Markov Chains Mixing Times **October 26, 2022**

Lecture 5: Stationary Time

Lecturer: Yuval Peres Week: 3

Remark 5.55. If τ is a strong stationary time starting from x, then

$$\begin{aligned}\mathbf{P}_x\{\tau \le t, X_t = y\} &= \sum_{s\le t}\sum_z \mathbf{P}_x\{\tau = s, X_s = z, X_t = y\} \\ &= \sum_{s\le t}\sum_z P^{t-s}(z,y)\mathbf{P}_x\{\tau = s\}\pi(z).\end{aligned}$$

It follows from the stationarity of π that $\sum_z \pi(z)P^{t-s}(z,y) = \pi(y)$, whence for all $t \ge 0$ and y

$$\mathbf{P}_x\{\tau \le t, X_t = y\} = \mathbf{P}_x\{\tau \le t\}\pi(y) \tag{5.27}$$

Example 5.56. Top to random shuffle: Consider a pile of n cards. Now, take the top card and insert it at a random location chosen uniformly. This is an irreducible and aperiodic chain on S_n. The time taken to move the bottom most card to 1 position up is distributed as $Geometric(\frac{1}{n})$. The time taken to move up by one more is distributed as $Geometric(\frac{2}{n})$. Continuing this, the time taken to shift the bottom card to top from second position will be distributed as $Geometric(\frac{n-1}{n})$. Hence, τ_{top} has a distribution of coupon collector time,

$$\begin{aligned}|\mathbf{E}\tau_{\text{top}} - nlogn| &\le n \\ \text{Var}\tau_{\text{top}} &\le 2n^2\end{aligned} \tag{5.28}$$

Given at time t that there is only 1 card under the original bottom card, there are two possibilities: either the top card is placed beneath the bottom card or above the bottom card. If the card is placed beneath then there are 2 possible permutations possible. Whereas if the card is placed above the bottom card then the cards below have only one possible ordering. This argument can be extended with induction. Define,

$$\tau_{\text{top}} := 1 + \text{time taken for the bottom card to reach top} \tag{5.29}$$

So, given $\tau_{\text{top}} - 1$ then there will be $(n-1)!$ possible arrangements of cards beneath the original bottom card. In the next move there is n choices so at τ_{top} there are $n!$ orderings each being equally likely. This shows that τ_{top} is a strong stationary time.

Definition 5.57. The ***separation distance*** to stationarity for x is defined as,

$$\begin{aligned}s_x(t) &= 1 - \max\{r : P^t(x,\cdot) \ge r\pi\} \\ &= 1 - \min_{y\in\mathcal{X}} \frac{P^t(x,y)}{\pi(y)}\end{aligned} \tag{5.30}$$

Lemma 5.58. *The total variation distance is upper bounded by the maximum of the separation distances from each of the states in $\mathcal{X}$, i.e. $d(t) \le s(t)$ where $s(t) := \max_x s_x(t)$*

Proof.

$$\left\|P^t(x,\cdot) - \pi\right\|_{\mathrm{TV}} = \sum_y \left[\pi(y) - P^t(x,y)\right]_+ \le \sum_y \pi(y)(1-r)$$
$$r = s_x(t).$$

Hence, we have $d(t) \le \max_x s_x(t) = s(t)$ □

Lemma 5.59. *If τ is a strong stationary time for starting state x, then*

$$s_x(t) \le \mathbf{P}_x\{\tau > t\}. \tag{5.31}$$

Proof.

$$\begin{aligned} \mathbf{P}_x\left\{\tau \le t, X_t = y\right\} &= \mathbf{P}_x\{\tau \le t\}\pi(y) \\ \implies P^t(x,y) &\ge \mathbf{P}_x\{\tau \le t\}\pi(y) \\ \implies s_x(t) &\le 1 - \mathbf{P}_x(\tau \le t) \\ &= \mathbf{P}_x(\tau \ge t) \end{aligned} \tag{5.32}$$

where the first equality is from Remark 5.55. □

Definition 5.60. If τ is stopping time for a chain P starting at x. A ***halting state*** $y \in \mathcal{X}$ is a state such that, if $X_t = y$ then $\tau \le t$.

Also, it can be verified that the inequality (5.31) achieves equality for a halting state.

Example 5.61. Consider the lazy random walk on hypercube. For this case τ_{refresh} is a strong stationary time. If the walk starts on $(0,\ldots,0)$ then halting state is $(1,1,\ldots,1)$.

$$s(t) = \mathbf{P}_0(\tau_{\text{refresh}} > t) \le e^{-cn} \text{ if } t = n\log n + cn$$

And if $t = n\log n - c_1 n$ then,

$$\begin{aligned} s(t) = \mathbf{P}_0(\tau_{\text{refresh}} > t) &= 1 - \mathbf{P}_0(\tau_{\text{refresh}} \le t) \\ &\ge 1 - \frac{\mathrm{Var}(\tau_{\text{refresh}})}{c_1^2 n^2} \\ &\ge 1 - \frac{2}{c_1^2} \end{aligned}$$

This shows that when the time is just a little less than $n\log n$ then the separation distance is quite close to 1. On the other hand, when it is little more than $n\log n$ then the separation distance is quite close to 0. Hence, the separation distance makes a sharp transition from 1 to 0. This phenomenon is known as cutoff. It will be defined rigorously in later part of the course.

Lemma 5.62. *For reversible Markov chain P,*

$$1 - s(2t) \ge \left(1 - \bar{d}(t)\right)^2. \tag{5.33}$$

Proof.

$$\begin{aligned}
\frac{P^{2t}(x,y)}{\pi(y)} &= \frac{1}{\pi(y)}\sum_{z\in\mathcal{X}} P^t(x,z)P^t(z,y) \\
&= \sum_{z\in\mathcal{X}} \pi(z)\frac{P^t(x,z)}{\pi(z)}\frac{P^t(y,z)}{\pi(z)} \\
&\geq \left(\sum_{z\in\mathcal{X}} \pi(z)\sqrt{\frac{P^t(x,z)P^t(z,y)}{\pi(z)^2}}\right)^2 \\
&= \left(\sum_{z\in\mathcal{X}} \sqrt{P^t(x,z)P^t(z,y)}\right)^2 \\
&\geq \left(\sum_{z\in\mathcal{X}} P^t(x,z)\wedge P^t(z,y)\right)^2 \\
&= \left(1-||P^t(x,\cdot)-P^t(y,\cdot)||\right)^2 \\
&\geq (1-\bar{d}(t))^2
\end{aligned} \tag{5.34}$$

□

The third inequality is by Cauchy-Schwarz. The final inequality is obtained by minimizing over x, y.

Example 5.63. For top-to-random shuffle, τ_{top} is a strong stationary time. Then,

$$||P^t(x,\cdot)-\pi||_{\text{TV}} \leq s_x(t) \leq \mathbf{P}_x(\tau_{\text{top}} > t).$$

Hence, $t_{\text{mix}} \leq n\log n + cn$, where $e^{-c} < \frac{1}{4}$.

Example 5.64. Lazy Simple Random Walk on $\{0,1\}^n$

The fact that $d(t_{mix}) \leq \frac{1}{4}$ implies that

$$\begin{aligned}
&\bar{d}(t_{mix}) \leq \frac{1}{2} \\
&\Rightarrow 1-s(2t_{mix}) \geq \frac{1}{4}.
\end{aligned} \tag{5.35}$$

Now, take $t < \frac{n\log n}{2} - \tilde{c}n$ which implies that:

$$\begin{aligned}
&2t < n\log n - 2\tilde{c}n \\
&\Rightarrow s(2t) \geq 1-\frac{1}{2\tilde{c}^2}.
\end{aligned} \tag{5.36}$$

Now, take $\tilde{c}=2 \Rightarrow 1-s(2t) \leq \frac{1}{8} \Rightarrow t < t_{mix}$.
Hence,

$$t_{mix} > \frac{n\log n}{2} - 2n.$$

Definition 5.65. Given a distribution π on χ and $1 \leq p \leq \infty$, the $l^p(\pi)$ norm of a function $f : \chi \to \mathbb{R}$ is defined as

$$||f||_p := \begin{cases} [\sum_{y \in \chi} |f(y)|^p \pi(y)]^{\frac{1}{p}} & 1 \leq p < \infty \\ \max_{y \in \chi} |f(y)| & p = \infty \end{cases} \tag{5.37}$$

Definition 5.66. The l^p distance $d^{(p)}(t)$ is defined as

$$d^{(p)}(t) = \max_x \left|\left| \frac{P^t(x,.) - \pi(.)}{\pi(.)} \right|\right|.$$

Proposition 5.67. *For a reversible markov chain,*

$$\left|\left| \frac{P^t(x,.) - \pi(.)}{\pi(.)} \right|\right|^2_{L^2(\pi)} = \frac{1}{\pi(x)} P^{2t}(x,x) - 1.$$

Proof.

$$\begin{aligned} \left|\left| \frac{P^t(x,.) - \pi(.)}{\pi(.)} \right|\right|^2_{L^2(\pi)} &= \sum_y \frac{[p^t(x,y) - \pi(y)]^2}{\pi(y)} \\ &= \frac{[P^t(x,y)]^2 - 2P^t(x,y)\pi(y) + \pi^2(y)}{\pi(y)} \\ &= \left(\sum_y \frac{[P^t(x,y)]^2}{\pi(y)} \right) - 1 \\ &= \sum_y \frac{P^t(x,y)P^t(y,x)}{\pi(x)} - 1 \\ &= \frac{1}{\pi(x)} P^{2t}(x,x) - 1. \end{aligned} \tag{5.38}$$

□

Markov Chains Mixing Times **October 28, 2022**

Lecture 6: Lower bounds on Mixing times

Lecturer: Yuval Peres Week: 3

So far we have discussed about finding upper bounds on t_{mix}. It is natural to ask if we can give lower bounds on t_{mix}.

Example 6.68. Row Operation on Matrices mod 2
$G = Gl_n(2) =$ invertible $n \times n$ Matrices mod 2.
Pick any two rows uniformly, say ith and jth rows where $i \neq j$ in $\{1, \ldots, n\}$.
Add row i to row j (mod 2). This chain is irreducible and we can give a lower bound to the mixing time of this Markov chain.

This is where we introduce ***Counting Bounds.*** Let (X_t) be a Markov chain with irreducible and aperiodic transition matrix P on the state space χ, and suppose that the stationary distribution π is uniform on χ. Define $d_{out}(x) = |\{y : P(x, y) > 0\}|$ to be the number of states accessible from x in one step, and let

$$\Delta := \max_{x \in \chi} d_{out}(x).$$

Denote by χ_t^x the set of states accessible from x in exactly t steps, and observe that $|\chi_t^x| \leq \Delta^t$. If $\Delta^t < (1 - \epsilon)|\chi|$, then from the definition of total variation distance we have that

$$||P^t(x, .) - \pi||_{TV} \geq P^t(x, \chi_t^x) - \pi(\chi_t^x) \geq 1 - \frac{\Delta^t}{|\chi|} > \epsilon. \tag{6.39}$$

This implies that

$$t_{mix}(\epsilon) \geq \frac{\log(|\chi|(1 - \epsilon))}{\log \Delta}. \tag{6.40}$$

Now in the above example 6.68, we can think of the markov chain as a graph of degree atmost $n(n-1)$. And as the cardinality of the state space is 2^{n^2}, we can conclude that:

$$t_{mix}(\epsilon) \geq \frac{\log(2^{n^2}(1 - \epsilon))}{\log(n(n-1))}.$$

And taking $\epsilon = 1/4$, we get that $t_{mix} > \frac{cn^2}{\log n}$ for some constant c.

Definition 6.69. The ***edge measure*** Q is defined by:

$$Q(x, y) := \pi(x)P(x, y), \qquad Q(A, B) = \sum_{x \in A, y \in B} Q(x, y). \tag{6.41}$$

Here, $Q(A, B)$ is the probability of moving from A to B in one step when starting from the stationary distribution.

Proposition 6.70. *$Q(S, S^c) = Q(S^c, S)$ for any $S \subset \chi$.*

Proof.

$$\begin{aligned} Q(S, S^c) &= \sum_{x\in S}\sum_{y\in S^c} \pi(x)P(x,y) \\ &= \sum_{x\in S} \pi(x) - \sum_{x\in S}\sum_{y\in S} \pi(x)P(x,y) \\ &= \sum_{x\in S} \pi(x) - \sum_{y\in S}(\pi(y) - \sum_{x\in S^c} \pi(x)P(x,y)) \\ &= \sum_{y\in S}\sum_{x\in S^c} \pi(x)P(x,y) = Q(S^c, S) \end{aligned} \tag{6.42}$$

□

Definition 6.71. The ***bottleneck ratio*** of the set S is defined to be

$$\Phi(S) := \frac{Q(S, S^c)}{\pi(S)}, \tag{6.43}$$

while the bottleneck ratio of the whole chain is:

$$\Phi_* = \max_{S:\pi(S)\le \frac{1}{2}} \Phi(S). \tag{6.44}$$

For simple random walk on graph with vertices χ and edge set E, the bottleneck ratio becomes

$$\Phi(S) = \frac{|\partial S|}{\sum_{x\in S} deg(x)}. \tag{6.45}$$

Theorem 6.72. *If Φ_* is the bottleneck ratio, then*

$$t_{mix} = t_{mix}(1/4) \ge \frac{1}{4\Phi_*}. \tag{6.46}$$

Proof. When $X_0 \sim \pi$, $\mathbb{P}(X_0 \in A, X_1 \notin A) = \sum_{x\in A}\sum_{y\notin A} \pi(x)P(x,y) = Q(A, A^c)$.

$$\mathbb{P}(X_0 \in A, X_t \notin A) \le \sum_{j=1}^{t} \mathbb{P}(X_{j-1} \in A, X_j \notin A) = tQ(A, A^c) \tag{6.47}$$

This implies that

$$\begin{aligned} &\mathbb{P}(X_t \notin A | X_0 \in A) \le t\Phi(A) \\ &\Rightarrow \exists\ X \in A \text{ such that } \mathbb{P}_x(X_t \notin A) \le t\Phi(A). \end{aligned} \tag{6.48}$$

Therefore,

$$d(t) \ge 1 - t\Phi(A) - \pi(A).$$

If $\pi(A) \le 1/2$ and $t < 1/[4\Phi(A)]$, then $d(t) > 1/4$. Therefore, $t_{mix} \ge 1/[4\Phi(A)]$. Maximizing over A with $\pi(A) \le 1/2$ completes the proof.

□