Advanced topics in Markov-chain Monte Carlo

Lecture 2:
Diameters and conductances, liftings, path graph
Part 1/2: Theoretical properties

Werner Krauth

ICFP -Master Course Ecole Normale Supérieure, Paris, France

26 January 2022

References

- D. A. Levin, Y. Peres, E. L. Wilmer, "**Markov Chains and Mixing Times**" (American Mathematical Society, 2008) Second edition: http://pages.uoregon.edu/dlevin/MARKOV/mcmt2e.pdf
- A. Sinclair, M. Jerrum **Approximate Counting, Uniform Generation and Rapidly Mixing Markov Chains** Information and Computation 82, 93-133 (1989) (We only need Lemma 3.3, and its proof) https://people.eecs.berkeley.edu/~sinclair/approx.pdf
- F. Chen, L. Lovasz, I. Pak, **Lifting Markov Chains to Speed up Mixing.** Proceedings of the 17th Annual ACM Symposium on Theory of Computing, 275 (1999) http://www.math.ucla.edu/~pak/papers/stoc2.pdf

Total variation distance, mixing time (reminder)

- Total variation distance:

$$||\pi^{\{t\}} - \pi||_{\text{TV}} = \max_{A \subset \Omega} |\pi^{\{t\}}(A) - \pi(A)| = \frac{1}{2} \sum_{i \in \Omega} |\pi_i^{\{t\}} - \pi_i|.$$

- (Above) first eq.: definition; second eq.: (tiny) theorem
- Distance:

$$d(t) = \max_{\pi^{\{0\}}} ||\pi^{\{t\}}(\pi^{\{0\}}) - \pi||_{\text{TV}}$$

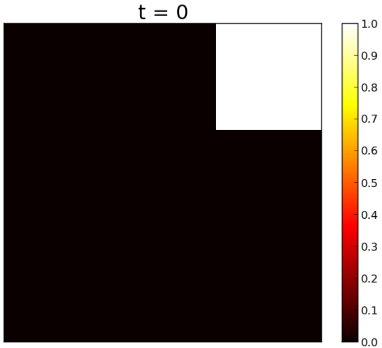
- Mixing time:

$$t_{\text{mix}}(\epsilon) = \min\{t : d(t) \leq \epsilon\}$$

- Usually $\epsilon = 1/4$ is taken (arbitrary, must be smaller than $\frac{1}{2}$): $t_{\text{mix}} = t_{\text{mix}}(1/4)$

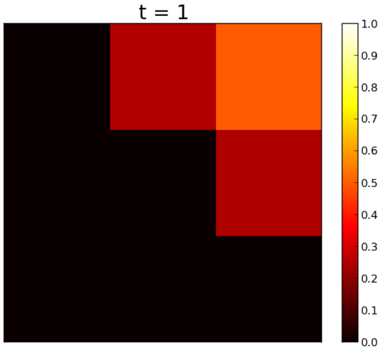
Mixing (reminder)

- Distribution $\pi^{t=0}$ (starting from upper right)



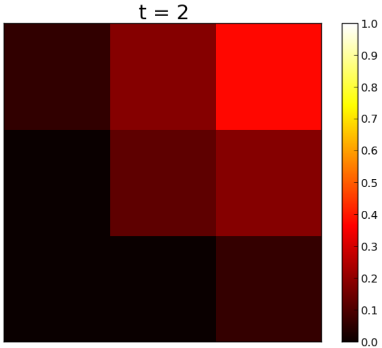
Mixing

- Distribution $\pi^{t=1}$ (starting from upper right)



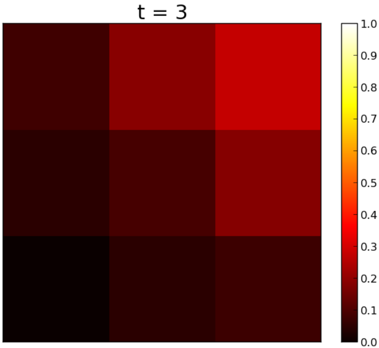
Mixing

- Distribution $\pi^{t=2}$ (starting from upper right)



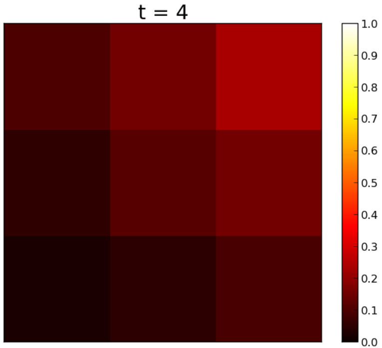
Mixing

- Distribution $\pi^{t=3}$ (starting from upper right)



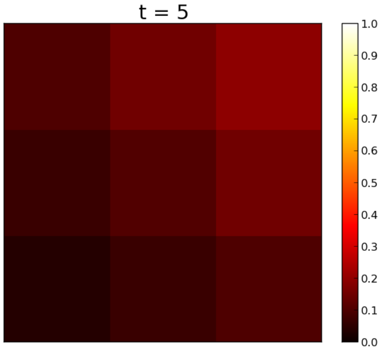
Mixing

- Distribution $\pi^{t=4}$ (starting from upper right)



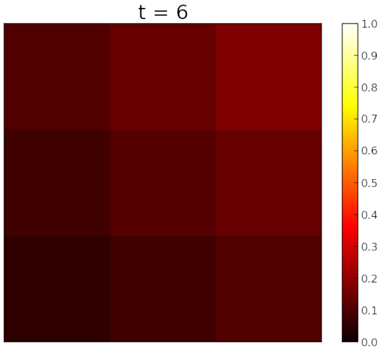
Mixing

- Distribution $\pi^{t=5}$ (starting from upper right)



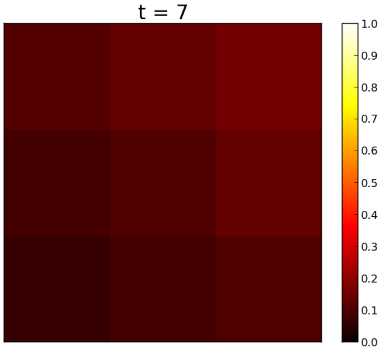
Mixing

- Distribution $\pi^{t=6}$ (starting from upper right)



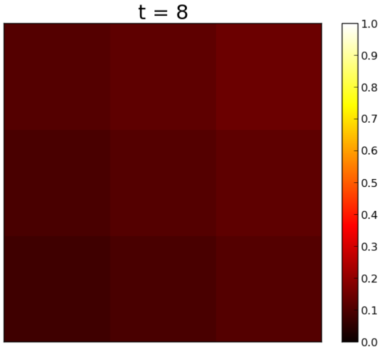
Mixing

- Distribution $\pi^{t=7}$ (starting from upper right)



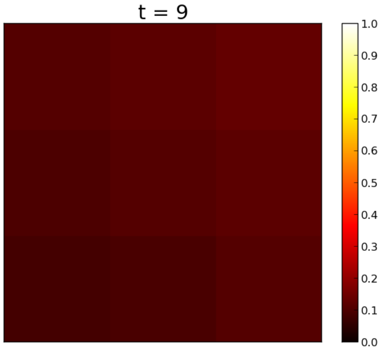
Mixing

- Distribution $\pi^{t=8}$ (starting from upper right)



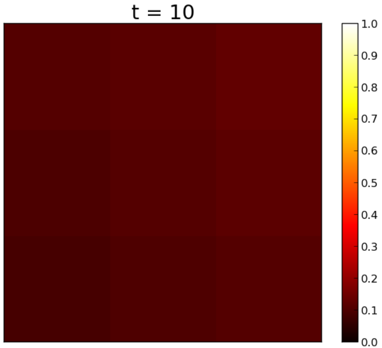
Mixing

- Distribution $\pi^{t=9}$ (starting from upper right)



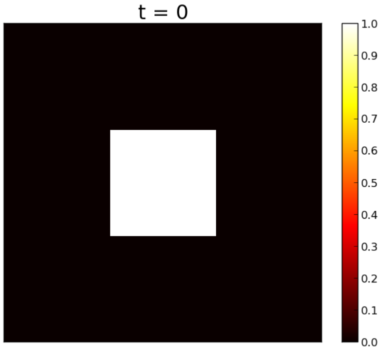
Mixing

- Distribution $\pi^{t=10}$ (starting from upper right)



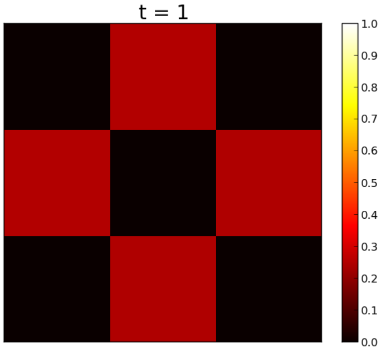
Mixing

- Distribution $\pi^{t=0}$ (starting from center)



Mixing

- Distribution $\pi^{t=1}$ (starting from center)



Diameter bounds, conductance

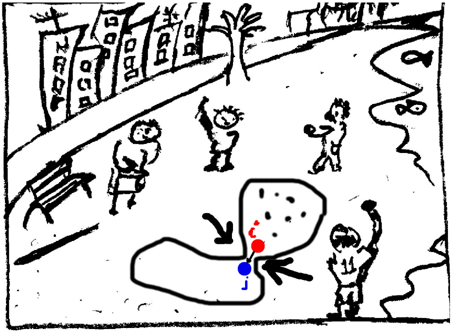
- Graph diameter L: minimum number of moves to travel between any $i, j \in \Omega$.
- NB: $L = 5$ for 3×3 pebble game.
- Diameter bound: or any $\epsilon < 1/2$, trivially satisfies

$$t_{\text{mix}} \geq L/2.$$

- Conductance (bottleneck ratio):

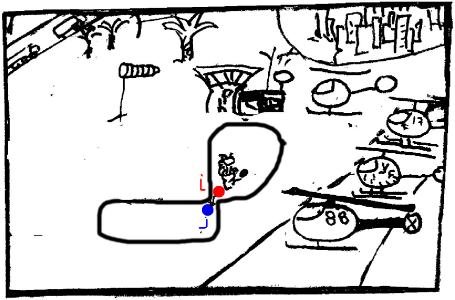
$$\Phi \equiv \min_{S \subset \Omega, \pi_S \leq \frac{1}{2}} \frac{\mathcal{F}_{S \to \overline{S}}}{\pi_S} = \min_{S \subset \Omega, \pi_S \leq \frac{1}{2}} \frac{\sum_{i \in S, j \notin S} \pi_i P_{ij}}{\pi_S}.$$

Direct sampling with bottleneck



NB: ... reaches a boundary site $i \in S$ with probability π_i / π_S

Direct sampling with bottleneck



NB: ... reaches a boundary site $i \in S$ less than with π_i/π_S

Conductance and correlations

Remember:

$$\Phi \equiv \min_{S \subset \Omega, \pi_S \le \frac{1}{2}} \frac{\mathcal{F}_{S \to \bar{S}}}{\pi_S} = \min_{S \subset \Omega, \pi_S \le \frac{1}{2}} \frac{\sum_{i \in S, j \notin S} \pi_i P_{ij}}{\pi_S}.$$

- Reversible Markov chains:

$$\frac{1}{\Phi} \le \tau_{\text{corr}} \le \frac{8}{\Phi^2}$$

(second relation see Sinclair & Jerrum, Lemma (3.3) (p 15-17))

- Arbitrary Markov chain (see Chen et al):

$$\frac{1}{4\Phi} \le \mathcal{A} \le \frac{20}{\Phi^2},$$

(set time: Expectation of $\max_S (t_S \times \pi_S)$ from equilibrium)

NB: One bottleneck, not many. Lower *and* upper bound.

Conductance and mixing

$$\Phi \equiv \min_{S\subset\Omega,\pi_S\leq\frac{1}{2}} \frac{\mathcal{F}_{S\to\overline{S}}}{\pi_S} = \min_{S\subset\Omega,\pi_S\leq\frac{1}{2}} \frac{\sum_{i\in S,j\notin S}\pi_i P_{ij}}{\pi_S}.$$

- Mixing-time bounds:

$$\frac{\text{const}}{\Phi} \leq t_{\text{mix}} \leq \frac{\text{const}'}{\Phi^2}\log\left(1/\pi_0\right)$$

  const and const′ depend on whether reversible or non-reversible. π_0: smallest weight (see Chen et al 1999).

NB: One bottleneck, not many. Lower *and* upper bound.
NNB: Conductance: more general than transition matrices

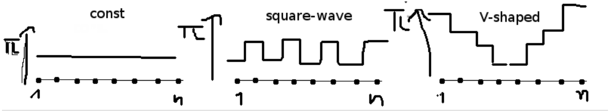
Metropolis algorithm on path graph

- path graph $P_n = (\Omega_n, E_n) \dots$
- vertices $\Leftrightarrow$ sample space $\Omega_n = \{1, \dots, n\}$
- edges $E_n = \{(1,2), \dots, (n-1, n)\} \Leftrightarrow$ non-zero P_{ij}
- one-d n-site lattice without pbc.
- stationary distribution $\pi = \{\pi_1, \dots, \pi_n\} \Leftrightarrow$ GBC.

Metropolis algorithm:

1. From vertex $i \in \Omega$ choose $j = i \pm 1$ with probability $1/2$
2. if $j \notin \Omega$: stay at i
3. otherwise: accept $i \to j$ with probability $\min(1, \pi_j/\pi_i)$
4. else stay at i.

Path-graph stationary distributions and capacities



- Constant: $\pi_i = \frac{1}{n} \; \forall i \in \Omega$.
  Conductance $\Phi = \mathcal{O}(1/n)$.
- Square wave: $\pi_{2k-1} = \frac{2}{3n}$, $\pi_{2k} = \frac{4}{3n}$.
  Conductance $\Phi = \frac{2}{3n}$ (for $n \to \infty$).
- V-shaped: $\pi_i = \text{const}|\frac{n+1}{2} - i| \; \forall i \in \Omega$, where const $= \frac{4}{n^2}$.
  Conductance $\Phi = \frac{2}{n^2}$.

NB: Graph diameter n.
NNB: π normalized.
NNNB: Bottleneck between $i = \frac{n}{2}$ and $j = \frac{n}{2} + 1$.