Spectral Theory of Finite Markov Chains

Austin Eide

University Of Nebraska – Lincoln

Spring 2020

Markov chains

Definition (Markov Chain)

A Markov chain on state space $\mathcal{X}$ is a sequence of $\mathcal{X}$-valued r.v.'s $(X_0, X_1, \dots)$ satisfying the *Markov property*:

$$\mathbf{P}(X_{t+1} = y | (X_t, \dots, X_0)) = \mathbf{P}(X_{t+1} = y | X_t = x) =: P(x, y)$$

Markov chains

Definition (Markov Chain)

A Markov chain on state space $\mathcal{X}$ is a sequence of $\mathcal{X}$-valued r.v.'s $(X_0, X_1, \dots)$ satisfying the *Markov property*:

$$\mathbf{P}(X_{t+1} = y | (X_t, \dots, X_0)) = \mathbf{P}(X_{t+1} = y | X_t = x) =: P(x, y)$$

A chain is thus entirely described by an initial distribution $\mu_0 \in \mathbb{R}^{|\mathcal{X}|}$ for X_0 and a $|\mathcal{X}| \times |\mathcal{X}|$ row-stochastic matrix P which stores *transition probabilities*.

The transition matrix

The transition matrix

If today's distribution (i.e., the distribution on X_t) is μ_t, then tomorrow's distribution is $\mu_{t+1} = \mu_t P$.

The transition matrix

If today's distribution (i.e., the distribution on X_t) is μ_t, then tomorrow's distribution is $\mu_{t+1} = \mu_t P$.

Given the initial distribution μ_0, inductively we have $\mu_t = \mu_0 P^t$.

The transition matrix

If today's distribution (i.e., the distribution on X_t) is μ_t, then tomorrow's distribution is $\mu_{t+1} = \mu_t P$.

Given the initial distribution μ_0, inductively we have $\mu_t = \mu_0 P^t$.

Note: almost always, we'll think of μ_0 as a point mass on some state $x \in \mathcal{X}$.

Irreducibility & Aperiodicity

Irreducibility & Aperiodicity

For $x, y \in \mathcal{X}$ and $t \geq 0$, $P^t(x,y)$ is the probability of traveling from x to y in t steps.

Irreducibility & Aperiodicity

For $x, y \in \mathcal{X}$ and $t \geq 0$, $P^t(x, y)$ is the probability of traveling from x to y in t steps.

Definition (I. & A.)

A chain is *irreducible* if $\forall$ pairs $x, y \in \mathcal{X}$, $\exists$ integer t with $P^t(x, y) > 0$.

Irreducibility & Aperiodicity

For $x, y \in \mathcal{X}$ and $t \geq 0$, $P^t(x, y)$ is the probability of traveling from x to y in t steps.

Definition (I. & A.)

A chain is *irreducible* if $\forall$ pairs $x, y \in \mathcal{X}$, $\exists$ integer t with $P^t(x, y) > 0$.

A chain is *aperiodic* if

$$\gcd\{t \geq 1 : P^t(x, x) > 0\} = 1.$$

Irreducibility & Aperiodicity

For $x, y \in \mathcal{X}$ and $t \geq 0$, $P^t(x,y)$ is the probability of traveling from x to y in t steps.

Definition (I. & A.)

A chain is *irreducible* if $\forall$ pairs $x, y \in \mathcal{X}$, $\exists$ integer t with $P^t(x,y) > 0$.

A chain is *aperiodic* if

$$\gcd\{t \geq 1 : P^t(x,x) > 0\} = 1.$$

(For example, a "bipartite" chain *is* periodic, since then the above quantity is 2.)

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Proof.

I + A $\implies$ $P^t > 0$ for all t sufficiently large.

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Proof.

I + A $\implies$ $P^t > 0$ for all t sufficiently large.

Easy to show that $\sigma(P) \leq 1$, and that 1 is an eigenvalue.

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Proof.

I + A $\implies$ $P^t > 0$ for all t sufficiently large.

Easy to show that $\sigma(P) \leq 1$, and that 1 is an eigenvalue.

Thus, by Perron-Frobenius (and a corollary thereof):

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Proof.

I + A $\implies$ $P^t > 0$ for all t sufficiently large.

Easy to show that $\sigma(P) \leq 1$, and that 1 is an eigenvalue.

Thus, by Perron-Frobenius (and a corollary thereof):

- P has a unique, strictly positive left eigenvector π with eigenvalue 1—the *stationary distribution* of P

Perron-Frobenius

Theorem

If P is irreducible and aperiodic, then $\exists!$ distribution π such that $\pi P = \pi$, and moreover for any μ_0 we have $\mu_0 P^t \to \pi$.

Proof.

I + A $\implies$ $P^t > 0$ for all t sufficiently large.

Easy to show that $\sigma(P) \leq 1$, and that 1 is an eigenvalue.

Thus, by Perron-Frobenius (and a corollary thereof):

- P has a unique, strictly positive left eigenvector π with eigenvalue 1—the *stationary distribution* of P
- For any distribution μ_0 on $\mathcal{X}$, $\mu_0 P^t \to \pi$

□

Convergence in...?

Usually,

Convergence in...?

Usually,

Definition (Total Variation Distance)

For probability distributions $\mu, \nu \in \mathbb{R}^{|\mathcal{X}|}$ on $\mathcal{X}$, define

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \frac{1}{2} \|\mu - \nu\|_1 .$$

Convergence in...?

Usually,

Definition (Total Variation Distance)

For probability distributions $\mu, \nu \in \mathbb{R}^{|\mathcal{X}|}$ on $\mathcal{X}$, define

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| = \frac{1}{2} \|\mu - \nu\|_1 .$$

Equivalent to $\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$.

Mixing times

For $x \in \mathcal{X}$, let $\mu_x \in \mathbb{R}^{|\mathcal{X}|}$ be the point-mass distribution at x.

Mixing times

For $x \in \mathcal{X}$, let $\mu_x \in \mathbb{R}^{|\mathcal{X}|}$ be the point-mass distribution at x.

Define $d(t) := \max_x \left\|\mu_x P^t - \pi\right\|_{TV}$.

Mixing times

For $x \in \mathcal{X}$, let $\mu_x \in \mathbb{R}^{|\mathcal{X}|}$ be the point-mass distribution at x.

Define $d(t) := \max_x \left\| \mu_x P^t - \pi \right\|_{TV}$.

For $\varepsilon > 0$, define $t_{\text{mix}}(\varepsilon) = \min\{t \in \mathbb{Z}_{\geq 0} : d(t) < \varepsilon\}$.

Mixing times

For $x \in \mathcal{X}$, let $\mu_x \in \mathbb{R}^{|\mathcal{X}|}$ be the point-mass distribution at x.

Define $d(t) := \max_x \left\| \mu_x P^t - \pi \right\|_{TV}$.

For $\varepsilon > 0$, define $t_{\text{mix}}(\varepsilon) = \min\{t \in \mathbb{Z}_{\geq 0} : d(t) < \varepsilon\}$.

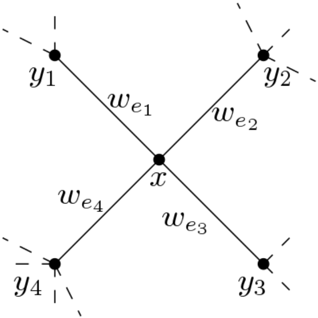
$t_{\text{mix}}(\varepsilon)$ is the (ε)-*mixing time* of the chain.

Reversibility & r.w.'s on graphs

Henceforth, we'll restrict attention to chains which are *random walks on edge-weighted graphs*

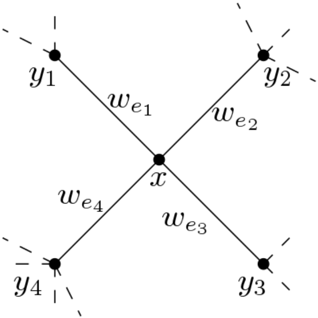
Reversibility & r.w.'s on graphs

Henceforth, we'll restrict attention to chains which are *random walks on edge-weighted graphs*



Reversibility & r.w.'s on graphs

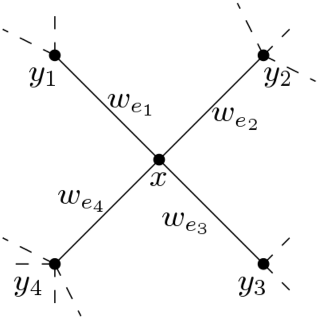
Henceforth, we'll restrict attention to chains which are *random walks on edge-weighted graphs*



Proceed by $P(x, y_i) = \frac{w_i}{\sum w_i}$.

Reversibility & r.w.'s on graphs

Henceforth, we'll restrict attention to chains which are *random walks on edge-weighted graphs*



Proceed by $P(x, y_i) = \frac{w_i}{\sum w_i}$. What do we get when all edges have weight 1?

Reversibility & r.w.'s on graphs

Reversibility is the core property relating general chains to r.w.'s on graphs.

Reversibility & r.w.'s on graphs

Reversibility is the core property relating general chains to r.w.'s on graphs.

Definition (Reversibility)

A Markov chain is *reversible* with respect to stationary distribution π if $\forall x, y \in \mathcal{X}$,

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

Reversibility & r.w.'s on graphs

Reversibility is the core property relating general chains to r.w.'s on graphs.

Definition (Reversibility)

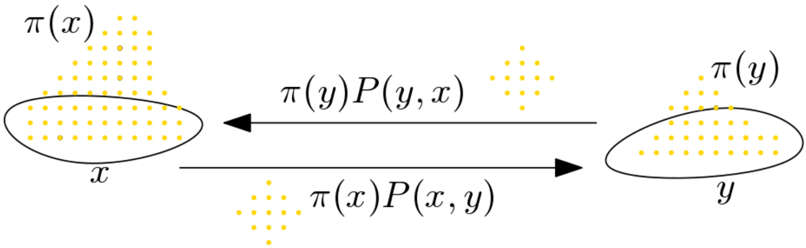
A Markov chain is *reversible* with respect to stationary distribution π if $\forall x, y \in \mathcal{X}$,

$$\pi(x)P(x,y) = \pi(y)P(y,x).$$

$$\{\text{reversible chains } P\} \iff \{\text{weighted graphs}\}$$

$$P \mapsto G_P \text{ where } V(G_P) = \mathcal{X},$$
$$\text{edge weights } \pi(x)P(x,y) = \pi(y)P(y,x)$$

Reversibility aka "Detailed Balance"



The π-inner product

If P is irreducible and reversible w.r.t. π, then $\langle \cdot, \cdot \rangle_\pi : \mathbb{R}^{|\mathcal{X}|} \to \mathbb{R}$ by

$$\langle f, g \rangle_\pi = \sum_{x \in \mathcal{X}} f(x) g(x) \pi(x)$$

is an inner product on $\mathbb{R}^{|\mathcal{X}|}$, which is a Hilbert space with respect to $\langle \cdot, \cdot \rangle_\pi$.

The π-inner product

If P is irreducible and reversible w.r.t. π, then $\langle \cdot, \cdot \rangle_\pi : \mathbb{R}^{|\mathcal{X}|} \to \mathbb{R}$ by

$$\langle f, g \rangle_\pi = \sum_{x \in \mathcal{X}} f(x)g(x)\pi(x)$$

is an inner product on $\mathbb{R}^{|\mathcal{X}|}$, which is a Hilbert space with respect to $\langle \cdot, \cdot \rangle_\pi$.

So...

The spectral representation of the chain

Lemma

Let P be aperiodic, irreducible, and reversible with respect to π. Then:

1. P *is a self-adjoint operator on* $(\mathbb{R}^{|\mathcal{X}|}, \langle \cdot, \cdot \rangle_\pi)$.
2. 1 *has multiplicity* 1 *as an eigenvalue of* P*, and the corresponding (right) eigenspace is spanned by the all* 1*'s vector* $\mathbf{1}$.
3. -1 *is not an eigenvalue of* P.

The spectral representation of the chain

Lemma

Let P be aperiodic, irreducible, and reversible with respect to π. Then:

1. *P is a self-adjoint operator on $(\mathbb{R}^{|\mathcal{X}|}, \langle \cdot, \cdot \rangle_\pi)$.*
2. *1 has multiplicity 1 as an eigenvalue of P, and the corresponding (right) eigenspace is spanned by the all 1's vector $\mathbf{1}$.*
3. *-1 is not an eigenvalue of P.*

Let $\lambda_* = \max\{|\lambda| : \lambda \in \text{spec}(P), \lambda \neq 1\}$.

The spectral representation of the chain

Lemma

Let P be aperiodic, irreducible, and reversible with respect to π. Then:

1. *P is a self-adjoint operator on $(\mathbb{R}^{|\mathcal{X}|}, \langle \cdot, \cdot \rangle_\pi)$.*
2. *1 has multiplicity 1 as an eigenvalue of P, and the corresponding (right) eigenspace is spanned by the all 1's vector $\mathbf{1}$.*
3. *-1 is not an eigenvalue of P.*

Let $\lambda_* = \max\{|\lambda| : \lambda \in \text{spec}(P), \lambda \neq 1\}$. By the above and fact $\sigma(P) = 1$, have $0 \leq \lambda_* < 1$.

Bounding mixing time with λ_*

Recall

$$d(t) = \max_{x \in \mathcal{X}} \left\| \mu_x P^t - \pi \right\|_{TV}.$$

Bounding mixing time with λ_*

Recall

$$d(t) = \max_{x \in \mathcal{X}} \left\| \mu_x P^t - \pi \right\|_{TV}.$$

λ_* controls the asymptotic (in t) rate of convergence of $d(t)$ to 0, i.e., for some c and C which depend on P we have

$$c\lambda_*^t \leq d(t) \leq C\lambda_*^t.$$

Bounding mixing time below with λ_*

A statistical perspective: think of vector $f \in \mathbb{R}^{|\mathcal{X}|}$ as a function ("statistic") on $\mathcal{X}$.

Bounding mixing time below with λ_*

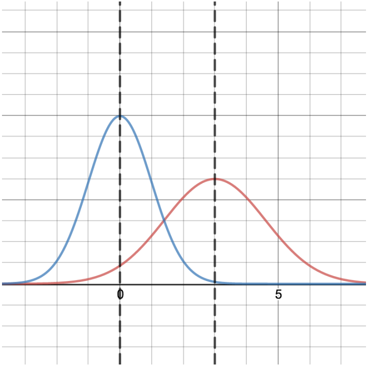
A statistical perspective: think of vector $f \in \mathbb{R}^{|\mathcal{X}|}$ as a function ("statistic") on $\mathcal{X}$.

Distinguish the distributions $\mu_x P^t$ and π using the statistic f.

Bounding mixing time below with λ_*

A statistical perspective: think of vector $f \in \mathbb{R}^{|\mathcal{X}|}$ as a function ("statistic") on $\mathcal{X}$.

Distinguish the distributions $\mu_x P^t$ and π using the statistic f.



Bounding mixing time below with λ_*

Theorem (Spectral Lower Bound)

For P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1 \right) \log \frac{1}{2\varepsilon}.$$

Bounding mixing time below with λ_*

Theorem (Spectral Lower Bound)

For P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \geq \left(\frac{1}{1 - \lambda_*} - 1 \right) \log \frac{1}{2\varepsilon}.$$

Proof.

For any $f \in \mathbb{R}^{|\mathcal{X}|}$ and $x \in \mathcal{X}$,

Bounding mixing time below with λ_*

Theorem (Spectral Lower Bound)

For P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1 \right) \log \frac{1}{2\varepsilon}.$$

Proof.

For any $f \in \mathbb{R}^{|\mathcal{X}|}$ and $x \in \mathcal{X}$,

$$|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| =$$

Bounding mixing time below with λ_*

Theorem (Spectral Lower Bound)

For P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon}.$$

Proof.

For any $f \in \mathbb{R}^{|\mathcal{X}|}$ and $x \in \mathcal{X}$,

$$|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| = \left| \sum_{y \in \mathcal{X}} (\mu_x P^t(y) - \pi(y)) f(y) \right|$$

Bounding mixing time below with λ_*

Theorem (Spectral Lower Bound)

For P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon}.$$

Proof.

For any $f \in \mathbb{R}^{|\mathcal{X}|}$ and $x \in \mathcal{X}$,

$$|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| = \left|\sum_{y \in \mathcal{X}} (\mu_x P^t(y) - \pi(y)) f(y)\right|$$

$$\leq \|f\|_\infty 2d(t)$$

where $\mathbb{E}_\nu(\cdot)$ is expected value taken against distribution ν.

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) =$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f =$

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f =$

Bounding mixing time below with λ_*

Spectral Theory of Finite Markov Chains

Austin Eide

Preliminaries

Spectral Rep. Chains

Examples

Intuition: the Dirichlet Energy

References

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f =$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f)$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0.$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0.$
2. $\mathbb{E}_{\mu_x P^t}(f) =$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0.$
2. $\mathbb{E}_{\mu_x P^t}(f) = \mu_x P^t f =$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0$.
2. $\mathbb{E}_{\mu_x P^t}(f) = \mu_x P^t f = \mu_x \lambda^t f =$

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0$.
2. $\mathbb{E}_{\mu_x P^t}(f) = \mu_x P^t f = \mu_x \lambda^t f = \lambda^t f(x)$.

Bounding mixing time below with λ_*

Proof.

We have $|\mathbb{E}_{\mu_x P^t}(f) - \mathbb{E}_\pi(f)| \leq \|f\|_\infty 2d(t)$. So any lower bound on the LHS gives a lower bound on $d(t)$.

If f is an eigenvector of P with eigenvalue $\lambda \neq 1$, we know two things:

1. $\mathbb{E}_\pi(f) = \pi f = \pi P f = \lambda \pi f = \lambda \mathbb{E}_\pi(f) \implies \mathbb{E}_\pi(f) = 0$.
2. $\mathbb{E}_{\mu_x P^t}(f) = \mu_x P^t f = \mu_x \lambda^t f = \lambda^t f(x)$.

So $|\lambda^t f(x)| \leq \|f\|_\infty 2d(t)$ for any x and eigenvalue $\lambda \neq 1$.

Bounding mixing time below with λ_*

Proof.

Optimizing over x and λ gives $\frac{\lambda_*^t}{2} \leq d(t)$.

Bounding mixing time below with λ_*

Proof.

Optimizing over x and λ gives $\frac{\lambda_*^t}{2} \leq d(t)$.

Setting the LHS to be at least ε and solving for t yields

$$t_{\text{mix}}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon}.$$

□

Bounding mixing time below with λ_*

Proof.

Optimizing over x and λ gives $\frac{\lambda_*^t}{2} \leq d(t)$.

Setting the LHS to be at least ε and solving for t yields

$$t_{\text{mix}}(\varepsilon) \geq \left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon}.$$

□

This can be understood as a "first moment" bound, i.e., relying only on expectations. If variances are computable, better bounds sometimes exist.

Bounding mixing time above with λ_*

Theorem (Spectral Upper Bound)

P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

Bounding mixing time above with λ_*

Theorem (Spectral Upper Bound)

P as before, $\varepsilon > 0$:

$$t_{mix}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

Proof.

A bit more technical, uses the diagonalization of P. □

Remarks

$$\left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon} \leq t_{\mathsf{mix}}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

Remarks

$$\left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon} \leq t_{\mathsf{mix}}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

- For a fixed chain, these bounds are quite tight...

Remarks

$$\left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon} \leq t_{\mathsf{mix}}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

- For a fixed chain, these bounds are quite tight...
- But common to have $|\mathcal{X}| = n$ and $n \to \infty$.

Remarks

$$\left(\frac{1}{1-\lambda_*} - 1\right) \log \frac{1}{2\varepsilon} \leq t_{\mathsf{mix}}(\varepsilon) \leq \frac{1}{1-\lambda_*} \log \frac{1}{\varepsilon \pi_{\min}}$$

- For a fixed chain, these bounds are quite tight...
- But common to have $|\mathcal{X}| = n$ and $n \to \infty$. Here, you pay a price for the $\log \frac{1}{\pi_{\min}}$.

Remarks

$$\left(\frac{1}{1-\lambda_*}-1\right)\log\frac{1}{2\varepsilon} \leq t_{\mathsf{mix}}(\varepsilon) \leq \frac{1}{1-\lambda_*}\log\frac{1}{\varepsilon\pi_{\min}}$$

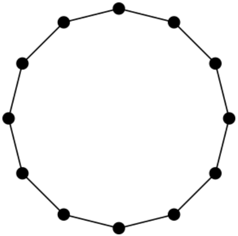
- For a fixed chain, these bounds are quite tight...
- But common to have $|\mathcal{X}| = n$ and $n \to \infty$. Here, you pay a price for the $\log\frac{1}{\pi_{\min}}$.
- In many chains like this, a *cutoff phenomenon* is observed: as $n \to \infty$, $d(t)$ approaches a step function which jumps from 1 (completely unmixed) to 0 (completely mixed) at a critical threshold $t_* = t_*(n)$.

The Cycle

Random walk on the (odd) n-cycle has eigenvalues $\{\cos \frac{2\pi j}{n}\}_{j=0}^{\frac{n-1}{2}}$.

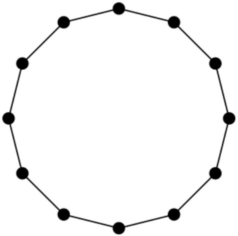
The Cycle

Random walk on the (odd) n-cycle has eigenvalues $\{\cos \frac{2\pi j}{n}\}_{j=0}^{\frac{n-1}{2}}$. So $\lambda_* = \cos \frac{2\pi}{n} = 1 - \frac{2\pi^2}{n^2} + O(n^{-4})$.



The Cycle

Random walk on the (odd) n-cycle has eigenvalues $\{\cos \frac{2\pi j}{n}\}_{j=0}^{\frac{n-1}{2}}$. So $\lambda_* = \cos \frac{2\pi}{n} = 1 - \frac{2\pi^2}{n^2} + O(n^{-4})$.



Since stationary is uniform, our bounds give

$$\frac{\pi^2 n^2}{2} \log \frac{1}{2\varepsilon} \lesssim t_{\mathsf{mix}}(\varepsilon) \lesssim \frac{\pi^2 n^2}{2} \log \frac{n}{\varepsilon}$$

Card Shuffling

Think of S_n as set of orderings of an n-card deck, laid side-by-side on a table.

Card Shuffling

Think of S_n as set of orderings of an n-card deck, laid side-by-side on a table.

Consider the Markov chain on S_n obtained by iterating the following rule: pick a random pair of cards and transpose them.

Card Shuffling

Think of S_n as set of orderings of an n-card deck, laid side-by-side on a table.

Consider the Markov chain on S_n obtained by iterating the following rule: pick a random pair of cards and transpose them.

Theorem (Diaconis & Shashashani '81)

For this chain, for any $\varepsilon > 0$

$$t_{mix}(\varepsilon) \sim \frac{1}{2} n \log n$$

(independent of ε).

A combinatorial optimization problem

Let P be reversible with respect to π.

A combinatorial optimization problem

Let P be reversible with respect to π.

Challenge: pick a function $f \in \{\pm 1\}^{|\mathcal{X}|}$ minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$, where

A combinatorial optimization problem

Let P be reversible with respect to π.

Challenge: pick a function $f \in \{\pm 1\}^{|\mathcal{X}|}$ minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$, where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathcal{X}} [f(x) - f(y)]^2 \pi(x) P(x,y)$$

A combinatorial optimization problem

Let P be reversible with respect to π.

Challenge: pick a function $f \in \{\pm 1\}^{|\mathcal{X}|}$ minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$, where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathcal{X}} [f(x) - f(y)]^2 \pi(x) P(x,y) = \langle (I - P)f, f \rangle_\pi$$

A combinatorial optimization problem

Let P be reversible with respect to π.

Challenge: pick a function $f \in \{\pm 1\}^{|\mathcal{X}|}$ minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$, where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathcal{X}} [f(x) - f(y)]^2 \pi(x) P(x,y) = \langle (I-P)f, f \rangle_\pi$$

s.t. $f \cdot \mathbf{1} = 0$.

A combinatorial optimization problem

Let P be reversible with respect to π.

Challenge: pick a function $f \in \{\pm 1\}^{|\mathcal{X}|}$ minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$, where

$$\mathcal{E}(f) := \frac{1}{2} \sum_{x,y \in \mathcal{X}} [f(x) - f(y)]^2 \pi(x) P(x,y) = \langle (I - P) f, f \rangle_\pi$$

s.t. $f \cdot \mathbf{1} = 0$.

If we identify P with it's edge weighted graph G_P, this is equivalent to finding a balanced labeling of the vertices of G_P with ± 1 minimizing the above.

The Dirichlet Energy

Minimizing $\mathcal{E}(f)$ is analogous to the problem of minimizing

The Dirichlet Energy

Minimizing $\mathcal{E}(f)$ is analogous to the problem of minimizing

$$\int_{\Omega} \|\nabla u\|^2 \, dx$$

over $u : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ s.t. some boundary conditions.

The Dirichlet Energy

Minimizing $\mathcal{E}(f)$ is analogous to the problem of minimizing

$$\int_{\Omega} \|\nabla u\|^2 \, dx$$

over $u : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ s.t. some boundary conditions.

To solve the continuous version, one solves Laplace's Equation $\Delta u = 0$.

The (Discrete) Dirichlet Energy

We can relax our combinatorial problem to minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$ over *any* $f \in \mathbb{R}^{|\mathcal{X}|}$ s.t. $\langle f, \mathbf{1} \rangle_\pi = 0$ (and $f \neq 0$).

The (Discrete) Dirichlet Energy

We can relax our combinatorial problem to minimizing $\frac{\mathcal{E}(f)}{\|f\|_2^2}$ over *any* $f \in \mathbb{R}^{|\mathcal{X}|}$ s.t. $\langle f, \mathbf{1} \rangle_\pi = 0$ (and $f \neq 0$).

Theorem

Let P have eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{|\mathcal{X}|}$ with eigenvectors $f_1, f_2, \ldots, f_{|\mathcal{X}|}$. The above optimization problem is solved by taking $f = 1 - f_2$, and thus has minimum value $\gamma = 1 - \lambda_2$.

References

1. *Markov Chains and Mixing Times*. Levin, D., and Peres, Y.