Summing up and generalizing:

Suppose we have a process $X_1, X_2, \dots, X_t, \dots$ (sequence of random variables)

If $P(X_{t+1} = x \mid X_1 = x_1, \dots, X_t = x_t) = P(X_{t+1} = x \mid X_t = x_t)$

we say this is a Markov process on Ω (before $\Omega = S_n$)

$P = (P(x,y))_{x,y \in \Omega}$ if $\exists$ measure $\pi : \Omega \to [0,1]$ such that $P^t_x \to \pi$ as $t \to \infty$.

$t_{mix}(\varepsilon) \overset{def}{=} \min\{t > 0 : \max_x d_x(t) < \varepsilon\}$

(some distance)

$T : \Omega \times \dots \times \Omega \times \dots \to \mathbb{N}$ random variable, is a stopping time for (X_t) if $\{T \leq t\}$ depends only on $X_1, X_2, \dots, X_t$.

Example: T = first time that card 1 is on top.

Definition: T is a strong stationary time for (X_t) if $\mathbb{P}_x(X_t = y \mid T \leq t) = \pi(y)$

Lemma (A-D) $s(t) \leq P(T > t)$.

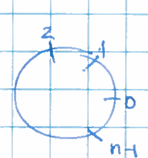
The proof is exactly the same as in the special case discussed above.

Tomorrow: 1) Eigenvalues and eigenfunctions of transition matrix in mixing
2) Comparison

Lecture 2 - Eigenvalues and eigenfunctions of the transition matrix

$C_n = \mathbb{Z}/n\mathbb{Z}$ $\quad S = \{0, 1, -1\}$

Start at 0 with probability 1/2 stay fixed
1/4 go 1 to the right or left



How long until we get "lost in C_n". This is another Markov chain that people know almost everything about like card shuffling. Transition matrix

$$P(x,y) = \begin{cases} 1/2 & y = x \\ 1/4 & y = x+1, x-1 \\ 0 & \text{otherwise} \end{cases} = P(y,x)$$

Definitions: Markov chain on X with transition matrix P (X always finite)

① A probability measure $\pi : X \to [0,1]$ is called stationary if $\pi P = \pi$

$X = \{x_1, \dots, x_n\}$ $\quad \pi : X \to [0,1]$ $\quad \pi = \begin{pmatrix} \pi(x_1) \\ \vdots \\ \pi(x_n) \end{pmatrix}^T$

② Markov chain graph G
vertex set X
Edges $x \sim y$ if $P(x,y) > 0$

For the example above the graph would be

③ Markov chain is irreducible if G is connected.

④ Markov chain is aperiodic if G is not bipartite

means we can write vertex set $V = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$ and there are no edges connecting vertices of V_1 nor V_2

i.e. If $x \sim y$ then $x \in V_1, y \in V_2$ or $x \in V_2$ and $y \in V_1$.

In the above example, if n is even and we remove the loops we would get a bipartite graph. The loops make sure that it is never bipartite.

5) (P, π) are reversible if $\pi(x) P(x,y) = \pi(y) P(y,x)$

If π is the uniform measure this means that the matrix is symmetric.

We expect these to have nice eigenfunctions.

Lemma: If (P,π) is reversible then π is stationary for P.

Proof:
$$\pi P(y) = \sum_x \pi(x) P(x,y) = \sum_x \pi(y) P(y,x) = \pi(y) \underbrace{\sum_x P(y,x)}_{1} \qquad \square$$

(Remark: In order to have $P_x^t \to \pi$ we will see that the Markov chain is irreducible and aperiodic)

Inner product: $f, g : X \to \mathbb{R}, \mathbb{C}$ $\qquad \langle f, g \rangle = \sum_{x \in X} f(x) g(x) \pi(x)$

Lemma: If (P,π) are reversible then $\langle Pf, g \rangle = \langle f, Pg \rangle$ (P is self-adjoint)

Proof: Exercise.

Hence P has real eigenvalues $\lambda_{|X|} \leq \dots \leq \lambda_2 \leq \lambda_1$ and we have a orthonormal basis of eigenfunctions ϕ_i
$$P\phi_i = \lambda_i \phi_i$$
with $\sum_x \phi_i(x) \phi_j(x) \pi(x) = 0$, $\sum_x \phi_i^2(x) \pi(x) = 1$.

Theorem (P, π) reversible, X finite

a) $\lambda_1 = 1$ and $|\lambda_i| \leq 1$ for all $i = 1, 2, \dots, |X|$

b) P is irreducible if and only if $\lambda_2 < 1$

c) P is aperiodic iff $\lambda_{|X|} > -1$

d) $\dfrac{1}{\pi(x)} = \sum_{i=1}^{|X|} \phi_i(x)^2$

e) $4 \| P_x^t - \pi \|_{TV}^2 \leq \sum_{i=2}^{|X|} (\phi_i(x))^2 \lambda_i^{2t}$ (will go to zero as $t \to \infty$ when P is aperiodic + irreducible)

Proof: a) Clearly $f = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ is an eigenfunction with eigenvalue 1

Suppose we have $P\phi = \lambda \phi$ so that $\lambda \phi(x) = (P\phi)(x) = \sum_{y \in X} P(x,y) \phi(y)$

$$|\lambda \phi(x)| \leq \sum_{y \in X} P(x,y) |\phi(y)|$$

choose $\bar{x}$ such that $|\phi(\bar{x})|$ is maximal. Then $|\lambda| |\phi(\bar{x})| \leq \sum_{y \in X} P(x,y) |\phi(\bar{x})| = |\phi(\bar{x})| \Rightarrow |\lambda| \leq 1$,

b) Assume $\lambda_2 < 1$. If P is not irreducible there are $X_1, X_2 \subset X$ non-empty so that we can't go between X_1 and X_2. Rearranging the transition matrix as
$$P = \begin{array}{c} \\ X_1 \\ X_2 \end{array} \begin{pmatrix} X_1 & X_2 \\ ? & 0 \\ 0 & ? \end{pmatrix}$$
we have that it is block diagonal

and we then have two independent eigenfunctions $\begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \begin{matrix} X_1 \\ X_2 \end{matrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \begin{matrix} X_1 \\ X_2 \end{matrix}$

Now assume that P is irreducible. Let f be an eigenfunction of P corresponding to 1. Then $Pf = f$

$f(x) = \sum_{y \in X} P(x,y) f(y)$ $\qquad$ Let $\bar{x}$ be the maximum point of f and assume without loss $f(\bar{x}) > 0$.

If $\tilde{y}$ is adjacent to $\bar{x}$ and $f(\tilde{y}) < f(\bar{x})$ we get a contradiction:

$$f(\bar{x}) = \sum_{y \in X} P(\bar{x},y) f(y) < \sum_{y\in X} P(\bar{x},y) f(\bar{x}) < f(\bar{x})$$

Now the same must be true for the neighbours of $\bar{y}$. Since the graph is connected we see that the eigenfunction f is constant, i.e.,

$$f = f(\bar{x}) \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

c) ~~If P is aperiodic, f, λ_n~~
If $\lambda_n = -1$ and $Pf = -f$ then ~~Pf~~ $f(x) = -\sum_y P(x,y) f(y)$

The same argument as above shows $f(x) = -f(y)$ for every y with $x \sim y$

Now $f = P^2 f$ $\quad f(x) = \sum_{y\in X} P^2(x,y) f(y) \Rightarrow f(x) = f(y)$ for all y at distance 2 from x

Continuing in this way $f(y) = \begin{cases} f(x) & \text{if the distance of } y \text{ to } x \text{ is even} \\ -f(x) & \text{'' if } y \text{ is odd} \end{cases}$

distance odd just means you can get to the point in a odd number of steps not that that is the least distance.

This separates X into 2 disjoint subsets: $X_1 = \{ y : \text{distance}(x,y) \text{ odd} \}$
$X_2 = \{ y : \text{distance}(x,y) \text{ is even} \}$

$\Rightarrow P$ is not aperiodic.

Conversely ~~[crossed out]~~ if G is bipartite then P has the form $\begin{pmatrix} 0 & ? \\ ? & 0 \end{pmatrix}$ define

$$f(x) = \begin{cases} 1 & x \in X_1 \\ -1 & x \in X_2 \end{cases} \quad \text{then} \quad Pf = -f \Rightarrow \lambda_n = -1$$

d) $\dfrac{1}{\pi(x)} = \dfrac{\delta_{xx}}{\pi(x)}$ where $\delta_{xy} = \begin{cases} 1 & \text{if } x=y \\ 0 & \text{otherwise} \end{cases}$

Write $\dfrac{\delta_{x,y}}{\pi(x)} = \sum_{i=1}^{n} \left\langle \dfrac{\delta_x}{\pi(x)}, \phi_i \right\rangle \phi_i(y)$

$$\left\langle \frac{\delta_x}{\pi(x)}, \phi_i \right\rangle = \sum_{y\in X} \frac{\delta_{x,y}}{\pi(x)} \phi_i(y)\pi(y) = \phi_i(x)$$

Hence $\dfrac{\delta_{x,y}}{\pi(x)} = \sum_{i=1}^{n} \phi_i(x)\phi_i(y)$ and setting $x=y$ we get the desired expression.

We used $f = \sum a_i \phi_i$ with $a_i = \langle f, \phi_i \rangle$. For e) we will also use $\langle f, f \rangle = \sum_i a_i^2 = \sum_{i=1}^{n} \langle f, \phi_i \rangle^2$

e) $$2\|P_x^t - \pi\|_{TV} = \sum_{y\in X} |P_x^t(y) - \pi(y)| = \sum_{y\in X} \pi(y) \left| \frac{P_x^t(y)}{\pi(y)} - 1 \right| = \sum_{y\in X} \sqrt{\pi(y)}\sqrt{\pi(y)} \left| \frac{P_x^t(y)}{\pi(y)} - 1 \right|$$

Cauchy-Schwarz
$$\leq \left(\sum_{y\in Y} \pi(y) \right)^{1/2} \left(\sum_{y\in X} \pi(y) \left| \frac{P_x^t(y)}{\pi(y)} - 1 \right|^2 \right)^{1/2} = \left\langle \frac{P_x^t}{\pi} - 1, \frac{P_x^t}{\pi} - 1 \right\rangle^{1/2} = \left\| \frac{P_x^t}{\pi} - 1 \right\|_2$$

Plancherel
$$= \left(\sum_i \left\langle \frac{P_x^t}{\pi} - 1, \phi_i \right\rangle^2 \right)^{1/2}$$

Now $$\frac{P_x^t(y)}{\pi(y)} = \sum_{i=1}^{n} \left(\sum_{z\in X} \frac{P_x^t(z)}{\pi(z)} \phi_i(z) \pi(z) \right) \phi_i(y) = \sum_{i=1}^{n} \lambda_i^t \phi_i(x)\phi_i(y) = 1 + \sum_{i=2}^{n} \lambda_i^t \phi_i(x)\phi_i(y)$$

Hence $\dfrac{P_x^t(y)}{\pi(y)} - 1 = \sum_{i=2}^{n} \lambda_i^t \phi_i(x)\phi_i(y)$ which concludes the proof. □

<u>Example:</u> Consider again G_n with $P(x,y) = \begin{cases} \frac{1}{2} & \text{if } x=y \\ \frac{1}{4} & \text{if } y = x\pm1 \\ 0 & \text{otherwise} \end{cases}$

Exercise: eigenvalues are $\lambda_i = \frac{1}{2} + \cos\frac{2\pi i}{n}$ $\quad f_i(x) = \cos\left(\frac{2\pi i x}{n}\right)$

There are ways of finding eigenvalues and functions using rep. theory + Fourier analysis - See Persi's book on Rep theory in Probability and Statistics, chapter 2 for more on this.

$\pi(y) = \frac{1}{n}$ is the stationary measure (by symmetry)

$$4 \| P_0^t - \pi \|_{TV}^2 \leq \sum_{j=1}^{n-1} \left(\frac{1}{2} + \frac{1}{2}\cos \frac{2\pi j}{n} \right)^{2t}$$ (See "Threads through group theory" by Diaconis)

[note that $f_j(0) = 1$]

For $\cos \frac{2\pi j}{n} \leq 0$ we have $0 \leq \frac{1}{2} + \frac{1}{2}\cos\frac{2\pi j}{n} \leq \frac{1}{2}$ so

$$\sum_{j=1}^{n-1} \left(\frac{1}{2} + \frac{1}{2}\cos\frac{2\pi j}{n}\right)^{2t} \leq n\left(\frac{1}{2}\right)^{2t} + \sum_{\cos\frac{2\pi j}{n} > 0} \left(\frac{1}{2} + \frac{1}{2}\cos\frac{2\pi j}{n}\right)^{2t} = n\left(\frac{1}{2}\right)^{2t} + 2\sum_{1 \leq j \leq \frac{n}{4}} \left(\frac{1}{2} + \frac{1}{2}\cos\frac{2\pi j}{n}\right)^{2t}$$

Now $\frac{1}{2} + \frac{1}{2}\cos\frac{2\pi j}{n} \leq \cos\left(\frac{1}{2}0 + \frac{1}{2}\frac{2\pi j}{n}\right) \leq e^{-\pi^2 j^2 / 2n^2}$ ($\cos x \leq e^{-x^2/2}$)

↑ cos is concave in this interval

We conclude that

$$4 \| P_x^t - \pi \|_{TV}^2 \leq n\left(\frac{1}{2}\right)^{2t} + 2\sum_{j=1} e^{-\frac{\pi^2 j^2}{n^2} t}$$

<u>Theorem</u>: If $t = cn^2$ then $4 \| P_x^t - \pi \|_{TV}^2 \leq e^{-c} A$ where A is some number that comes from the expression above.

<u>Note</u>: For very specific examples we can find eigenvalues - typically this happens when the Markov process is determined by a set of conjugacy classes as in the example above where we have $S = \{-1, 0, 1\}$ [here the group is abelian so any set is a set of conj. classes].

<u>Comparison Theory for random walks on groups</u> (Diaconis, Saloff-Coste)

Setup: Let G be a finite group. $S = \{s_1, s_1^{-1}, \ldots, s_k, s_k^{-1}\} \cup \{id\}$ a generating set
$\tilde{S} = \{\tilde{s}_1, \tilde{s}_1^{-1}, \ldots, \tilde{s}_\ell, \tilde{s}_\ell^{-1}\} \cup \{id\}$ another generating set.

Define Markov chains by setting: Fix $\mu : G \to [0,1]$ probability measure $\mu(g) > 0$ for all $s \in S$, $\mu(s) = 0$ if $s \notin S$
$\nu : G \to [0,1]$. " " $\nu(s) \neq 0$ if $s \in \tilde{S}$, 0 otherwise.

assumption $\mu(s) = \mu(s^{-1})$, $\nu(\tilde{s}) = \nu(\tilde{s}^{-1})$ ← needed for reversibility

$P(x, xs) = \mu(s)$
$Q(x, x\tilde{s}) = \nu(\tilde{s})$

can one compare the eigenvalues of P, Q?

So we have $P(x,y) = P(y,x)$

Irreducibility comes from $S, \tilde{S}$ being generating sets.

Adding identities ensures aperiodicity.

<u>Statement</u>: Assume P has eigenvalues $-1 < \lambda_n \leq \ldots \leq \lambda_2 < \lambda_1 = 1$ $n =$ size of G
Q " " $-1 < q_n < \ldots \leq q_2 < q_1 = 1$

Then $\boxed{q_i \leq 1 - \frac{1-\lambda_i}{\ }}$ where to find A we take $s \in S$ and write it as a word in $\tilde{S}$, $s = \tilde{s}_1 \cdots \tilde{s}_b$, length$(s) = b$

Let $N(s,\tilde{s})$ be the number of times $\tilde{s}$ is used in s.

Then
$$A = \max_{\tilde{s} \in \tilde{S}} \left\{ \frac{1}{\nu(\tilde{s})} \sum_{s \in S} \text{length}(s)\, N(s,\tilde{s})\, \mu(s) \right\}$$

Example: Consider S_n with random transpositions (Diaconis, Shub ...)

deck of n cards

left + right hand picks cards uniformly at random (maybe the same)

Then transpose the cards.

[This is related to how DNA encodes information]

$$S = \{ (a,b) \mid a \neq b \} \cup \{ id \}$$

We have $\mu(id) = \frac{1}{n}$ and $\mu(ab) = \frac{2}{n^2}$

Since again S is a union of conjugacy classes, rep theory can be used to find eigenvalues and eigenvectors.

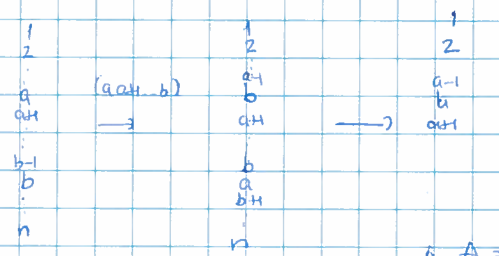
Can see that $\lambda_2 = 1 - \frac{2}{n}$

Random to random: deck of n cards. Pick a card uniformly at random, remove it from deck and insert it in a uniformly random position.

The generators of these transformations are $\tilde{S} = \{ \text{cycles} \} \cup \{ id \}$

Then $\nu(id) = \frac{1}{n}$ $\quad \nu(a\; a+1) = \frac{2}{n^2}$ $\quad \nu(a, a+1, \dots, a+k) = \frac{1}{n^2}$

Given a transposition (a,b) can write it as a product of elements in $\tilde{S}$:



hence length$(a,b) = 2$

$N((a,b), s) = 1$ for $s = (a, a+1, \dots, b)$

We get the following formula for A:

$$A = n^2 \cdot \frac{2}{n^2} \cdot 2 \cdot 2 \cdot 1 = 8$$

(2 transpositions at most)

$\tilde{s} = (a, a+1, \dots, b)$ (this is the maximum you can get)

The inequality says

$$q_2 \leq 1 - \frac{1 - \left(1 - \frac{2}{n}\right)}{8} = 1 - \frac{1}{4n}$$