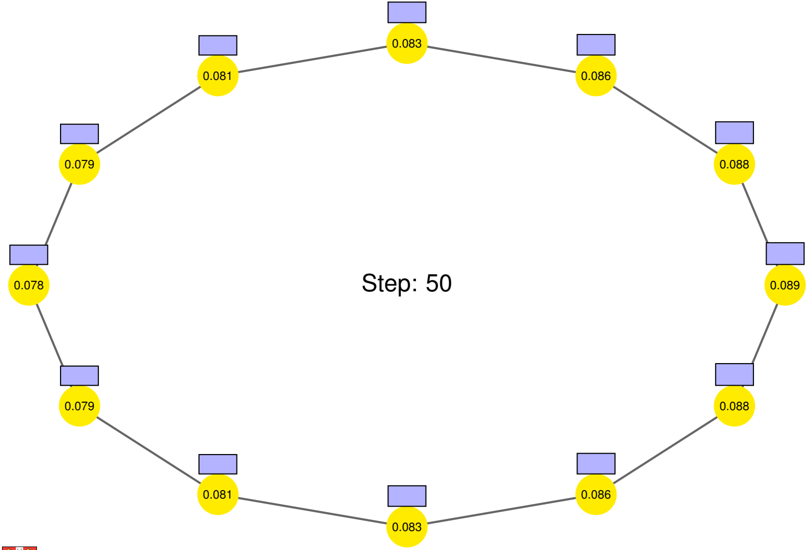
Convergence to Stationarity for the LRW on C_{12} from 0

At step t the value at vertex x is $P^t_{0,x}$.



Convergence to Stationarity for the LRW on C_{12} from 0

At step t the value at vertex x is $P^t_{0,x}$.



Outline

Periodicity and Convergence

Total Variation Distance

Mixing Times

How Similar are Two Probability Measures?

Loaded Dice

- I present to you three *loaded* (unfair) dice A, B, C:

x	1	2	3	4	5	6
$\mathbf{P}[A = x]$	1/3	1/12	1/12	1/12	1/12	1/3
$\mathbf{P}[B = x]$	1/4	1/8	1/8	1/8	1/8	1/4
$\mathbf{P}[C = x]$	1/6	1/6	1/8	1/8	1/8	9/24

- Question 1 : Which dice is the least fair ?
- Question 2 : Which dice is the most fair ?

Question 1: Most of you choose A. Why?

Question 2: Dice B and C seem "fairer" than A but which is fairest?

Question 3 : What do we mean by "fair"?

Total Variation Distance

The *Total Variation Distance* between two probability distributions μ and η on a countable state space Ω is given by

$$\|\mu - \eta\|_{tv} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)|.$$

- Let $d(\mu, \nu) = \|\mu - \nu\|_{tv}$, then $d(\cdot, \cdot)$ is a metric on the space of measures.

Loaded Dice : let $D = Unif\{1, 2, 3, 4, 5, 6\}$ be the law of a fair dice:

$$\|D - A\|_{tv} = \frac{1}{2}\left(2\left|\frac{1}{6} - \frac{1}{3}\right| + 4\left|\frac{1}{6} - \frac{1}{12}\right|\right) = \frac{1}{3}$$

$$\|D - B\|_{tv} = \frac{1}{2}\left(2\left|\frac{1}{6} - \frac{1}{4}\right| + 4\left|\frac{1}{6} - \frac{1}{8}\right|\right) = \frac{1}{6}$$

$$\|D - C\|_{tv} = \frac{1}{2}\left(3\left|\frac{1}{6} - \frac{1}{8}\right| + \left|\frac{1}{6} - \frac{9}{24}\right|\right) = \frac{1}{6}.$$

Thus

$$\|D - B\|_{tv} = \|D - C\|_{tv} \qquad \text{and} \qquad \|D - C\|_{tv}, \|D - C\|_{tv} < \|D - A\|_{tv}.$$

So *A* is the least "fair" however *B* and *C* are equally "fair" (in TV distance).

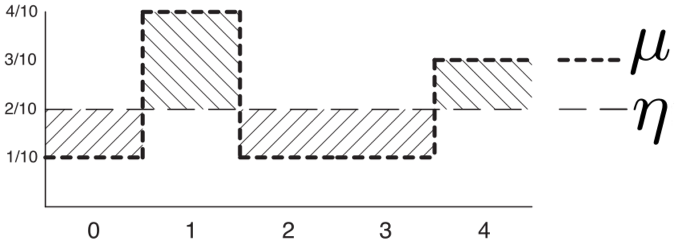
Total Variation Distance

Lemma

For any probability distributions μ and η on a countable state space Ω

$$\|\mu - \eta\|_{tv} := \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \eta(\omega)| = \sup_{A \subset \Omega} |\mu(A) - \eta(A)|.$$

Proof by picture.



□

TV Distances

Let P be a Markov Chain with stationary distribution π.

- Let μ be a prob. vector on $\mathcal{I}$ (might be just one vertex) and $t \geq 0$. Then

$$P_\mu^t := \mathbf{P}_\mu[X_t = \cdot] = \mathbf{P}[\, X_t = \cdot \mid X_0 \sim \mu\,],$$

  is a probability measure on $\mathcal{I}$.

- For any μ,

$$\left\| P_\mu^t - \pi \right\|_{tv} \leq \max_{x \in \mathcal{I}} \left\| P_x^t - \pi \right\|_{tv}.$$

Convergence Theorem (rephrased)

For any finite, irreducible, aperiodic Markov Chain

$$\lim_{t \to \infty} \max_{x \in \mathcal{I}} \left\| P_x^t - \pi \right\|_{tv} = 0.$$

Outline

Periodicity and Convergence

Total Variation Distance

Mixing Times

Applications of Markov Chain Convergence

Markov Chain Monte Carlo (MCMC): Sampling, Counting, Integration, ...

Example : Markov Chain for Sampling a Matching of G.

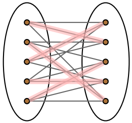
Pick some initial matching M (may have no edges)

1. With probability $1/2$ stay at M
2. Otherwise pick $uv \in E$ and let

$$M' = \begin{cases} M - \{uv\} & \text{if } uv \in M \\ M \cup \{uv\} & \text{if } uv \text{ can be added to } M \\ M \cup \{uv\} - \{e'\} & \text{if either } u \text{ or } v \text{ is matched by } e' \in M \\ M & \text{otherwise} \end{cases}$$

3. Let $M = M'$ and repeat steps $1 - 3$.

- Markov Chain on Matchings of G.
- Satisfies the Convergence theorem.
- Has uniform stationary distribution.
- Thus run it "long enough" then halt to return a uniform matching on G.



$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Mixing Time of a Markov Chain

Convergence Theorem: "Nice" Markov chains converge to stationarity.

Question How fast do they converge?

The *Mixing time* $\tau(\epsilon)$ of a Markov chain P with stationary distribution π is

$$\tau(\epsilon) = \min\left\{ t\colon\ \max_x \left\| P_x^t - \pi \right\|_{TV} \leq \epsilon \right\}.$$

- This is how long we need to wait until we are "ε close" to stationarity .
- We often take $\varepsilon = 1/4$, indeed let $t_{mix} := \tau(1/4)$.
- For any fixed $0 < \epsilon < \delta < 1/2$ we have

$$\tau(\epsilon) \leq \left\lceil \frac{\ln \epsilon}{\ln 2\delta} \right\rceil \tau(\delta).$$

Thus for any $\epsilon < 1/4$

$$\tau(\epsilon) \leq \left\lceil \log_2 \epsilon^{-1} \right\rceil t_{mix}.$$