# Mixing Times of Markov Chains: Lecture 1

**Background** $X$ is markov if the future is independent of the past.

**Def$^n$** $X$ is called a Markov Chain taking values in a space $E$ if $\forall x_0, \dots, x_n \in E$ s.t. $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) > 0$. A MC is defined by it's transition matrix $P$: $P(x,y) = \mathbb{P}(X_1 = y \mid X_0 = x)$

Note: Only studying time homogeneous MCs.

**Check:** $P(X_t = y \mid X_0 = x) = P^t(x,y)$ or $p_{xy}(t)$ or $p_t(x,y)$.

**Def$^n$** A MC is called irreducible if $\forall x,y \in E\ \exists n \geq 0$ s.t. $P^n(x,y) > 0$

recurrent if $P_x(T_x < \infty) = 1$, $T_x = \inf\{t \geq 1 : X_t = x\}$ $\forall x$

transient otherwise

**Def$^n$** An MC is aperiodic if $\gcd\{n \geq 1 : P^n(x,x) > 0\} = 1 \quad \forall x$.

**Def$^n$** $\pi$ is an invariant distribution if it is a prob. dist. s.t if $X_0 \sim \pi$, then $\forall n \ X_n \sim \pi \iff \pi = \pi P$

Let $X$ be a MC, $P, \pi$. Fix $N$ and let $Y_z = X_{N-z}$, $z \in \{0, \dots, N\}$, $X_0 \sim \pi$. Then $Y$ is a MC with transition matrix $P^*(x,y) = \pi(y) P(y,x) / \pi(x)$

$X$ is called reversible if $P^* = P \iff \forall x,y \quad \pi(x) P(x,y) = \pi(y) P(y,x)$.

Let $f, g : E \to \mathbb{R}$. Define $\langle f, g \rangle_\pi = \sum_x f(x) g(x) \pi(x)$

Then check: $\langle Pf, g \rangle_\pi = \langle f, P^* g \rangle_\pi$

**Example** SRW on a graph. Let $G=(V,E)$ be a finite graph. Then SRW on $G$ is the MC with $P(x,y) = 1/\deg(x)$ $x \sim y$

$\pi(x) = \dfrac{\deg(x)}{2|E|}$ is invariant and SRW is reversible.

**Theorem** (Convergence to Equilibrium) Let $X$ be a aperiodic & irreducible MC on a finite state space with $P$ and $\pi$. Then as $t \to \infty$,

$P^t(x,y) \to \pi(y) \quad \forall x,y.$

We need to define a notion of distance of MCs.

## Total variation distance

**Def$^n$** Let $\mu$ and $\nu$ be 2 prob. dist. on $E$. Define $\|\mu-\nu\|_{TV}$ as

$$\|\mu-\nu\|_{TV} = \max_{A\subset E} |\mu(A) - \nu(A)|$$

**Prop$^n$**

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

**Proof:** Let $B = \{x : \mu(x) \geq \nu(x)\}$ and $A\subset E$. Then

$$\begin{aligned}
\mu(A) - \nu(A) &= \mu(A\cap B) - \nu(A\cap B) + \underbrace{\mu(A\cap B^c) - \nu(A\cap B^c)}_{\leq 0 \text{ by def}^n \text{ of } B} \\
&\leq \mu(A\cap B) - \nu(A\cap B) \\
&= \mu(B) - \nu(B) - \underbrace{(\mu(A^c\cap B) - \nu(A^c\cap B))}_{\geq 0 \text{ again by def}^n \text{ of } B} \\
&\leq \mu(B) - \nu(B)
\end{aligned}$$

Similarly, $\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c) = \mu(B) - \nu(B)$

so $|\mu(A) - \nu(B)| \leq \mu(B) - \nu(B)$ $\forall A$ and taking $A=B$

$\to \max_A |\mu(A) - \nu(A)| = \mu(B) - \nu(B)$

So $\|\mu-\nu\|_{TV} = \mu(B) - \nu(B) = \sum_{x:\mu(x)\geq\nu(x)} (\mu(x)-\nu(x))$

**Proof: (Cont.)** $\Rightarrow \quad \|\mu-\nu\|_{TV} = \frac{1}{2}\sum_{x:\mu(x)\geq\nu(x)} (\mu(x)-\nu(x)) + \frac{1}{2}\sum_{x:\nu(x)>\mu(x)} (\nu(x)-\mu(x))$

$$= \frac{1}{2}\sum_{x} |\mu(x)-\nu(x)|$$

**Remark** TV satisfies the Triangle ineq.

**Def$^{n}$** A coupling of $\mu$ and $\nu$, two prob. distr, is a pair of random variables $(X,Y)$ s.t. $X\sim\mu$, $Y\sim\nu$ on the same prob. space.

**Example** Let $X\sim\mu$ and set $Y=X$. Another is to let $X,Y$ be indep. w/ $X\sim\mu$, $Y\sim\nu$.

**Theorem** Let $\mu,\nu$ be 2 prob. dist$^{n}$s. on $\Omega$. Then $\|\mu-\nu\|_{TV} = \inf\{\mathbb{P}(X\neq Y) : (X,Y) \text{ coupling of } \mu \text{ and } \nu\}$. The infimum is attained & the coupling achieving it is called the optimal coupling.

**Proof:** Let $X,Y$ be a coupling of $\mu$ and $\nu$. Then $\forall A\subset\Omega$, $|\mu(A)-\nu(A)|$

$$= |\mathbb{P}(X\in A) - \mathbb{P}(Y\in A)| = |\mathbb{P}(X\in A, Y\notin A) - \mathbb{P}(X\notin A, Y\in A)|$$

$$\leq \max\{\mathbb{P}(X\in A, Y\notin A), \mathbb{P}(X\notin A, Y\in A)\}$$

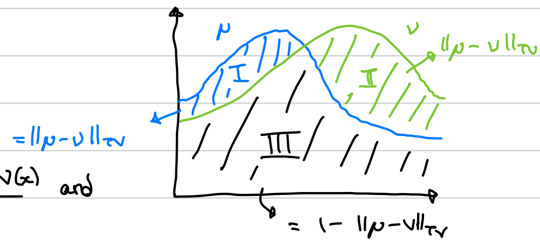
$$\leq \mathbb{P}(X\neq Y)$$

We will construct a coupling that achieves equality.

Toss a coin w.p of H equal to

$$p = \sum_x \mu(x)\wedge\nu(x)$$

$a\wedge b = \min\{a,b\}$
$a\vee b = \max\{a,b\}$




If H then sample $Z\sim\gamma_{III}(x) = \dfrac{\mu(x)\wedge\nu(x)}{p}$ and set $X=Y=Z$.

If T then sample $X\sim\gamma_{I}(x) = \dfrac{\mu(x)-\nu(x)}{1-p}\cdot\mathbb{1}(\mu(x)>\nu(x))$

and independently $Y\sim\gamma_{II}(x) = \dfrac{\nu(x)-\mu(x)}{1-p}\cdot\mathbb{1}(\nu(x)>\mu(x))$.

Since $\gamma_{I}$ & $\gamma_{II}$ have disjoint supports, $X$ and $Y$ will be equal only if the coin comes up heads.

$$p = \sum_x \mu(x) \wedge \nu(x) = \sum_{x:\mu(x)\geq \nu(x)} \nu(x) + \sum_{x:\nu(x)\geq\mu(x)} \mu(x) = 1 - \sum_{x:\mu(x)\geq\nu(x)} (\mu(x)-\nu(x))$$

$$= 1 - \|\mu - \nu\|_{TV}$$

$$\Rightarrow \quad \mathbb{P}(X=x) = p \cdot \frac{\mu(x)\wedge\nu(x)}{p} + (1-p)\left(\frac{\mu(x)-\nu(x)}{1-p}\right) \mathbb{1}(\mu(x) > \nu(x)) = \mu(x)$$

$\Rightarrow$ coupling of $\mu$ and $\nu$.

Finally, $\mathbb{P}(X \neq Y) = 1 - p = \|\mu-\nu\|_{TV}$ ■

**Def$^n$** Let $0$ and $\pi$ be a stochastic matrix & its invariant dist$^n$.

Define
$$d(t) = \max_x \|P^t(x,\cdot) - \pi\|_{TV}$$
(Worst TV distance from $\pi$ after $t$ steps)

$$\bar{d}(t) = \max_{x,y} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}$$
(worst distance between two starting points after $t$)

**Lemma** $\forall t$ we have $d(t) \leq \bar{d}(t) \leq 2d(t)$

**Proof:** $\bar{d}(t) \leq 2d(t)$ follows from $\Delta$ ineq.

$$\|P^t(x,\cdot) - \pi\|_{TV} = \max_A |P^t(x,A) - \pi(A)|$$
$$= \max_A \left|P^t(x,A) - \sum_y \pi(y) P^t(y,A)\right| \quad \text{since } \pi = \pi P$$
$$\leq \max_A \sum_y \pi(y) \left|P^t(x,A) - P^t(y,A)\right|$$
$$\leq \sum_y \pi(y) \underbrace{\max_A |P^t(x,A) - P^t(y,A)|}_{\|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}} \leq \bar{d}(t).$$

**Theorem** $P$ be aperiodic, irred. on finite state space and $\pi$ invar. dist$^n$. Then there exist $\beta \in (0,1)$ and $C > 0$ s.t. $\max_x \|P^t(x,\cdot) - \pi\|_{TV} \leq C\beta^t$

**Proof:** Because of irred. & aperiodic. $\exists\, r > 0$ s.t. $P^r$ has strictly +ve entries. (Finite state space). Set $\alpha = \min_{x,y} \frac{P^r(x,y)}{\pi(y)}$, then $\alpha > 0$.

Then $\forall x,y \quad P^r(x,y) \geq \alpha\pi(y)$ $\leftarrow$ $\begin{pmatrix} -\pi- \\ \vdots \\ -\pi- \end{pmatrix}$

$\Rightarrow \quad P^r(x,y) = \alpha\pi(y) + (1-\alpha)Q(x,y)$ where $Q$ is stochastic.

condition on first $r$ steps

$$P^{rk}(x,y) = (1-\alpha)^k Q^k(x,y) + (1-(1-\alpha)^k)\,\pi(y) \qquad \begin{pmatrix} - & \pi & - \\ - & \pi & - \\ & \vdots & \\ - & \pi & - \end{pmatrix}$$

never become stationary $\qquad\qquad$ become stationary

$$P^{rk+j} = (1-(1-\alpha)^k)\pi + (1-\alpha)^k Q^k P^j \qquad \text{since } \pi P = \pi$$

$$\Rightarrow \|P^{rk+j} - \pi\|_{TV} = \|(1-\alpha)^k (Q^k P^j - \pi)\|_{TV}$$

$$= (1-\alpha)^k \|Q^k P^j - \pi\|_{TV} \le (1-\alpha)^k \quad \blacksquare$$

**Def$^n$** For $\varepsilon \in (0,1)$ define $t_{mix}(\varepsilon) = \min\{t \ge 0 : d(t) \le \varepsilon\}$

**Proposition** $\bar{d}$ is submultiplicative, i.e. $\forall\, s,t$ $\bar{d}(s+t) \le \bar{d}(s)\bar{d}(t)$ where $\bar{d} = \max_{x,y} \|P^t(x,\cdot) - P^t(y,\cdot)\|_{TV}$

**Proof:** Fix $x,y$. Let $(X,Y)$ be the optimal coupling of $P^s(x,\cdot)$ and $P^s(y,\cdot)$.

Then $\|P^s(x,\cdot) - P^s(y,\cdot)\|_{TV} = \mathbb{P}(X \ne Y)$ and $P^{s+t}(x,z) = \mathbb{E}[P^t(X,z)]$

$P^{s+t}(y,z) = \mathbb{E}[P^t(Y,z)]$ by Chapman-Kolmogorov.

$$\Rightarrow \|P^{s+t}(x,\cdot) - P^{s+t}(y,\cdot)\|_{TV} = \frac{1}{2}\sum_z |P^{s+t}(x,z) - P^{s+t}(y,z)|$$

$$= \frac{1}{2}\sum_z |\mathbb{E}[P^t(x,z) - P^t(y,z)]|$$

$$\le \mathbb{E}\left[\frac{1}{2}\sum_z |P^t(X,z) - P^t(Y,z)|\right]$$

$$= \mathbb{E}\left[\mathbb{1}(X \ne Y) \cdot \frac{1}{2}\sum_z |P^t(X,z) - P^t(Y,z)|\right]$$

$$\le \mathbb{P}(X \ne Y)\,\mathbb{E}\left[\frac{1}{2}\max_{\omega,\omega'}\left[\sum_z |P^t(x,z) - P^t(Y,z)|\right]\right]$$

$$\le \bar{d}(t)\,\mathbb{P}(X \ne Y).$$

So $\forall x,y$ $\|P^{s+t}(x,\cdot) - P^{s+t}(y,\cdot)\|_{TV} \le \bar{d}(t)\|P^s(x,\cdot) - P^s(y,\cdot)\|_{TV}$

max over $x,y$ $\Rightarrow$ $\bar{d}(s+t) \le \bar{d}(s)\bar{d}(t)$ $\blacksquare$

Also have $d(t) \le \bar{d}(t) \le 2d(t)$ $\Rightarrow$ $d(s+t) \le \bar{d}(s+t) \le \bar{d}(s)\bar{d}(t) \le 2d(t)\bar{d}(s)$

**Remark:** Choose a diff coupling to get $d(s+t) \le \bar{d}(s)d(t)$ (Exercise!)

**Def^n (Mixing Time)** $t_{mix}(\varepsilon) = \min\{t \geq 0 : d(t) \leq \varepsilon\}$. When $\varepsilon = 1/4$ we just write $t_{mix}$.

Why $\varepsilon = 1/4$? $d(l\, t_{mix}) \underset{\text{sub-mult.}}{\leq} \bar{d}(l\, t_{mix}(\varepsilon)) \leq \bar{d}(t_{mix}(\varepsilon))^l \underset{\bar{d} \leq 2d}{\leq} (2\varepsilon)^l$

Taking $\varepsilon = 1/4$ gives $d(l\, t_{mix}) \leq \frac{1}{2^l}$ then $t_{mix}(\varepsilon) \leq \lceil \log_2 \frac{1}{\varepsilon} \rceil t_{mix}$

**Def^n (Coupling)** A coupling of MCs with transition matrix $P$ is a process $(X_t, Y_t)_t$ s.t. both $X$ and $Y$ are MCs w/ transition matrix $P$ and possibly different starting distr.

**Def^n (Markovian Coupling)** In addition, $\forall x, x', y, y'$ $\mathbb{P}(X_1 = x' \mid X_0 = x, Y_0 = y) = P(x, x')$ and $\mathbb{P}(Y_1 = y' \mid X_0 = x, Y_0 = y) = P(y, y')$.

**Def^n (Coalescent)** A coupling is called coalescent if whenever $\exists s$ s.t. $X_s = Y_s$ then $X_t = Y_t$ $\forall t \geq s$. When they touch they stay together.

**Remark** A Markovian coupling can be modified to make it a coalescent coupling. Run the MCs using the Markovian coupling until the first time they meet. Then contin. together.

**Theorem** Let $(X, Y)$ be a Markovian coupling w/ $X_0 = x$ and $Y_0 = y$. Let $\tau_{couple}$ be $\min\{t \geq 0 : X_t = Y_t\}$. Then $\| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV} \leq \mathbb{P}_{x,y}(\tau_{couple} > t)$
In particular if $\forall (x, y)$ $\exists$ Mark. coupling with $\tau_{couple}$ the coupling time, then $d(t) \leq \max_{x,y} P_{x,y}(\tau_{couple} > t)$

Note: Smaller than prev. bound since $X_t = Y_t \Rightarrow \tau_{couple} \leq t$.

Proof: • Fix $x, y$. Markovian $\Rightarrow$ $\mathbb{P}_{x,y}(X_t = x') = P(x, x')$ and same for $Y_t$.
• So $X_t, Y_t$ is a coupling of $P^t(x, \cdot)$ and $P^t(y, \cdot)$
So $\| P^t(x, \cdot) - P^t(y, \cdot) \|_{TV} \leq P_{x,y}(X_t \neq Y_t) \leq \mathbb{P}_{x,y}(\tau_{couple} > t)$
• $d(t) \leq \max_{x,y} \frac{\mathbb{E}_{x,y}[\tau_{couple}]}{t}$ by Markov ineq. $t = 4\mathbb{E}_{x,y}[\tau_{couple}] \Rightarrow d(t) \leq \frac{1}{4} \Rightarrow t_{mix} \leq t$.

**Notation** $f, g$ functions. $f, g : \mathbb{N} \to \mathbb{R}^+$

$f(n) \lesssim g(n)$ if $\exists c > 0$ s.t. $f(n) \leq C g(n)$ $\forall n$.

$f(n) \gtrsim g(n)$ if $g(n) \lesssim f(n)$

$f(n) \asymp g(n)$ if both hold.

**Def$^{n}$** (Lazy chain) Take the MC with transition matrix $\frac{P+I}{2}$ (i.e. with prob $1/2$ remain where you are, with prob $1/2$ move according to $P$.

Note: We use this to avoid aperiodicity.

**Example** SRW on $\mathbb{Z}_n = \{0, \dots, n-1\}$ $\mathbb{P}(i, (i \pm 1) \bmod n) = 1/2$ lazy chain $\frac{P+I}{2}$

Claim: $t_{mix} \asymp n^2$

# Lecture 4

**Claim** Let $X$ be a lazy SRW on $\mathbb{Z}_n$ then $t_{mix} \asymp n^2$

**Proof:** • Take $x,y \in \mathbb{Z}_n$. $X, Y$ two SRWs starting at $x, y$

• Couple as follows: Toss fair coin. If H move $X$ to random neighbour, if T move $Y$. When they meet, continue together.

• clockwise distance between $X, Y$ is SRW on $\{0, 1, .., n\}$ w/ absorption at $0$ and $n$.

Let $\tau = \min\{t \geq 0 : X_t = Y_t\}$. Then $\tau$ is first time distance gets absorbed at $\{0, n\}$

$|x-y| = k \Rightarrow \mathbb{E}_k[\tau] = k(n-k)$ (it is the hitting probs)

• $d(t) \leq \max_{x,y} \dfrac{\mathbb{E}_{x,y}[\tau]}{t} \leq \dfrac{n^2}{4t}$ so if $t = n^2$ then $d(t) \leq 1/4$

$\Rightarrow t_{mix} \leq n^2$

**Lower bound** Let $(S_t)$ be a lazy SRW on $\mathbb{Z}$ starting at $0$ then set $X_t = S_t \bmod n$.

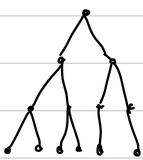
$$\mathbb{P}(X_t \in \underbrace{\{\lceil \tfrac{n}{4} \rceil + 1, \dots, \lceil \tfrac{3n}{4} \rceil\}}_{A}) \leq \mathbb{P}_0(|S_t| > \tfrac{n}{4}) \underset{\text{Chebyshev}}{\leq} \frac{Var(S_t)}{n^2/16} = \frac{8t}{n^2}$$

since $S_t = \sum_{i=1}^{t} \xi_i$ iid $\xi_i = \begin{cases} +1 & \text{wp } 1/4 \\ -1 & \text{wp } 1/4 \\ 0 & \text{ow} \end{cases}$

Take $t = n^2/32$ then $\mathbb{P}(X_t \in A) \leq 1/4$ but $\pi(A) \geq 1/2$ so $d(t) \geq \pi(A) - \mathbb{P}_0(X_t \in A)$

$\geq \frac{1}{4} \Rightarrow t_{mix} \geq n^2/32$

## RW on the finite binary tree



$n$ vertices, root has degree 2, offspring degree 3, leaves degree 1.

**Claim** $t_{mix} \asymp$

**Exercise** $t_{mix} \geq c \cdot \max_{\substack{x, A \\ \pi(A) \geq 1/8}} \mathbb{E}_x[\tau_A]$

**Check** If $\tau_e = \min\{t \geq 0 : X_t = \text{root}\}$ $\max_x \mathbb{E}_x[\tau_e] \leq C \cdot n$ so $t_{mix} \lesssim n$.

**Upper bound** Coupling: Toss fair coin. H $\Rightarrow$ X moves T $\Rightarrow$ Y moves until they reach same level.

So $\tau \leq 1^{st}$ time $X$ hits roots after having visited leaves.

so $\mathbb{E}_{x,y}[\tau] \leq C^{*} n + C^{**} \log n \leq C' n \Rightarrow t_{mix} \lesssim n$

## Top to random shuffle

Deck of $n$ cards. RW on $S_n$.

- $\tau_{top} = \{$time to bottom card get te top $+1\}$ then $X_{\tau_{top}}$ is uniform on $S_n$ and indep of $\tau_{top}$.

**Def$^n$** (Stopping time) A RV $\tau$ such that $\{\tau \le t\} \in \mathcal{F}_t \ \forall t$.

A stationary time is a stopping time $\tau$ (possibly dep. on $X_0$) s.t. $\mathbb{P}_x(X_\tau = y) = \pi(y)$.

A strong stationary time is a stat. time s.t. $\forall t, y \ \mathbb{P}_x(X_\tau = y, \tau = t) = \pi(y) \cdot \mathbb{P}_x(\tau = t)$

**Example** Lazy RW on $\{0,1\}^n$. $(x_1, \dots, x_n) \sim (y_1, \dots, y_n)$ if $\exists! \ i$ s.t. $y_i = 1 - x_i$ and $y_j = x_j \ \forall j \ne i$. Pick a coord. u.a.r (uniform. at. random) and refresh the bit by a uniform one $\{0,1\}$. This is lazy rw on $\{0,1\}^n$

$X_{t+1} = f(X_t, Z_{t+1})$ where $Z_i$ are iid$^n$ Bernoulli$(1/2)$ and indep. of $X_t$.

### Lecture 5

- Defined $s(t)$
- Proved $\| P^t(x, \cdot) - \pi \|_{TV} \le \mathbb{P}_x(\tau > t)$
- $s(2t) \le 1 - (1 - \bar{d}(t))^2$ using C-S.
- Proved some results for a coupon collector
- Compared LRW on $\{0,1\}^n$ to coupon collector to get $t_{mix}(\varepsilon) \le \lceil n \log n + c(\varepsilon) n \rceil$
- Top to random shuffle has that $\tau_{top}$ strong stationary time.
  $d(t) \le \mathbb{P}(\tau_{top} > t)$ Get $t_{mix}(\varepsilon) \le \lceil n \log n + c(\varepsilon) n \rceil$
- Proved lowerbound for $t_{mix}(\varepsilon)$ for $\tau_{top}$.
- Defined what it means for a sequence of MCs to exhibit cutoff.

Lecture 6

- Defined $\mathcal{L}^p$ distance: $f : E \to \mathbb{R}$, $\|f\|_p = \|f\|_{p,\pi} = \begin{cases} \left( \sum |f(x)|^p \pi(x) \right)^{1/p} & 1 \le p < \infty \\ \max_x |f(x)| & p = \infty \end{cases}$
- $d_p(t) = \max_x \| q_t(x, \cdot) - 1 \|_p$ where $q_t(x,y) = \frac{P^t(x,y)}{\pi(y)}$ for a reversible chain.

$$\underset{\substack{\uparrow \\ TV}}{2d(t)} = d_1(t) \le d_2(t) \le d_\infty(t) \quad \text{by Jensens.}$$

- $\mathcal{L}^p$-mixing time $t_{mix}^{(p)}(\varepsilon) = \min\{ t \ge 0 : d_p(t) \le \varepsilon \}$

  For $p = \infty$, $t_{mix}^{(\infty)}(\varepsilon)$ = uniform mixing time
- Let $P$ be a reversible chain wrt $\pi$. Then $\forall t$ $d_\infty(2t) = (d_2(t))^2 = \max \frac{P^{2t}(x,x)}{\pi} - 1$
- Spectral Techniques:
  - $E$ finite state space, $\pi$ prob dist$^{n}$, $f,g : E \to \mathbb{R}$. Defined $\langle f,g \rangle$ and $\langle f,g \rangle_\pi$