STATS 217: Introduction to Stochastic Processes I

Lecture 20

Convergence theorem

- Last time, we proved the convergence theorem for irreducible, aperiodic, finite-state Markov chains.
- Let $(X_n)_{n\geq 0}$ be a DTMC on S with transition matrix P. Suppose that P is irreducible and aperiodic with unique stationary distribution π.
- Let

$$\Delta(n) = \max_{x\in S} \Delta_x(n) = \max_{x\in S} \mathrm{TV}(X_n \mid X_0 = x, \pi).$$

- There exist constants $\epsilon > 0$ and $C > 0$ (depending on P) such that

$$\Delta(n) \leq C \cdot (1-\epsilon)^n.$$

Sub-multiplicativity

- In fact, we worked with the quantities

$$D_{x,y}(n) = \mathrm{TV}(X_n \mid X_0 = x, X_n \mid X_0 = y)$$

  and

$$D(n) = \max_{x,y \in S} D_{x,y}(n).$$

- We showed that

$$\Delta(n) \le D(n) \le 2\Delta(n)$$

  for all integers $n \ge 0$ and that for any integers $s, t \ge 0$,

$$D(s+t) \le D(s)D(t).$$

Mixing time

- For $\varepsilon \in [0,1]$, define the ε-**mixing time** of the chain to be

$$\tau_{\text{mix}}(\varepsilon) := \min\{t : \Delta(t) \leq \varepsilon\}.$$

- Since $\Delta(n+1) \leq \Delta(n)$ for all $n \geq 0$, it follows that for any $t \geq \tau_{\text{mix}}(\varepsilon)$ and for any $x \in S$,

$$\text{TV}(X_t \mid X_0 = x, \pi) \leq \varepsilon.$$

- It is convenient to define

$$\tau_{\text{mix}} := \tau_{\text{mix}}(1/4).$$

- The choice of the constant $1/4$ is not important and can be replaced by another constant which is strictly smaller than $1/2$.

Mixing time

- The reason that it's often enough to look only at τ_{mix} is because for any $\varepsilon \in (0, 1)$,

$$\tau_{\mathrm{mix}}(\varepsilon) \le \lceil \log_2 \varepsilon^{-1} \rceil \tau_{\mathrm{mix}}.$$

- Indeed,

$$\begin{aligned}
\Delta(\lceil \log_2 \varepsilon^{-1} \rceil \tau_{\mathrm{mix}}) &\le D(\lceil \log_2 \varepsilon^{-1} \rceil \cdot \tau_{\mathrm{mix}}) \\
&\le D(\tau_{\mathrm{mix}})^{\lceil \log_2 \varepsilon^{-1} \rceil} \\
&\le (2\Delta(\tau_{\mathrm{mix}}))^{\lceil \log_2 \varepsilon^{-1} \rceil} \\
&\le 2^{-\lceil \log_2 \varepsilon^{-1} \rceil} \\
&\le \varepsilon.
\end{aligned}$$

Coupling of Markov chains

- Consider a transition matrix P on a finite state space S.
- A **coupling of Markov chains** with transition matrix P and initial distributions μ and ν is a process

$$(\widehat{X}_t, \widehat{Y}_t)_{t=0}^{\infty}$$

such that for all $t \geq 0$,

$$\widehat{X}_t \sim (X_t \mid X_0 \sim \mu)$$
$$\widehat{Y}_t \sim (X_t \mid X_0 \sim \nu)\,,$$

and such that

$$\widehat{X}_t = \widehat{Y}_t \implies \widehat{X}_{t+1} = \widehat{Y}_{t+1}.$$

- We have already seen couplings of Markov chains in our proof of the convergence theorem

Coupling of Markov chains

- As we will soon see, couplings of Markov chains are a useful tool to bound the mixing time in applications.
- This is due to the following: Let $(\widehat{X}_t, \widehat{Y}_t)$ be a coupling of two Markov chains with transition matrix P and with $\widehat{X}_0 = x$, $\widehat{Y}_0 = y$. Let

$$\tau_{\text{couple}} := \min\{t : \widehat{X}_t = \widehat{Y}_t\}.$$

Recall that

$$D_{x,y}(n) = \mathrm{TV}(X_n \mid X_0 = x, X_n \mid X_0 = y).$$

Then,

$$D_{x,y}(n) \leq \mathbb{P}[\tau_{\text{couple}} \geq n].$$

Coupling of Markov chains

- The proof is a direct application of the coupling lemma.
- Indeed, since $\widehat{X}_n \sim X_n \mid X_0 = x$ and $\widehat{Y}_n \sim X_n \mid X_0 = y$, we have

$$D_{x,y}(n) \leq \mathbb{P}[\widehat{X}_n \neq \widehat{Y}_n] \leq \mathbb{P}[\tau_{\text{couple}} \geq n].$$

- Therefore, by Markov's inequality,

$$D_{x,y}(4 \cdot \mathbb{E}[\tau_{\text{couple}}]) \leq \mathbb{P}[\tau_{\text{couple}} \geq 4 \cdot \mathbb{E}[\tau_{\text{couple}}]] \leq \frac{1}{4}.$$

Example: Lazy random walk on the hypercube

- $S = \{0, 1\}^n$.
- The transitions are given as follows. Suppose the current state is x. With probability $1/2$, the chain remains at x; with probability $1/2$, it moves uniformly to one of the n possible vectors y which differ from x in exactly one coordinate.
- The transition matrix is clearly aperiodic and irreducible, and we have seen that the unique stationary distribution is the uniform distribution on $\{0, 1\}^n$.
- Here is an equivalent description of the transitions: suppose the current state is x. We choose a coordinate $i \in \{1, \ldots, n\}$ uniformly at random and an unbiased bit $b \in \{0, 1\}$, also uniformly at random, and independent of the coordinate i.
- Then, we set the value of coordinate i to b and keep all other coordinates unchanged.

Example: Lazy random walk on the hypercube

- Given this alternate description, there is a natural choice of coupling: for the two chains started from x and y, use the same i and b at every step.
- Let τ denote the first time that each coordinate i has been chosen to be updated. Then, clearly, $\widehat{X}_\tau = \widehat{Y}_\tau$.
- Moreover, τ is exactly the first time to collect all n coupons in the coupon-collector problem and

$$\mathbb{P}[\tau > t] \leq n\left(1 - \frac{1}{n}\right)^t \leq ne^{-t/n},$$

which gives $\tau_{\text{mix}} \leq n \log n + n \log(1/4)$.

Example: Lazy random walk on the cycle

- The states of the n-cycle can be identified with $\mathbb{Z}_n$, the integers modulo n.
- The transitions are given as follows. Suppose that the current state is x. With probability $1/2$, the chain remains at x; with probability $p/2$, it moves to $x+1$; with probability $q/2$, it moves to $x-1$. Here, $p+q=1$.
- Here is a natural choice of coupling: start the two chains at x and y. At each step, flip an unbiased coin. If the coin lands heads, then the x-chain stays put, and the y chain moves $+1$ with probability p and -1 with probability q. If the coin lands tails, then the y-chain stays put, and the x chain moves $+1$ with probability p and -1 with probability q.

Example: Lazy random walk on the cycle

- Let dist_t denote the (clockwise) distance between the states of the two chains at time t.
- Then, $(\text{dist}_t)_{t \geq 0}$ is a simple symmetric random walk on $\{0, \ldots, n\}$ with absorbing states 0 and n.
- From Gambler's ruin, we know that if the initial distance is k, then the expected time to absorption is $k(n-k) \leq n^2/4$.
- Hence,

$$\tau_{\text{mix}} \leq 4 \cdot \frac{n^2}{4} = n^2.$$