Mixing times of Markov chains

Jonathan Hermon

January 7, 2020

General information

- When and where: Tue and Thu 2-3:30, Math 126.
- My website and email: jhermon@math.ubc.ca, www.math.ubc.ca/∼jhermon/.
- Office hours: Math annex 1224, to be determined next week. (Contact me if you want to meet this week.)

Final grade

- 100% based on 4 homework assignments.
- There will probably be additional exercises that I will not grade. I will provide solutions to some of them.

About me

- I am new faculty.
- My research is mainly on mixing times of Markov chains.
- Only a small part of the course will be about my own research.

Ask questions!

- I expect you to ask questions!
- If what I am saying or my handwriting is unclear, or if I have a typo, let me know! You will be doing yourselves and your classmates a service.

Literature

- I will follow the lecture notes of Perla Sousi https://www.dpmms.cam.ac.uk/∼ps422/mixing.html. I will expand upon them throughout the semester. They will be available on my website.
- The presentation will be close to the 2nd edition of the book by Levin and Peres (with contributions by Wilmer) - available on Levin's website - very accessible, and has a lot of examples.

Literature

- I will follow the lecture notes of Perla Sousi https://www.dpmms.cam.ac.uk/∼ps422/mixing.html. I will expand upon them throughout the semester. They will be available on my website.
- The presentation will be close to the 2nd edition of the book by Levin and Peres (with contributions by Wilmer) - available on Levin's website - very accessible, and has a lot of examples.
- Another good book (with a heavy analytic flavour) is R. Montenegro and P. Tetali, Mathematical aspects of mixing times in Markov chains. Foundations and Trends in Theoretical Computer Science: Vol. 1: No. 3, pp 237-354, 2006. Available online at Prasad Tetali's website.
- D. Aldous and J. Fill, Reversible Markov Chains and Random Walks on Graphs. Unfinished manuscript, available online at David Aldous' website.

What are Markov chains?

- A sequence of random variables $(X_n)_{n\geq 0}$ taking values in a **state space** E is called a **Markov chain** if for all $x_0, \ldots, x_n \in E$ such that $\mathbb{P}(X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) > 0$ we have

$$\begin{aligned}\mathbb{P}(X_n = x_n \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}) &= \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) \\ &= P(x_{n-1}, x_n).\end{aligned}$$

  In other words, **the future of the process is independent of the past given the present**.

What are Markov chains?

- A sequence of random variables $(X_n)_{n\geq 0}$ taking values in a **state space** E is called a **Markov chain** if for all $x_0, \dots, x_n \in E$ such that $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) > 0$ we have

$$\mathbb{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = x_n \mid X_{n-1} = x_{n-1})$$
$$= P(x_{n-1}, x_n).$$

In other words, **the future of the process is independent of the past given the present**.

- A Markov chain is defined by its **transition matrix** P given by

$$P(i,j) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad \forall\, i,j \in E, n \in \mathbb{N}.$$

Note that P^n is the time n transition probabilities

$$P^n(i,j) = \mathbb{P}(X_n = j \mid X_0 = i) =: \mathbb{P}_i[X_n = j] \quad \forall\, i,j \in E.$$

- Under mild conditions there exists a unique invariant distribution π, and the law of the chain at time n converges to π as $n \to \infty$. That is, for all x, y

$$\lim_{n\to\infty} P^n(x,y) = \pi(y).$$

- In this course, the state space will almost always be finite. If the chain is periodic, we fix this by replacing P with its δ **lazy version** $\delta I + (1-\delta)P$, or by working in continuous-time.

- Under mild conditions there exists a unique invariant distribution π, and the law of the chain at time n converges to π as $n \to \infty$. That is, for all x, y

$$\lim_{n\to\infty} P^n(x,y) = \pi(y).$$

- But how quickly does this occur? This does not say anything about the rate of convergence!

  To talk about rate of convergence we need to pick a metric.

- To quantify this, we define the ε-**total-variaiton mixing time** to be

$$t_{\mathrm{mix}}(\varepsilon) := \inf\{n : \max_x \|P^n(x,\cdot) - \pi\|_{\mathrm{TV}} \le \varepsilon\},$$

  where

$$\|\mu - \nu\|_{\mathrm{TV}} := \frac{1}{2}\sum_x |\mu(x) - \nu(x)|.$$

- We will see that $t_{\mathrm{mix}}(2^{-k}) \le k t_{\mathrm{mix}}(1/4)$, and so the choice of $\varepsilon < 1/2$ is not important.
- So we can define the mixing time as $t_{\mathrm{mix}} := t_{\mathrm{mix}}(1/4)$.

- From a theoretical perspective, it is natural to look at a sequence of Markov chains and study the asymptotics of the mixing times for a sequence of Markov chains of increasing size.

- From a theoretical perspective, it is natural to look at a sequence of Markov chains and study the asymptotics of the mixing times for a sequence of Markov chains of increasing size.
- We will see that this is a natural also from the perspective of applications.
- The theory of mixing times is rich and has connections to other areas of mathematics, like statistical mechanics, combinatorics and representation theory.

- From a theoretical perspective, it is natural to look at a sequence of Markov chains and study the asymptotics of the mixing times for a sequence of Markov chains of increasing size.
- We will see that this is a natural also from the perspective of applications.
- The theory of mixing times is rich and has connections to other areas of mathematics, like statistical mechanics, combinatorics and representation theory.
- Nevertheless, it is an area with a relatively low entry threshold.

Card shuffling

- Consider a deck of n cards. At each step pick two cards at random and swap them w.p. $1/2$

  (or, e.g., at each step pick a random card and move it to the top of the deck).

- How many shuffles are required until the deck is shuffled well?

- We will see that some card shuffling schemes can be analyzed using a powerful technique called **coupling**.

Card shuffling

- Consider a deck of n cards. At each step pick two cards at random and swap them w.p. $1/2$

  (or, e.g., at each step pick a random card and move it to the top of the deck).

- How many shuffles are required until the deck is shuffled well?

- We will see that some card shuffling schemes can be analyzed using a powerful technique called **coupling**. This technique is indispensible part of the toolkit of anyone working in discrete-probability.

- Surprisingly, another useful technique is to use representation theory.

Random walks on graphs

- Simple random walk on a sequence of graphs. E.g., the n-cycle (at each step stay put w.p. $1/2$ and otherwise, move left or right with equal probability). What is the order mixing time?

Random walks on graphs

- Simple random walk on a sequence of graphs. E.g., the n-cycle (at each step stay put w.p. $1/2$ and otherwise, move left or right with equal probability). What is the order mixing time?

- Hint: if the state space is $\{0, 1, \ldots, n-1\}$ (this of $n = 0$) then $P^t(x, \cdot)$ is roughly the law of $[Y]$ mod n, where $Y \sim N(0, t/2)$, where $[a]$ is the integer closest to a.

Random walks on graphs

- Simple random walk on a sequence of graphs. E.g., the n-cycle (at each step stay put w.p. $1/2$ and otherwise, move left or right with equal probability). What is the order mixing time?
- Hint: if the state space is $\{0, 1, \ldots, n-1\}$ (this of $n = 0$) then $P^t(x, \cdot)$ is roughly the law of $[Y]$ mod n, where $Y \sim N(0, t/2)$, where $[a]$ is the integer closest to a.
- Answer: By the (local) CLT it is $n^2 \asymp (\text{diameter})^2$.

Random walks on graphs

- Simple random walk on a sequence of graphs. E.g., the n-cycle. What is the order mixing time?
  Answer: By the (local) CLT it is $n^2 \asymp (\text{diameter})^2$.

- As we shall see, this is the case for any sequence of vertex-transitive graphs of polynomial growth.

  ($G = (V, E)$ is vertex-transitive if for all $x, y \in V$ there is an automorphism mapping x to y.)

Random walks on graphs

- Simple random walk on a sequence of graphs. E.g., the n-cycle. What is the order mixing time?
  Answer: By the (local) CLT it is $n^2 \asymp (\text{diameter})^2$.
- As we shall see, this is the case for any sequence of vertex-transitive graphs of polynomial growth.

  ($G = (V, E)$ is vertex-transitive if for all $x, y \in V$ there is an automorphism mapping x to y.)
- To do so, we will develop general technique to transform isoperimetric estimates into bounds on mixing times.

Applications to sampling and counting

- Given a target distribution π, often a complicated one for which we have some local (Gibbs measure) or recursive description, there are standard ways of setting up a Markov chain whose stationary distribution is π.
- Hence, if we want to sample from a distribution close to π, we just run such a chain. We need to know the mixing time of the chain!

Applications to sampling and counting

- Given a target distribution π, often a complicated one for which we have some local (Gibbs measure) or recursive description, there are standard ways of setting up a Markov chain whose stationary distribution is π.
- Hence, if we want to sample from a distribution close to π, we just run such a chain. We need to know the mixing time of the chain!
- This can be used to simulate complicated distributions arising e.g. in Bayesian statistics or statistical mechanics, and approximate certain quantities that calculating their exact value is a hard problem.
- As we will see, this is used to estimate the size of complicated combinatorial sets (e.g., number of q-colorings of a large graph; this can be done even when exact counting is "hard").

Applications to sampling and counting

- Given a target distribution π, often a complicated one for which we have some local (Gibbs measure) or recursive description, there are standard ways of setting up a Markov chain whose stationary distribution is π.
- Hence, if we want to sample from a distribution close to π, we just run such a chain. We need to know the mixing time of the chain!
- This can be used to simulate complicated distributions arising e.g. in Bayesian statistics or statistical mechanics, and approximate certain quantities that calculating their exact value is a hard problem.
- As we will see, this is used to estimate the size of complicated combinatorial sets (e.g., number of q-colorings of a large graph; this can be done even when exact counting is "hard").
- The mixing time of an auxiliary Markov chain is the main component in the running time of the randomized algorithms.

Connections with statistical mechanics

- In statistical mechanics a model on a graph $G = (V, E)$ is just a distribution π over S^V, where S is some finite set.

- A generic way of introducing a dynamics (Markov chain) whose invariant distribution is π is by picking a site x at random and updating its spin ($=$ value) according to the distribution π conditioned on the values of the spins of the rest of the vertices. This is called a **Glauber dynamics**.

Connections with statistical mechanics

- In statistical mechanics a model on a graph $G = (V, E)$ is just a distribution π over S^V, where S is some finite set.

- A generic way of introducing a dynamics (Markov chain) whose invariant distribution is π is by picking a site x at random and updating its spin ($=$ value) according to the distribution π conditioned on the values of the spins of the rest of the vertices. This is called a **Glauber dynamics**.
- This can be simmulated efficiently when π has a local description, meaning that the update probabilities depend only on the spins of the neighbors of the chosen vertex.

Connections with statistical mechanics

- In statistical mechanics a model on a graph $G = (V, E)$ is just a distribution π over S^V, where S is some finite set.
- A generic way of introducing a dynamics (Markov chain) whose invariant distribution is π is by picking a site x at random and updating its spin (= value) according to the distribution π conditioned on the values of the spins of the rest of the vertices. This is called a **Glauber dynamics**.
- This can be simmulated efficiently when π has a local description, meaning that the update probabilities depend only on the spins of the neighbors of the chosen vertex. Examples:
- q-colorings of a graph $G = (V, E)$: Pick v at random and update its color from the uniform distribution over allowed colors. Here $S := \{1, \dots, q\}$ and π is the uniform distribution on all proper q-colorings.

Connections with statistical mechanics

- In statistical mechanics a model on a graph $G = (V, E)$ is just a distribution π over S^V, where S is some finite set.
- A generic way of introducing a dynamics (Markov chain) whose invariant distribution is π is by picking a site x at random and updating its spin (= value) according to the distribution π conditioned on the values of the spins of the rest of the vertices. This is called a **Glauber dynamics**.
- This can be simmulated efficiently when π has a local description, meaning that the update probabilities depend only on the spins of the neighbors of the chosen vertex. Examples:
- q-colorings of a graph $G = (V, E)$: Pick v at random and update its color from the uniform distribution over allowed colors. Here $S := \{1, \ldots, q\}$ and π is the uniform distribution on all proper q-colorings.
- Ising model: A model for ferromagnetism. Here $S = \{\pm 1\}$ and $\pi(\sigma) = \frac{1}{Z(\beta)} \exp\left(\beta \sum_{uv \in E} \sigma(u)\sigma(v)\right)$.
- Here the mixing time corresponds to the time it takes the system to relax to equilibrium. It depends on the inverse temperature β, and may exhibit a phase transition.

Particle systems - the interchange process

- Let $G = (V, E)$ be a connected graph of size n.
- Consider the model in which we have n distinct particles, one at each site.
- At each step we pick a random edge xy and swap the particles currently at x and y.

Particle systems - the interchange process

- Let $G = (V, E)$ be a connected graph of size n.
- Consider the model in which we have n distinct particles, one at each site.
- At each step we pick a random edge xy and swap the particles currently at x and y.
- We will develop general comparison techniques that would allow us to reduce the problem to the case that G is the complete graph.

- A sequence of MCs exhibits (TV) **cutoff** if the ε-mixing time is asymptotically indep. of ε:

$$\lim_{n\to\infty} t_{\mathrm{mix}}^{(n)}(\varepsilon)/t_{\mathrm{mix}}^{(n)}(1-\varepsilon)=1, \quad \forall\, 0<\varepsilon<1. \tag{1}$$

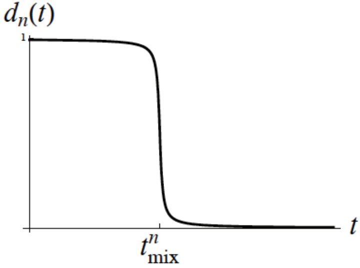

Figure: The worst case TV distance of the nth chain, $d_n(t) := \max_x \|P^t(x,\cdot)-\pi\|_{\mathrm{TV}}$ as a function of t when cutoff occurs.

Figure: David Aldous and Persi Diaconis - the founders of the modern study of Markov chains.

- In 86 they coined the term cutoff.
- While it appears that cutoff is more the norm than the exception, it is extremely challenging to prove, and only relatively few cases are completely understood.

Figure: David Aldous and Persi Diaconis

- We will prove a necessary and sufficient condition for cutoff.
- We will transform it into a simple spectral condition for chains for which the graph supporting the transition probabilities is a tree.

Figure: David Aldous and Persi Diaconis

- We will prove a necessary and sufficient condition for cutoff.
- We will transform it into a simple spectral condition for chains for which the graph supporting the transition probabilities is a tree.
- We will prove cutoff for some card shuffling schemes and for random walk on some random graphs.

Transition matrix and continuouos-time

- We sometimes work with continuous-time Markov chains.
- Consult Levin-Peres Ch. 20 for more details.
- In the finite state space setup, cts-time chains can be described as follows: When at state x wait $\mathrm{Exp}(c)$ time units, and then move to state y w.p. $P(x, y)$. (Possibly $y = x$!)

Continuouos-time Markov chains

- Continuous-time Markov chains: When at state x wait $\mathrm{Exp}(c)$ time units, and then move to state y (possibly $y = x!$) w.p. $P(x, y)$.
  The time t transition probabilities are given by the matrix
  $P_t := e^{ct(P-I)} = \sum_{n \geq 0} \mathbb{P}[\mathrm{Poisson}(ct) = n] P^n$, where $e^Q := \sum_{n \geq 0} Q^n / n!$.

Continuouos-time Markov chains

- Continuous-time Markov chains: When at state x wait $\mathrm{Exp}(c)$ time units, and then move to state y (possibly $y = x$!) w.p. $P(x, y)$.
  The time t transition probabilities are given by the matrix
  $P_t := e^{ct(P-I)} = \sum_{n \geq 0} \mathbb{P}[\mathrm{Poisson}(ct) = n] P^n$, where $e^Q := \sum_{n \geq 0} Q^n / n!$.

  Thus for all $x, y \in E$

$$P_t(x, y) = \mathbb{P}[X_t = y \mid X_0 = x] = \sum_{n \geq 0} \mathbb{P}[\mathrm{Poisson}(ct) = n] P^n(x, y).$$

Continuouos-time Markov chains

- Continuous-time Markov chains: When at state x wait $\mathrm{Exp}(c)$ time units, and then move to state y (possibly $y = x$!) w.p. $P(x, y)$.
  The time t transition probabilities are given by the matrix $P_t := e^{ct(P-I)} = \sum_{n \geq 0} \mathbb{P}[\mathrm{Poisson}(ct) = n] P^n$, where $e^Q := \sum_{n \geq 0} Q^n / n!$.

  Thus for all $x, y \in E$

$$P_t(x, y) = \mathbb{P}[X_t = y \mid X_0 = x] = \sum_{n \geq 0} \mathbb{P}[\mathrm{Poisson}(ct) = n] P^n(x, y).$$

- We say that P is the **jump matrix** (of the cts-time chain), and that $Q := c(P - I)$ is its infinitesimal (Markov) **generator**.
- The name comes from the fact that $\frac{d}{dt} P_t = Q P_t = P_t Q$ and thus $\frac{d}{dt} P_t(x, y) \mid_{t=0} = Q(x, y)$.

Irreducibility, aperiodicity, recurrence

- A Markov chain is called **irreducible** if for all $x, y \in E$ there exists $n \geq 0$ such that $P^n(x, y) > 0$.
- A Markov chain is called **aperiodic**, if for all x we have $\mathrm{g.c.d.}\{n \geq 1 : P^n(x, x) > 0\} = 1$.

Irreducibility, aperiodicity, recurrence

- A Markov chain is called **irreducible** if for all $x, y \in E$ there exists $n \geq 0$ such that $P^n(x, y) > 0$.

- A Markov chain is called **aperiodic**, if for all x we have $\mathrm{g.c.d.}\{n \geq 1 : P^n(x, x) > 0\} = 1$.

- A Markov chain is **recurrent** if $\mathbb{P}_x[T_x^+ < \infty] = 1$ for all $x \in E$ and is transient otherwise, where $T_x^+ := \inf\{n > 0 : X_n = x\}$.
- It is **positively recurrent** if $\mathbb{E}_x[T_x^+] < \infty$ for all x.

Irreducibility, aperiodicity, recurrence

- A Markov chain is called **irreducible** if for all $x, y \in E$ there exists $n \geq 0$ such that $P^n(x, y) > 0$.
- A Markov chain is called **aperiodic**, if for all x we have $\text{g.c.d.}\{n \geq 1 : P^n(x, x) > 0\} = 1$.
- A Markov chain is **recurrent** if $\mathbb{P}_x[T_x^+ < \infty] = 1$ for all $x \in E$ and is transient otherwise, where $T_x^+ := \inf\{n > 0 : X_n = x\}$.
- It is **positively recurrent** if $\mathbb{E}_x[T_x^+] < \infty$ for all x.
- Finite state space + irreducibility $\implies$ positive recurrence.

Stationary = invariant = equilibrium distribution

- We call π an **invariant distribution** if $\pi P = \pi$.

  (Recall: For a signed measure (row vector) σ,

  $$(\sigma P^n)(x) := \sum_y \sigma(y) P^n(y,x).$$

  If σ is a distribution and $X_0 \sim \sigma$, then $X_n \sim \sigma P^n$.)

- This means that if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \in \mathbb{N}$.

Stationary = invariant = equilibrium distribution

- We call π an **invariant distribution** if $\pi P = \pi$. This means that if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \in \mathbb{N}$.
- For a continuous-time chain $(X_t)_{t \geq 0}$ with jump matrix P, the condition $\pi P = \pi$ is equivalent to $\pi P_t = \pi$ for all $t \geq 0$ (check!) and then $X_t \sim \pi$ for all $t \in \mathbb{R}_+$ whenever $X_0 \sim \pi$.

Stationary = invariant = equilibrium distribution

- We call π an **invariant distribution** if $\pi P = \pi$. This means that if $X_0 \sim \pi$, then $X_n \sim \pi$ for all $n \in \mathbb{N}$.
- For a continuous-time chain $(X_t)_{t \geq 0}$ with jump matrix P, the condition $\pi P = \pi$ is equivalent to $\pi P_t = \pi$ for all $t \geq 0$ (check!) and then $X_t \sim \pi$ for all $t \in \mathbb{R}_+$ whenever $X_0 \sim \pi$.
- If the chain is irreducible and positively reccurent, it has a unique invariant distribution π given by

$$\pi(x) := \frac{1}{\mathbb{E}_x[T_x^+]}$$

(in continuous-time $\pi(x) := \frac{1}{c(1-P(x,x))\mathbb{E}_x[T_x^+]}$).

Stationary = invariant = equilibrium distribution

Consider an irreducible, positively reccurent Markov chain.

- $\pi(x)$ is the asymptotic frequency of the time spent at x:

$$\pi(x) = \lim_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} P^i(y,x).$$

- If the chain is also aperiodic then $\lim_{n\to\infty} \mathbb{P}[X_n = x \mid X_0 = y] = \pi(x)$ for all $x, y \in E$.

- In continuous-time this does not require aperiodicity!

Reversibility

- We say that P is **reversible** wrt π if for all x, y

$$\text{(Detailed balance equation)} \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

This implies that $(\pi P)(x) = \sum_y \pi(y)P(y,x) = \pi(x)\sum_y P(x,y) = \pi(x)$.

Reversibility

- We say that P is **reversible** wrt π if for all x, y

$$\text{(Detailed balance equation)} \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

This implies that $(\pi P)(x) = \sum_y \pi(y)P(y,x) = \pi(x)\sum_y P(x,y) = \pi(x)$.

- Generally, if π is invariant for P, we define the **time reversal** P^* by

$$P^*(x,y) = \frac{\pi(y)}{\pi(x)}P(y,x).$$

This is a transition matrix for which π is also invariant (check!).

Reversibility

- We say that P is **reversible** wrt π if for all x, y

$$\text{(Detailed balance equation)} \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

This implies that $(\pi P)(x) = \sum_y \pi(y)P(y,x) = \pi(x)\sum_y P(x,y) = \pi(x)$.

- Generally, if π is invariant for P, we define the **time reversal** P^* by

$$P^*(x,y) = \frac{\pi(y)}{\pi(x)}P(y,x).$$

This is a transition matrix for which π is also invariant (check!).

- It is the dual operator of P wrt to the inner-product induced by π, $\langle f, g\rangle_\pi := \sum_x \pi(x)f(x)g(x) = \mathbb{E}_\pi[fg]$. That is,

$$\langle Pf, g\rangle_\pi = \langle f, P^*g\rangle_\pi.$$

Recall: for a function f : State-space $\to \mathbb{R}$ (column vector)

$$(P^n f)(x) := \sum_y P^n(x,y)f(y) = \mathbb{E}[f(X_n) \mid X_0 = x] =: \mathbb{E}_x[f(X_n)].$$

Reversibility

- We say that P is **reversible** wrt π if for all x, y

$$\text{(Detailed balance equation)} \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

- Generally, if π is invariant for P, we define the **time reversal** P^* by

$$P^*(x,y) = \frac{\pi(y)}{\pi(x)}P(y,x).$$

- Hence P is reversibile iff $P = P^*$.
- By induction: for all $a_0, \ldots, a_n$ we have

$$\pi(a_0)P(a_0,a_1)\cdots P(a_{n-1},a_n) = \pi(a_n)P^*(a_n,a_{n-1})\cdots P^*(a_1,a_0).$$

Reversibility

- We say that P is **reversible** wrt π if for all x, y

$$\text{(Detailed balance equation)} \quad \pi(x)P(x,y) = \pi(y)P(y,x).$$

- Generally, if π is invariant for P, we define the **time reversal** P^* by

$$P^*(x,y) = \frac{\pi(y)}{\pi(x)} P(y,x).$$

- Hence P is reversibile iff $P = P^*$.
- By induction: for all $a_0, \ldots, a_n$ we have

$$\pi(a_0)P(a_0,a_1)\cdots P(a_{n-1},a_n) = \pi(a_n)P^*(a_n,a_{n-1})\cdots P^*(a_1,a_0).$$

- Reversibility means that at equilibrium the chain looks the same forward and backwards: $(X_0, \ldots, X_N)$ has the same distribution as $(X_N, \ldots, X_0)$, when $X_0 \sim \pi$.

Reversibility = weighted random walks

- Call a chain reversible if its transition matrix (jump matrix) is reversible.
- Most of the chains that we shall study will be reversible.

Reversibility = weighted random walks

- Call a chain reversible if its transition matrix (jump matrix) is reversible.
- Most of the chains that we shall study will be reversible.
- This is partially due to the fact that the theory is nicer for them, as they admit a spectral-decomposition.
- It is a huge class of chains: it is the collection of weighted random walks on graphs.

Reversibility = weighted random walks

- A finite connected (undirected) graph $G = (V, E)$ and a collection of symmetric positive weights $\mathbf{c} := (c_e : e \in E)$ is called a **network**.

  (Symmetric means $c_{xy} = c_{yx}$; if $xy \notin E$ set $c_{xy} = 0$.)

Reversibility = weighted random walks

- A finite connected (undirected) graph $G = (V, E)$ and a collection of symmetric positive weights $\mathbf{c} := (c_e : e \in E)$ is called a **network**.

  (Symmetric means $c_{xy} = c_{yx}$; if $xy \notin E$ set $c_{xy} = 0$.)

- A **random walk on the network** $(G, \mathbf{c})$
  (or a weighted random walk on G with weights $\mathbf{c}$)
  is the Markov chain with transitions proportional to the weights $\mathbf{c}$:

  $$P(x,y) := c_{xy}/c_x,$$

  where $c_v := \sum_u c_{vu}$ for $v \in V$.

- It is reversible w.r.t. $\pi(x) := \frac{c_x}{\sum_y c_y}$ as

  $$\pi(x)P(x,y) = \frac{c_{xy}}{\sum_y c_y} = \frac{c_{yx}}{\sum_y c_y} = \pi(y)P(x,y).$$

Reversibility = weighted random walks

- A finite connected (undirected) graph $G = (V, E)$ and a collection of symmetric positive weights $\mathbf{c} := (c_e : e \in E)$ is called a **network**.

  (Symmetric means $c_{xy} = c_{yx}$; if $xy \notin E$ set $c_{xy} = 0$.)

- A **random walk on the network** $(G, \mathbf{c})$
  (or a weighted random walk on G with weights $\mathbf{c}$)
  is the Markov chain with transitions proportional to the weights $\mathbf{c}$:

  $$P(x, y) := c_{xy}/c_x,$$

  where $c_v := \sum_u c_{vu}$ for $v \in V$.

- It is reversible w.r.t. $\pi(x) := \frac{c_x}{\sum_y c_y}$ as

  $$\pi(x)P(x, y) = \frac{c_{xy}}{\sum_y c_y} = \frac{c_{yx}}{\sum_y c_y} = \pi(y)P(x, y).$$

- Conversely, if P is reversible, then the chain is a weighted random walk w.r.t. weights $c_{xy} := \pi(x)P(x, y)$, on the graph in which $x \sim y$ iff $P(x, y) > 0$. (Check!)

Reversibility = weighted random walks

- A finite connected (undirected) graph $G = (V, E)$ and a collection of symmetric positive weights $\mathbf{c} := (c_e : e \in E)$ is called a **network**.

- A **random walk on the network** $(G, \mathbf{c})$ is the Markov chain with transitions proportional to the weights $\mathbf{c}$:

$$P(x, y) := c_{xy} / \sum_u c_{xu}.$$

- Taking $c_e = 1$ for all $e \in E$ gives rise to **simple random walk** on G

$$P(x, y) = \frac{\mathbf{1}\{x \sim y\}}{\deg(x)}.$$