

Short Course: Markov chains & mixing times

Lecture 1: Yuval Peres, Microsoft Research

Definition of Markov chains

- A process $(X_t)_{t=0,1,\dots}$ is a (time-homogeneous) *Markov chain* if

$$\mathbf{P}(X_{k+1} = y \mid X_0 = x_0, \dots, X_{k-1} = x_{k-1}, X_k = x) = p(x, y),$$

where $p(x, y) \geq 0$ and $\sum_y p(x, y) = 1$ for all x .

- The *transition kernel* $P = (p(x, y))_{x,y}$.
- The *stationary measure* π satisfies that $\pi P = \pi$.
- Aperiodic: $\gcd\{t \geq 1 : P^t(x, x) > 0\} = 1$ for all x .
- Irreducible: for all x, y , there exists $t \in \mathbf{N}$ such that $P^t(x, y) > 0$.
- Reversible: $\pi(x)p(x, y) = \pi(y)p(y, x)$ for all x, y .

Most examples in this talk are aperiodic, irreducible and reversible.

Classical theory vs. modern focus

- An important feature of aperiodic and irreducible (finite) Markov chains: the distribution at time t converges to the (unique) stationary distribution as $t \rightarrow \infty$.
- Classical theory: Fix the chain and study the rate of convergence of the distribution at time t to stationarity, as $t \rightarrow \infty$.
- Modern focus: Fix the target distance to stationarity and study the asymptotics of the required time to reach that target, for a family of chains as the size goes to infinity. (Aldous, Diaconis, ...)

Motivations

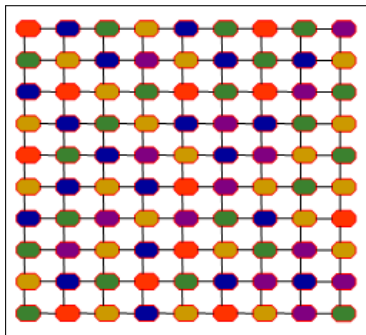
- Statistical physics.
 - Monte Carlo simulation.
 - Models of dynamical processes.
 - Deep connections between the convergence rate and the spatial properties of the physical model.
- Biology.
 - Models of DNA evolution.
 - A much simplified chain of Durrett: Given a permutation, take a random segment of fixed length and reverse it.
- Computer science.
 - Sampling.
 - Approximate counting of combinatorial structures.
- Card shuffling.

...

More on sampling

- Random permutation: an example easy to sample directly.
 - Knuth Algorithm: For $i = 1, \dots, n - 1$, select U_i out of $\{i, \dots, n\}$ uniformly at random and swap the two elements at positions i and U_i .
- Many other models are hard to sample directly.
 - Ising model : more in later lectures.
 - Coloring model: a uniform measure on all the proper colorings of a graph with q colors, where in a proper coloring, any two adjacent vertices are assigned different colors.

An illustration of coloring model



Total variation distance

For two distributions μ and ν on a (discrete) space Ω , the total variation distance $\|\mu - \nu\|_{\text{TV}}$ is defined in the following equivalent ways:

- $\|\mu - \nu\|_{\text{TV}} := \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)|.$
- $\|\mu - \nu\|_{\text{TV}} := \sup_{A \subset \Omega} (\mu(A) - \nu(A)).$
- For all couplings (X, Y) where $X \sim \mu$ and $Y \sim \nu$, we have $\mathbf{P}(X \neq Y) \geq \|\mu - \nu\|_{\text{TV}}$. Furthermore, there exists a coupling such that equality holds.

An important feature: advancing a Markov chain can only decrease the total variation distance. That is,

$$\|\mu P - \nu P\|_{\text{TV}} \leq \|\mu - \nu\|_{\text{TV}}.$$

Mixing time and its first property

- We are interested in the decay of $d(t) := \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{\text{TV}}$.
- $\|\mu P - \nu P\| \leq \|\mu - \nu\| \Rightarrow d(t)$ is decreasing in t .
- The *mixing time* $t_{\text{MIX}}(\epsilon) := \min\{t \geq 0 : d(t) \leq \epsilon\}$, for some $\epsilon \in (0, 1)$. Furthermore, $t_{\text{MIX}} := t_{\text{MIX}}(1/4)$ by convention.
- Since $d(t + s) \leq 2d(t)d(s)$, we have $t_{\text{MIX}}(2^{-k}) \leq k \cdot t_{\text{MIX}}$.

Mixing time v.s. relaxation time

- The transition kernel P always has 1 as the largest eigenvalue.
- Define $\lambda_\star := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P\}$ and the spectral gap $1 - \lambda_\star$. Furthermore, define the relaxation time $t_{\text{REL}} := 1/(1 - \lambda_\star)$.
- An important relation between relaxation time and mixing time:

$$(t_{\text{REL}} - 1) \log(1/2\epsilon) \leq t_{\text{MIX}}(\epsilon) \leq \log(1/\epsilon\pi_{\min})t_{\text{REL}},$$

where $\pi_{\min} = \min_{x \in \Omega} \pi(x)$.

- Seemingly, t_{REL} and t_{MIX} are roughly the same. However, the possible difference of t_{REL} and t_{MIX} can reveal deep features of a Markov chain.

Random walks on cycles and hypercubes

A chain is called lazy if $P(x, x) \geq 1/2$ for all x . All the eigenvalues of a lazy chain are non-negative.

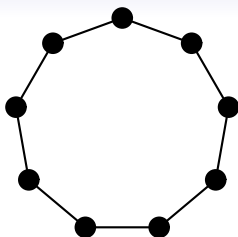
- Lazy random walk on a cycle of length n .
 - The second eigenvalue $\lambda_2 = (1 + \cos(2\pi/n))/2$ and the corresponding eigenfunction is $f_2(k) = \cos(4\pi k/n)$.
 - The relaxation time $t_{\text{REL}} \asymp n^2$.
 - The mixing time $t_{\text{MIX}} \asymp n^2$.
- Lazy random walk on hypercube $\{-1, 1\}^n$.
 - The second eigenvalue $\lambda_2 = 1 - \frac{1}{n}$ and one of the corresponding eigenfunctions is $f_2 = \sum_{i=1}^n x_i$, where $x_i \in \{-1, 1\}$ denotes the i -th coordinate.
 - The relaxation time $t_{\text{REL}} = n$.
 - The mixing time $t_{\text{MIX}} \asymp n \log n$.
- In the cycle, the relaxation time and the mixing time have the same order, while in the hypercube the mixing time is of larger magnitude.

Couplings of Markov Chains

Let P be a transition matrix for a Markov chain. A coupling of a P -Markov-chain started at x and a P -Markov-chain started at y is a sequence $\{(X_n, Y_n)\}_{n=0}^{\infty}$ such that

- all variables X_n and Y_n are defined on the same probability space,
- $\{X_n\}$ is a P -Markov-chain started at x , and
- $\{Y_n\}$ is a P -Markov-chain started at y .

Example: The lazy random walk on the n -cycle.



- This chain remains at its current position with probability $1/2$, and moves to each of the two adjacent site with probability $1/4$.
- Can couple the chains started from x and y as follows:
 - Flip a fair coin to decide if the X -chain moves or the Y -chain moves,
 - Move the selected chain to one of its two neighboring sites, chosen with equal probability.
- Both the x -particle and the y -particle are performing lazy simple random walks on the n -cycle.

Mixing and Coupling

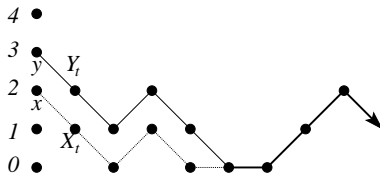
- Let $(X_t, Y_t)_{t=0}^{\infty}$ be a coupling of a P -chain started from x and a P -chain started at y .
- Let

$$\tau := \min\{t \geq 0 : X_t = Y_t\}.$$

If the coupling is Markovian (and we will only consider those), it can always be redefined so that

$$X_t = Y_t \text{ for } t \geq \tau,$$

So, let us assume this.



- The pair (X_t, Y_t) (for given t) is a coupling of $P^t(x, \cdot)$ and $P^t(y, \cdot)$.

Mixing and Coupling

- Since X_t has distribution $P^t(x, \cdot)$ and Y_t has distribution $P^t(y, \cdot)$, using the coupling characterization of total variation distance,

$$\mathbf{P}(\tau > t) = \mathbf{P}(X_t \neq Y_t) \geq d_{\text{TV}}(\mathbf{P}^t(x, \cdot), \mathbf{P}^t(y, \cdot)).$$

- Combined with the inequality

$$d_{\text{TV}}(P^t(x, \cdot), \pi) \leq \max_{y \in \Omega} d_{\text{TV}}(P^t(x, \cdot), P^t(y, \cdot)),$$

if there is a coupling (X_t, Y_t) for every pair of initial states (x, y) , then this shows that

$$\begin{aligned} d(t) &= \max_{x \in \Omega} d_{\text{TV}}(P^t(x, \cdot), \pi) \leq \max_{x, y} d_{\text{TV}}(P^t(x, \cdot), P^t(y, \cdot)) \\ &\leq \max_{x, y} \mathbf{P}_{x, y}(\tau > t). \end{aligned}$$

Mixing for lazy random walk on the n -cycle

- Use the coupling which selects at each move one of the “particles” at random; the chosen particle is equally likely to move clockwise as counter-clockwise.
- The clockwise difference between the particles, $\{D_t\}$, is a simple random walk on $\{0, 1, \dots, n\}$.
- When $D_t \in \{0, n\}$, the two particles have collided.
- If τ is the time until a simple random walk on $\{0, 1, \dots, n\}$ hits an endpoint when started at k , then

$$\mathbf{E}_k \tau = k(n - k) \leq \frac{n^2}{4}.$$

RW on n -cycle, continued

- By Markov's inequality,

$$\mathbf{P}(\tau > t) \leq \frac{\mathbf{E}\tau}{t} \leq \frac{n^2}{4t}.$$

- Using the coupling inequality,

$$d(t) \leq \max_{x,y} \mathbf{P}(\tau > t) \leq \frac{n^2}{4t}.$$

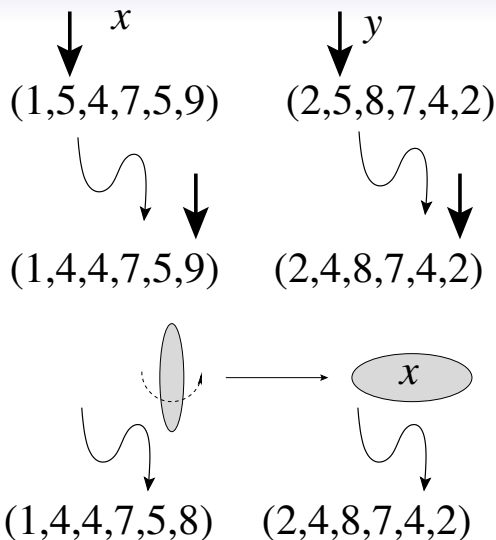
- Taking $t \geq n^2$ yields $d(t) \leq 1/4$, whence

$$t_{\text{MIX}} \leq n^2.$$

Random Walk on d -dimensional Torus



- $\Omega = (\mathbf{Z}/n\mathbf{Z})^d$. The walk remains at current position with probability $1/2$.
- Couple two particles as follows:
 - Select among the d coordinates at random.
 - If the particles agree in the selected coordinate, move the walks together in this coordinate. Thus both walks together either make a clockwise move, a counterclockwise move, or remain put.
 - If the particles disagree in the chosen coordinate, flip a coin to decide which walker will move. Move the selected walk either clockwise or counterclockwise, each with probability $1/2$.



- Consider the clockwise difference between the i -th coordinate of the two particles. It moves at rate $1/d$, and when it does move, it performs simple random walk on $\{0, 1, \dots, n\}$, with absorption at 0 and n . Thus the expected time to couple the i -th coordinate is bounded above by $dn^2/4$.
- Since there are d coordinates, the expected time for all of them to couple is not more than

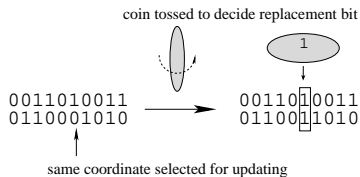
$$d \times d \frac{n^2}{4} = \frac{d^2 n^2}{4}.$$

- By the coupling theorem,

$$t_{\text{MIX}} \leq d^2 n^2.$$

Exercise: Improve the d -dependence in this bound.

RW on hypercube



- Consider the lazy random walk on the hypercube $\{0, 1\}^n$. Sites are neighbors if they differ in exactly one coordinate.
- To update the two walks, first pick a coordinate at random. *The same coordinate is used for both walks.*
- Toss a coin to determine if the bit at the chosen coordinate is replaced by a 1 or a 0. *The same bit is used for both walks.*
- No matter the initial positions of the two walks, when every coordinate has been selected, the two walks agree.
- Reduces to a "coupon collector's" problem: how many times must a coordinate be drawn at random before every coordinate is chosen?

Coupon collector

- Let $A_k(t)$ be the event that the k -th coupon has *not* been collected by time t .
- Observe

$$\mathbf{P}(A_k(t)) = \left(1 - \frac{1}{n}\right)^t \leq e^{-t/n}.$$

- Consequently,

$$\mathbf{P}\left(\bigcup_{k=1}^n A_k(t)\right) \leq \sum_{k=1}^n e^{-t/n} = ne^{-t/n}.$$

- In other words, if τ is the time until all coupons have been collected,

$$\mathbf{P}(\tau > n \log n + cn) = \mathbf{P}\left(\bigcup_{k=1}^n A_k(n \log n + cn)\right) \leq e^{-c}.$$

Returning to the hypercube,

$$d(n \log n + cn) \leq \mathbf{P}(\tau > n \log n + cn) \leq e^{-c},$$

whence

$$t_{\text{MIX}}(\epsilon) \leq n \log n + n \log(1/\epsilon).$$

Strong stationary times

- Random mapping representation of Markov chain P : an i.i.d. sequence (Z_t) and a map f such that $X_t = f(X_{t-1}, Z_t)$ with $X_0 = x$.
- The sequence (Z_t) has more information than the chain (X_t) .
- A *randomized stopping time* is a stopping time for the sequence (Z_t) . Not necessarily a stopping time of (X_t) !
- A *strong stationary time* is a randomized stopping time τ such that X_τ has distribution π and is independent of τ . That is,

$$\mathbf{P}_x(\tau = t, X_\tau = y) = \mathbf{P}_x(\tau = t)\pi(y).$$

Proposition

If τ is a strong stationary time, then

$$d(t) = \max_x \|P^t(x, \cdot) - \pi\|_{\text{TV}} \leq \max_x \mathbf{P}_x(\tau > t).$$

Top-to-random shuffle

- *Top-to-random shuffle*: take the top card and insert it uniformly at random in the deck.
- Strong stationary time τ_{top} : the time one move after the the first occasion when the original bottom card has moved to the top of the deck.
- τ_{top} is the same as the coupon collector's time and hence $\mathbf{P}(\tau_{\text{top}} > n \log n + cn) \leq e^{-c}$.
- $t_{\text{MIX}}(\epsilon) \leq n \log n + \log(\epsilon^{-1})n$.

Random walk on the hypercube - revisited

- Lazy walk on a hypercube can be viewed as a dynamics on $\{-1, 1\}^n$, where at each step a coordinate is selected and updated uniformly at random.
- The strong stationary time τ_{refresh} : the first time when all the coordinates have been selected at least once for updating.
- τ_{refresh} is the same as the coupon collector's time.
- $t_{\text{MIX}}(\epsilon) \leq n \log n + \log(\epsilon^{-1})n$.
- Is this tight?

Answer: In fact $t_{\text{MIX}} = \frac{1}{2}n \log n + O(n)$ and there is a cutoff.