Stats 318: Lecture #4

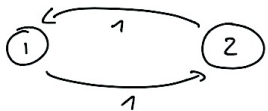
Agenda:

- Aperiodicity
- Strong Irreducibility
- Couplings
- Convergence of Markov Chains: Second Ergodic Theorem

Irreducible chains may not converge!

Irreducibility does not imply that $X_t \xrightarrow{\mathrm{d}} \pi$

Example: $\mathcal{X} = \{1, 2\}$



$$X_0 = 1 \Rightarrow \qquad X_{2t} = 1 \ \& \ X_{2t+1} = 2$$

Need aperiodicity

Aperiodicity

Definition

Chain is aperiodic iff $\mathrm{GCD}\{t > 0 : P^t(x, x) > 0\} = 1 \ \forall \ x \in \mathcal{X}$

Definition

P is strongly irreducible if $\exists \ t$ s.t. $P^s(x, y) > 0 \ \forall \ x, y \in \mathcal{X}, \ \forall \ s \geq t$

It is possible to reach every state from any initial state in t steps

Proposition

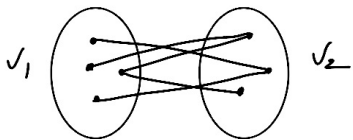
P is strongly irreducible iff P is aperiodic & irreducible

Example: random walk on graph

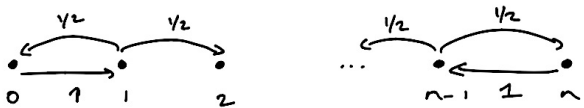
Irreducibility $\Rightarrow$ graph is connected

$$x \sim y \qquad P^2(x,x) \geq P(x,y)P(y,x) > 0 \qquad \therefore \quad \text{period is either } 1 \text{ or } 2$$

Period is 2 if graph is bipartite
iff $V = V_1 \cup V_2$ (V_1 & V_2 disjoint) and all edges connect one vertex of V_1 & one vertex of V_2



e.g. RW with reflecting boundaries



Bipartite graph

$$V_1 = \{0, 2, 4, \dots\}$$
$$V_2 = \{1, 3, 5, \dots\}$$

Example: Ehrenfest model

- Chain is irreducible with period 2
- X_0 even $\implies$ $\left.\begin{matrix} X_{2t} \text{ even} \\ X_{2t+1} \text{ odd} \end{matrix}\right\}$ $\implies$ $\{X_t\}$ cannot converge in dist.
- Consider $\widetilde{\mathcal{X}} = \{x \in \mathcal{X} : x \text{ even}\}$ and $\widetilde{X}_t = X_{2t}$
  $\{\widetilde{X}_t\}$ irreducible & aperiodic and invariant distribution is (x even)

$$\widetilde{\pi}(x) = \pi(x) \Big/ \sum_{y \text{ even}} \pi(y)$$

- Why? $\widetilde{X}$ has transition matrix P^2 and since $\pi P^2 = \pi$, $\forall$ y even

$$\sum_{x \text{ even}} \pi(x)P^2(x,y) + \underbrace{\sum_{x \text{ odd}} \pi(x)P^2(x,y)}_{0} = \sum_{x \text{ even}} \pi(x)P^2(x,y) = \pi(y)$$

Total-variation distance

$$\|\mu - \nu\|_{\mathsf{TV}} = \max_A \ |\mu(A) - \nu(A)|$$

Max diff. between prob. assigned to a single event

Proposition

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

$A = \{x : \mu(x) \geq \nu(x)\}$

- $$\mu(A) - \nu(A) = \sum_{x \in A} \mu(x) - \nu(x) = \sum_{x \in A^c} \nu(x) - \mu(x) = \frac{1}{2} \sum_x |\mu(x) - \nu(x)|$$

- Consider any event $B \subset \mathcal{X}$

$$\begin{aligned}\mu(B) - \nu(B) &= \sum_{x \in B} \mu(x) - \nu(x) \\ &= \sum_{x \in B \cup A} |\mu(x) - \nu(x)| - \sum_{x \in B \cup A^c} |\mu(x) - \nu(x)| \leq \mu(A) - \nu(A)\end{aligned}$$

- Similarly, $\nu(B) - \mu(B) \leq \mu(A) - \nu(A)$
- $\therefore$

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

Coupling and total-variance distance

Definition

(X, Y) s.t. $X \sim \mu$, $Y \sim \nu$ called coupling between μ & ν

Proposition

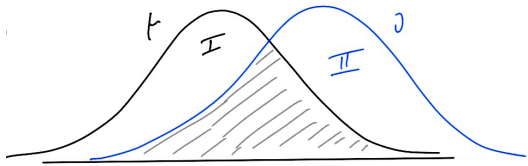
$$\|\mu - \nu\|_{TV} = \inf \{\mathbb{P}(X \neq Y) : (X, Y) \text{ coupling of } \mu \ \& \ \nu\}$$

Proof: For any coupling & any event A

$$\mu(A) - \nu(A) = \mathbb{P}(X \in A) - \mathbb{P}(Y \in A) \leq \mathbb{P}(X \in A, Y \notin A) \leq \mathbb{P}(X \neq Y)$$

$$\implies \qquad \|\mu - \nu\|_{\mathsf{TV}} \leq \ \inf \{\mathbb{P}(X \neq Y) : \ldots\}$$

$$p = \sum_x \mu(x) \wedge \nu(x) = 1 - \|\mu - \nu\|_{\mathsf{TV}} \qquad \text{[Why?]}$$



Flip coin $\mathbb{P}(H) = p$

(i) If 'H', $z \sim q_{//}(x) = \frac{\mu(x) \wedge \nu(x)}{p}$
& $X = Y = Z$

(ii) If 'T', $X \sim q_I,\ Y \sim q_{II}$

$$q_I(x) = \begin{cases} \frac{\mu(x) - \nu(x)}{1-p} & \mu(x) > \nu(x) \\ 0 & \text{o.w.} \end{cases}$$

$$q_{II}(x) = \begin{cases} \frac{\nu(x) - \mu(x)}{1-p} & \nu(x) > \mu(x) \\ 0 & \text{o.w.} \end{cases}$$

IN $X = Y$

OTHERWISE $X \neq Y$

$X \sim \mu$, $Y \sim \nu$ AND

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

Convergence of Markov chains

Theorem

P irreducible & aperiodic. For any dist. $\pi^{(0)}$,

$$\pi^{(t)} \xrightarrow[t\to\infty]{} \pi$$

Let k s.t. $P^k(x,y) > 0 \ \forall\ x, y \in \mathcal{X}$ and set $\beta = \sum_y \{\inf_x P^k(x,y)\}^2 > 0$. Then

$$\|\pi^{(t)} - \pi\|_{\mathsf{TV}} \leq (1-\beta)^{\lfloor t/k \rfloor}$$

$\lfloor a \rfloor$ integer part of a

- Chain approaches equilibrium
- Speed of convergence is 'exponential' ⚠

Proof by coupling (modern)

- $X_0 \sim \pi^{(0)}$ & $X_t = f_{\theta_t} \circ \ldots \circ f_{\theta_1}(X_0) \qquad X_t = f_{\theta_t}(X_{t-1})$
- Second chain
  $\widetilde{X}_0 \sim \pi$ & $\widetilde{X}_t = f_{\widetilde{\theta}_t} \circ \ldots \circ f_{\widetilde{\theta}_1}(\widetilde{X}_0) \qquad \widetilde{X}_t = f_{\widetilde{\theta}_t}(\widetilde{X}_{t-1}) \quad \{\theta_t \ \& \ \widetilde{\theta}_t\}$ indep.
  - Step 1: Two chains meet w.p. 1

$$\tau = \text{ inf } \{t : X_t = \widetilde{X}_t\} \quad \Longrightarrow \quad P(\tau < \infty) = 1$$

  - Step 2: Third chain

$$Y_0 = X_0 \ \& \ Y_{t+1} = \begin{cases} f_{\theta_{t+1}}(X_t) & t < \tau \\ f_{\widetilde{\theta}_{t+1}}(\widetilde{X}_t) & t \geq \tau \end{cases}$$

$\{Y_t\}$ evolves like $\{X_t\}$ until τ, then switches to evolve like $\{\widetilde{X}_t\}$

$\{Y_t\}$ is a Markov Chain with transition matrix P & initial dist. $\pi^{(0)}$

$$Y_0 \sim \pi^{(0)} \qquad \Longrightarrow \qquad Y_t \sim \pi^{(t)}$$

$$1\{Y_t \neq \widetilde{X}_t\} \longrightarrow 0 \text{ a.s.} \qquad \Longrightarrow \qquad Y_t \xrightarrow{d} \widetilde{X}_t = \pi$$

$$\therefore \qquad \pi^{(t)} \longrightarrow \pi$$

Proof of Step 1: $\quad P^k(x,y) > 0 \qquad \forall\ x, y \in \mathcal{X}$

$$\begin{aligned}\mathbb{P}(\tau \le k) \ge \mathbb{P}(X_k = \widetilde{X}_k) &= \sum_y \mathbb{P}(X_k = y, \widetilde{X}_k = y) \\ &= \sum_y \mathbb{P}(X_k = y)\mathbb{P}(\widetilde{X}_k = y) \\ &= \sum_y \Big[\sum_x \mathbb{P}(X_k = y \mid X_0 = x)\mathbb{P}(X_0 = x)\Big]\big[\ \ldots\big]\end{aligned}$$

$$\alpha(y) = \inf_x\ P^k(x,y) \Longrightarrow \mathbb{P}(\tau \le k) \ge \sum_y \alpha^2(y) = \beta$$

$$\mathbb{P}(X_{2k} \neq \widetilde{X}_{2k} \mid \tau > k) \leq 1 - \beta$$

$$\begin{aligned} \therefore \qquad \mathbb{P}(\tau > 2k) &= \mathbb{P}(\tau > 2k \mid \tau > k)\mathbb{P}(\tau > k) \\ &\leq (1-\beta)(1-\beta) \end{aligned}$$

Likewise

$$\mathbb{P}(\tau > \ell k) \leq (1-\beta)^{\ell} \underset{\ell \to \infty}{\longrightarrow} 0 \qquad \text{\& chains meet wp } 1$$

About TV distance

$$\|\pi^{(t)} - \pi\|_{\mathsf{TV}} \leq \mathbb{P}(Y_t \neq \widetilde{X}_t) = \mathbb{P}(\tau > t)$$

Conclusion: $t = \ell k \qquad \|\pi^{(t)} - \pi\|_{\mathsf{TV}} \leq (1-\beta)^{\ell}$

About couplings

$Y_1 \sim \text{Bin}(n, p_1)$, $Y \sim \text{Bin}(n, p_2)$ $\qquad p_2 \geq p_1$

To show that for all $t \geq 0$, $\mathbb{P}(Y_2 \geq t) \geq \mathbb{P}(Y_1 \geq t)$

Proof by coupling:

$$Y_2 = \sum_{i=1}^{n} I_i, \quad I_i \overset{i.i.d.}{\sim} \text{Ber}(p_2)$$

$$Y_1 = \sum_{i=1}^{n} I_i B_i \quad B_i \overset{i.i.d.}{\sim} \text{Ber}(p_1/p_2) \quad \{B_i\} \perp\!\!\!\perp \{I_i\}$$

$$Y_1 \sim \text{Bin}(n, p_1)$$

$$Y_2 \geq Y_1 \qquad \Longrightarrow \qquad \mathbb{P}(Y_2 \geq t) \geq \mathbb{P}(Y_1 \geq t)$$