6.896 Probability and Computation	February 14 , 2011
Lecture 4	
Lecturer: Constantinos Daskalakis	*Scribe: Georgios Papadopoulos*

NOTE: The content of these notes has not been formally reviewed by the lecturer. It is recommended that they are read critically.

1 Introduction

In the previous lecture we proved that an irreducible and aperiodic Markov chain converges to its unique stationary distribution. Our proof used the idea of coupling two copies of the Markov Chain that start from arbitrary states x and y, move on the state space independently from each other, until they meet and stick together forever. We argued that such a sticky coupling satisfies

$$\max_x ||p_x^t - \pi||_{TV} \leq \max_{x,y} Pr[T_{x,y} > t], \tag{1}$$

where $T_{x,y}$ is a random variable that equals the time that it takes the two coupled copies that start at x, y to meet, i.e. $T_{x,y} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}$.

Today we move one step forward and examine more general couplings for showing convergence to the stationary distribution. We are also concerned with bounding the rate of convergence to the stationary distribution.

Definition 1. *A* **coupling** *of a Markov chain is a pair process $(X_t, Y_t)_t$ on $\mathbb{R}^{|\Omega\times\Omega|}$ such that:*

- *$(X_t, .)_t$ and $(., Y_t)_t$ are faithful copies of the Markov chain,*
- *$X_t = Y_t => X_{t+1} = Y_{t+1}$, i.e. if the two copies are on the same state at some time t they are on the same state also at the next time, $t+1$.*

Couplings are useful because a comparison between distributions is reduced to a comparison between random variables, and therefore upper bounds on distance can be obtained easily.

Lemma 1. *$\Delta(t) \leq \max_{x,y} Pr[T_{x,y} > t]$, where $\Delta(t) := \max_x ||p_x^t - \pi||_{TV}$ is the worst possible distance from the stationary distribution after t steps, and $T_{x,y} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}$.*

Proof: Recall from last lecture that

$$\Delta(t) \leq D(t),$$

where

$$D(t) = \max_{x,y} ||p_x^t - p_y^t||_{TV}.$$

Using these and the coupling lemma we have:

$$\begin{aligned} \Delta(t) &\leq \max_{x,y} ||p_x^t - p_y^t|| \\ &\leq \max_{x,y} Pr[X_t \neq Y_t | X_0 = x, Y_0 = y] \\ &= \max_{x,y} Pr[T_{x,y} > t] \end{aligned} \tag{2}$$

□

The above lemma provides an upper bound on the distance of the Markov Chain from its stationary distribution in t steps.

2 Example: Simple Random Walk on $\{0,1\}^n$

The n-dimensional hypercube is a graph whose vertices are the binary n-tuples $\{0,1\}^n$. Two vertices are connected by an edge when they differ in exactly one coordinate. We define a random walk on a hypercube as follows:

- with probability $p = 0.5$ we do not move;
- with probability $p = 0.5$ we pick a coordinate uniformly at random and we flip its value.

This Markov Chain is aperiodic and irreducible. Hence it convergences to its unique stationary distribution; in this case, this is the uniform distribution over $\{0,1\}^n$ (why?).

To bound the mixing time of this chain consider two copies $(X_t)_t$ and $(Y_t)_t$ starting at arbitrary vertices of the cube and couple them as follows at all t:

- pick a coordinate c uniformly at random;
- if X_t and Y_t are equal at coordinate c, then with probability $p = 0.5$ do nothing in both X_t and Y_t, and with probability 0.5 flip the value of coordinate c in both X_t and Y_t;
- if X_t and Y_t differ at coordinate c, then pick value 0 or 1 uniformly at random, and set coordinate c to that value in both X_t and Y_t.

Our goal is to study the time $T_{x,y}$. Notice that this is upper bounded by the Coupon Collector's time for n coupons, and therefore

$$Pr[T_{x,y} > n\ln(n) + cn] \leq e^{-c}.$$

Applying Lemma 1 we have:

$$\Delta(n\log(n) + cn) \leq Pr[T_{x,y} > n\log(n) + cn] \leq e^{-c}.$$

Setting $\tau(\epsilon) := n(\log(n) + \log(1/\epsilon))$ we get

$$\Delta(\tau(\epsilon)) \leq \epsilon.$$

Remark: The fact that our chain was lazy with probability 1/2 was very convenient to design our coupling. How would you modify the coupling if the probability of being lazy was smaller than 1/2? What would be the effect on the mixing time?

3 Rates of Convergence

In this section, we study the rate of convergence of a Markov Chain to its stationary distribution π. We show first that the convergence is monotonic and proceed to show that the convergence is exponential. Recall the following definitions from last lecture:

$$\Delta(t) = \max_x ||p_x^t - \pi||_{TV}; \tag{3}$$

$$D(t) = \max_{x,y} ||p_x^t - p_y^t||_{TV}. \tag{4}$$

Given these definitions and the triangle inequality we have shown that:

$$\Delta(t) \leq D(t) \leq 2\Delta(t) \tag{5}$$

Lemma 2. *$\Delta(t)$ is non-increasing in t.*

Proof: Let us consider two faithful copies of an aperiodic, irreducible Markov chain $(X_t)_t$ and $(Y_t)_t$, where $(X_t)_t$ starts from an arbitrary initial state $X_0 = x$ and $(Y_t)_t$ starts from a state drawn from the stationary distribution π. (In particular, for all t, Y_t is distributed according to π).

Now pick an arbitrary time T. Using the coupling lemma we couple the evolution of X_t with Y_t so that at time T we have:

$$||X_T - Y_T||_{TV} = Pr[X_T \neq Y_T]. \tag{6}$$

Conditioning on X_T and Y_T we couple the next step $(T+1)$ as follows:

- if $X_T = Y_T$ then sample X_{T+1} according to the chain and set $Y_{T+1} = X_{T+1}$;
- otherwise, the X_{T+1} and Y_{T+1} are sampled independently.

Using the coupling lemma once more we have:

$$||X_{T+1} - Y_{T+1}||_{TV} \leq Pr[X_{T+1} \neq Y_{T+1}] \leq Pr[X_T \neq Y_T] = ||X_T - Y_T||_{TV}. \tag{7}$$

Given that the Y_t starts in the stationary distribution:

$$||X_{T+1} - Y_{T+1}||_{TV} = ||P_x^{T+1} - \pi||_{TV} \tag{8}$$

$$||X_T - Y_T||_{TV} = ||P_x^T - \pi||_{TV} \tag{9}$$

and therefore from the above $\Delta(T+1) = ||X_{T+1} - Y_{T+1}||_{TV} \leq ||X_T - Y_T||_{TV} = \Delta(T)$. Since T was arbitrary this concludes the proof. □

Definition 2. *The* **mixing time** *of an irreducible, aperiodic Markov Chain is the first time beyond which the distribution of the Markov Chain is guaranteed to stay within* $1/2e$ *from its stationary distribution no matter where it started, i.e.,*

$$\tau_{mix} = \min\left\{t : \Delta(t) \leq \frac{1}{2e}\right\}; \tag{10}$$

in general, we define:

$$\tau(\epsilon) = \min\{t : \Delta(t) \leq \epsilon\}. \tag{11}$$

Notice that the Fundamental Theorem of Markov Chains and the above lemma imply that these times exist and are finite. The following lemma shows that τ_{mix} captures the the rate of convergence to the stationary distribution.

Lemma 3. $\Delta(t) \leq \exp\left(-\left\lfloor \frac{t}{\tau_{mix}} \right\rfloor\right)$, *i.e. the distance from the stationary distribution reduces exponentially with rate* τ_{mix}.

Proof: We consider two faithful copies of an aperiodic and irreducible Markov chain $(X_t)_t$ and $(Y_t)_t$, starting at arbitrary states x and y. By the coupling lemma, we can couple the first τ steps of the chains so that:

$$||p_x^\tau - p_y^\tau||_{TV} = Pr[X_t \neq Y_t] \tag{12}$$

Conditioning on the first τ steps of the chain, we construct a coupling for the next τ' time steps as follows:

- if $X_\tau = Y_\tau$ then $X_{\tau+s} = Y_{\tau+s}$ for all $s = 1, 2, 3, \ldots, \tau'$;
- otherwise, suppose $X_t = x'$ and $Y_t = y'$, where $x' \neq y'$; use the coupling lemma to couple the chains $(X_{\tau+s})_s$ and $(Y_{\tau+s})_s$ so that

$$||p_{x'}^{\tau'} - p_{y'}^{\tau'}||_{\text{TV}} = \Pr[X_{\tau+\tau'} \neq Y_{\tau+\tau'} | X_\tau = x' \neq y' = Y_\tau].$$

Applying the coupling lemma once more we have:

$$\begin{aligned}
||p_x^{\tau+\tau'} - p_y^{\tau+\tau'}||_{TV} &\leq Pr[X_{\tau+\tau'} \neq Y_{\tau+\tau'} | X_\tau = x, Y_\tau = y] \\
&\leq Pr[X_\tau \neq Y_\tau] Pr[X_{\tau+\tau'} \neq Y_{\tau+\tau'} | X_\tau \neq Y_\tau] \\
&\leq ||p_x^\tau - p_y^\tau||_{TV} \max_{x',y'} ||p_{x'}^{\tau'} - p_{y'}^{\tau'}||_{TV} \\
&\leq D(\tau) D(\tau').
\end{aligned}$$

Since the above is true for an arbitrary pair of states x, y:

$$D(\tau + \tau') = \max_{x,y} ||p_x^{\tau+\tau'} - p_y^{\tau+\tau'}||_{TV} \leq D(\tau) D(\tau').$$

Now taking $t = k \cdot \tau_{\text{mix}}$, $k \geq 1$, we get:

$$\Delta(k \cdot \tau_{\text{mix}}) \leq D(k \cdot \tau_{\text{mix}}) \leq D(\tau_{\text{mix}})^k \leq (2\Delta(\tau_{\text{mix}}))^k \leq e^{-k}. \tag{13}$$

□

4 Shuffle Cards: Random Transpositions

Recall the random transposition shuffle from Lecture 1:

- Pick two cards C and C' uniformly at random;
- switch them;
- repeat.

It is easy to see that an equivalent description of the transposition shuffle is the following:

- pick a card C and a position P uniformly at random;
- exchange card C with whatever card is at position P;
- repeat.

To bound the mixing time, we consider 2 copies of the Markov Chain $(X_t)_t$ and $(Y_t)_t$ and couple them to pick the same card C and position P at all time steps t. Denoting by $d(t)$ the number of positions where X_t and Y_t differ, we analyze our coupling as follows:

- if card C is at the same position in X_t, Y_t, then we have $d(t+1) = d(t)$;
- if card C is at different positions in X_t and Y_t then we distinguish the subcases:
  - if the card at position P is the same in X_t and Y_t, then $d(t+1) = d(t)$;
  - otherwise, $d(t+1) \leq d(t) - 1$. (In what case can I improve by more than 1?)

Hence,

$$Pr[d_{t+1} < d_t] = \left(\frac{d_t}{n}\right)^2. \tag{14}$$

Hence, the expected time to decrease the distance by 1 is at most $(\frac{n}{d_t})^2$, and therefore, we can bound

$$E[T_{x,y}] \leq \sum_{d=1}^{n} \left(\frac{n}{d}\right)^2 = c \cdot n^2,$$

for some constant c. Using Markov's inequality we have, for all $c' > 0$:

$$Pr[T_{x,y} \geq c'n^2] < \frac{c}{c'}. \tag{15}$$

Choosing $c' = 2e \cdot c$ we have

$$Pr[T_{x,y} \geq 2e \cdot cn^2] < \frac{1}{2e}.$$

Hence,

$$\tau_{\text{mix}} = O(n^2).$$

2 points problem: Design a better coupling that gives $\tau_{\text{mix}} = O(n \log(n))$.