STATS 217: Introduction to Stochastic Processes I

Lecture 19

Convergence theorem

Today, we will prove 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. There exists some $\epsilon > 0$ (depending on P) such that

but does not depend on n.

$$\max_{x \in S} \mathrm{TV}(X_n \mid X_0 = x, \pi) \leq (1-\epsilon)^n.$$

choose the worst possible starting point.

$X_0 = x, X_1, X_2, \ldots, X_n$

Key quantities

- For any $x \in S$, let

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

- Let

$$\Delta(n) = \max_{x \in S} \Delta_x(n).$$

- Therefore, our goal is to show that there exists some $\epsilon > 0$ such that

$$\Delta(n) \le (1 - \epsilon)^n.$$

$$\max_{x \in S} \underbrace{TV(X_n | X_0 = x, \pi)}_{= \Delta_x(n)}$$

Key quantities

- For any $x \in S$, let
$$\Delta_x(n) = \mathrm{TV}(X_n \mid X_0 = x, \pi).$$
- Let
$$\Delta(n) = \max_{x \in S} \Delta_x(n).$$
- Therefore, our goal is to show that there exists some $\epsilon > 0$ such that
$$\Delta(n) \leq (1-\epsilon)^n.$$

we will show this along the subsequence $n = n_0 k$ for some $n_0 \geq 1$.

0 1 0 1 0 1 0 1 . . . .

- On the homework, you will show that $\Delta(n+1) \leq \Delta(n)$ for all integers $n \geq 0$.

Key quantities

- For any $x \in S$, let
$$\Delta_x(n) = \mathrm{TV}(X_n \mid X_0 = x, \pi).$$
- Let
$$\Delta(n) = \max_{x \in S} \Delta_x(n).$$
- Therefore, our goal is to show that there exists some $\epsilon > 0$ such that
$$\Delta(n) \leq (1-\epsilon)^n.$$
- On the homework, you will show that $\Delta(n+1) \leq \Delta(n)$ for all integers $n \geq 0$.
- Therefore, it suffices to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that
$$\Delta(r_0 n) \leq (1-\epsilon)^n.$$
$$\leq (1-\delta)^{r_0 n} \text{ for some } \delta > 0$$

Key quantities

$$\Delta_x(n) = TV(X_n \mid X_0 = x, \pi)$$

- It will be a bit more convenient to work with the following quantities: for any $x, y \in S$, let

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

- Let

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

start chain at x → $X_n \mid X_0 = x$

start chain at y → $X_n \mid X_0 = y$

TV b/w these.

Recall:

$$\Delta(n) = \max_{x \in S} \Delta_x(n)$$

Key quantities

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

- It will be a bit more convenient to work with the following quantities: for any $x, y \in S$, let

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

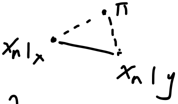
- Let

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

$$D_{x,y}(n) \le TV(X_n \mid x, \pi) + TV(\pi, X_n \mid y) \le 2\Delta(n)$$

$$\max_{x,y} D(n) \le 2\Delta(n)$$

- $\Delta(n) \le D(n)$ for all integers n. Why?



So it will suffice to show that

$$D(r_0 n) \le (1-\epsilon)^n \Rightarrow \Delta(r_0 n) \le (1-\epsilon)^n.$$

Key quantities

- It will be a bit more convenient to work with the following quantities: for any $x, y \in S$, let
$$D_{x,y}(n) = \mathrm{TV}(X_n \mid X_0 = x, X_n \mid X_0 = y).$$
- Let
$$D(n) = \max_{x,y \in S} D_{x,y}(n).$$

- $\Delta(n) \le D(n)$ for all integers n. Why?
- It suffices to show that for any $x \in S$ and any $A \subseteq \Omega$,
$$\mathbb{P}[X_n \in A \mid X_0 = x] - \pi(A) \le D(n).$$

if we can show this,
(1) then take max over $A \subseteq \Omega$,
(2) take max over $x \in S$

$$\mathrm{TV}(\mu, \nu) = \max_{A \subseteq \Omega} |\nu(A) - \mu(A)|$$

Key quantities

$P^n(x,\cdot)$ is the distribution of $X_n \mid X_0 = x$

We have

$$\mathbb{P}[X_n \in A \mid X_0 = x] - \pi(A) = P^n(x, A) - \sum_{y\in S} \pi(y) P^n(y, A)$$

$\pi = \pi P^n$

$\pi(A)$

$= \sum_{y \in S} \pi(y) P^n(y, A)$

$\pi(\cdot) = \sum_{y\in S} \pi(y) P^n(y, \cdot)$

$$= \sum_{y\in S} \pi(y) P^n(x, A) - \sum_{y\in S} \pi(y) P^n(y, A)$$

$\sum_{y\in S} \pi(y) = 1$

Key quantities

We have

$$
\begin{aligned}
\mathbb{P}[X_n \in A \mid X_0 = x] - \pi(A) &= P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) [P^n(x, A) - P^n(y, A)] \\
&\leq \max_{y \in S} |P^n(x, A) - P^n(y, A)|
\end{aligned}
$$

Key quantities

We have

$$\begin{aligned}
\mathbb{P}[X_n \in A \mid X_0 = x] - \pi(A) &= P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) [P^n(x, A) - P^n(y, A)] \\
&\leq \max_{y \in S} |P^n(x, A) - P^n(y, A)| \\
&\leq \max_{y \in S} D_{x,y}(n) \leq D(n).
\end{aligned}$$

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

-

Key quantities

We have

$$
\begin{aligned}
\mathbb{P}[X_n \in A \mid X_0 = x] - \pi(A) &= P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) P^n(x, A) - \sum_{y \in S} \pi(y) P^n(y, A) \\
&= \sum_{y \in S} \pi(y) [P^n(x, A) - P^n(y, A)] \\
&\leq \max_{y \in S} |P^n(x, A) - P^n(y, A)| \\
&\leq \max_{y \in S} D_{x,y}(n) \\
&\leq D(n).
\end{aligned}
$$

Overview

$\Delta_x(n) = TV(X_n \mid X_0 = x, \pi)$, $\Delta(n) = \max_x \Delta_x(n)$.

- We want to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$\Delta(r_0 n) \leq (1-\epsilon)^n.$$

this is enough b/c $\Delta(n+1) \leq \Delta(n)$.

- Since $\Delta(n) \leq D(n)$ for all integers $n \geq 0$, it suffices to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$D(r_0 n) \leq (1-\epsilon)^n.$$

$D(n) = \max_{x,y} TV(X_n \mid X_0 = x, X_n \mid X_0 = y)$

Overview

- We want to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$\Delta(r_0 n) \leq (1-\epsilon)^n.$$

- Since $\Delta(n) \leq D(n)$ for all integers $n \geq 0$, it suffices to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$D(r_0 n) \leq (1-\epsilon)^n.$$

- For this, we will first show that D is sub-multiplicative i.e., for any integers $s, t \geq 0$,

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

suppose $D(r_0) \leq 1-\varepsilon$

$$D(r_0 n) = D(r_0 + r_0 + \dots + r_0) \leq D(r_0)^n$$
$$\leq (1-\varepsilon)^n.$$

Overview

- We want to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that
$$\Delta(r_0 n) \leq (1-\epsilon)^n.$$
- Since $\Delta(n) \leq D(n)$ for all integers $n \geq 0$, it suffices to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that
$$D(r_0 n) \leq (1-\epsilon)^n.$$
- For this, we will first show that D is sub-multiplicative i.e., for any integers $s, t \geq 0$,
$$D(s+t) \leq D(s)D(t).$$
- This implies that for any integer $r \geq 1$, $D(nr) \leq D(r)^n$.

Overview

- We want to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$\Delta(r_0 n) \leq (1-\epsilon)^n.$$

- Since $\Delta(n) \leq D(n)$ for all integers $n \geq 0$, it suffices to show that there is some integer $r_0 \geq 1$ and some $\epsilon > 0$ such that

$$D(r_0 n) \leq (1-\epsilon)^n.$$

- For this, we will first show that D is sub-multiplicative i.e., for any integers $s, t \geq 0$,

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

- This implies that for any integer $r \geq 1$, $D(nr) \leq D(r)^n$. | until here, abstract argument.
- Finally, using the irreducibility and aperiodicity of P, we will show that there exists an integer $r_0 \geq 1$ such that $D(r_0) < 1$.

Sub-multiplicativity of D

$\{0\}$ $\{1\}$ $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$

- Let us prove the key sub-multiplicativity property

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

- The left hand side is

$$\max_{x,y\in S} \mathrm{TV}(X_{t+s} \mid X_0 = x, X_{t+s} \mid X_0 = y).$$

- For now, fix $x, y \in S$. Later, we will take the maximum.

Sub-multiplicativity of D

- Let us prove the key sub-multiplicativity property

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

- The left hand side is

$$\max_{x,y \in S} \mathrm{TV}(X_{t+s} \mid X_0 = x, X_{t+s} \mid X_0 = y).$$

- For now, fix $x, y \in S$. Later, we will take the maximum.
- We will bound the left hand side by constructing a coupling $(\widehat{X}_{t+s}, \widehat{Y}_{t+s})$ of $X_{t+s} \mid X_0 = x$ and $X_{t+s} \mid X_0 = y$.

Recall: coupling lemma

$$\mathrm{TV}(\mu, \nu) = \min_{(X,Y) \text{ coupling of } \mu \,\&\, \nu} \mathbb{P}[X \neq Y]$$

Constructing a coupling

we are trying to construct $\hat{X}_{t+s}, \hat{Y}_{t+s}$

$\mathbb{P}[\hat{X}_{t+s} \neq \hat{Y}_{t+s}] \leq D(s)D(t).$

Here is our coupling:

- First, use the coupling lemma to find a coupling $(\widehat{X}_t, \widehat{Y}_t)$ of the distributions $X_t \mid X_0 = x$ and $X_t \mid X_0 = y$ such that

$$\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] = \mathrm{TV}(X_t \mid X_0 = x, X_t \mid X_0 = y) = D_{x,y}(t).$$

imagine

$x \rightsquigarrow X_t \mid X_0 = x$ $\quad$ $y \rightsquigarrow X_t \mid X_0 = y$

$$\begin{aligned} TV(X_t \mid X_0 = x, X_t \mid X_0 = y) &= D_{x,y}(t) \\ &\leq D(t) \end{aligned}$$

Constructing a coupling

Here is our coupling:

- First, use the coupling lemma to find a coupling $(\widehat{X}_t, \widehat{Y}_t)$ of the distributions $X_t \mid X_0 = x$ and $X_t \mid X_0 = y$ such that

$$\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] = \mathrm{TV}(X_t \mid X_0 = x, X_t \mid X_0 = y) = D_{x,y}(t).$$

- If $\widehat{X}_t = \widehat{Y}_t$, then set $\widehat{X}_{t+s} = \widehat{Y}_{t+s}$. $\sim$ $\hat{X}_t P^s = \hat{Y}_t P^s$

Constructing a coupling

Here is our coupling:

- First, use the coupling lemma to find a coupling $(\widehat{X}_t, \widehat{Y}_t)$ of the distributions $X_t \mid X_0 = x$ and $X_t \mid X_0 = y$ such that

$$\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] = \mathrm{TV}(X_t \mid X_0 = x, X_t \mid X_0 = y) = D_{x,y}(t).$$

- If $\widehat{X}_t = \widehat{Y}_t$, then set $\widehat{X}_{t+s} = \widehat{Y}_{t+s}$.
- Else, if $x' = \widehat{X}_t \neq \widehat{Y}_t = y'$, use the coupling lemma to find a coupling $(\widehat{U}_s, \widehat{W}_s)$ of the distributions $X_{t+s} \mid X_t = x'$ and $X_{t+s} \mid X_t = y'$ such that

$$\mathbb{P}[\widehat{U}_s \neq \widehat{W}_s] = \mathrm{TV}(X_{t+s} \mid X_t = x', X_{t+s} \mid X_t = y')$$

$x' \quad y'$

$X_s \mid X_0 = x' \quad X_s \mid X_0 = y'$

$(\hat{U}_s, \hat{W}_s)$

$X_s \mid X_0 = x' \sim X_{t+s} \mid X_t = x'$

$X_s \mid X_0 = y' \sim X_{t+s} \mid X_t = y'$

Constructing a coupling

Here is our coupling:

- First, use the coupling lemma to find a coupling $(\widehat{X}_t, \widehat{Y}_t)$ of the distributions $X_t \mid X_0 = x$ and $X_t \mid X_0 = y$ such that

$$\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] = \mathrm{TV}(X_t \mid X_0 = x, X_t \mid X_0 = y) = D_{x,y}(t).$$

- If $\widehat{X}_t = \widehat{Y}_t$, then set $\widehat{X}_{t+s} = \widehat{Y}_{t+s}$.
- Else, if $x' = \widehat{X}_t \neq \widehat{Y}_t = y'$, use the coupling lemma to find a coupling $(\widehat{U}_s, \widehat{W}_s)$ of the distributions $X_{t+s} \mid X_t = x'$ and $X_{t+s} \mid X_t = y'$ such that

$$\mathbb{P}[\widehat{U}_s \neq \widehat{W}_s] = \mathrm{TV}(X_{t+s} \mid X_t = x', X_{t+s} \mid X_t = y') = D_{x',y'}(s) \leq D(s).$$

  Here, the second equality uses the Markov property.

Constructing a coupling

Here is our coupling:

- First, use the coupling lemma to find a coupling $(\widehat{X}_t, \widehat{Y}_t)$ of the distributions $X_t \mid X_0 = x$ and $X_t \mid X_0 = y$ such that

$$\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] = \mathrm{TV}(X_t \mid X_0 = x, X_t \mid X_0 = y) = D_{x,y}(t).$$

- If $\widehat{X}_t = \widehat{Y}_t$, then set $\widehat{X}_{t+s} = \widehat{Y}_{t+s}$.
- Else, if $x' = \widehat{X}_t \neq \widehat{Y}_t = y'$, use the coupling lemma to find a coupling $(\widehat{U}_s, \widehat{W}_s)$ of the distributions $X_{t+s} \mid X_t = x'$ and $X_{t+s} \mid X_t = y'$ such that

$$\mathbb{P}[\widehat{U}_s \neq \widehat{W}_s] = \mathrm{TV}(X_{t+s} \mid X_t = x', X_{t+s} \mid X_t = y') = D_{x',y'}(s) \leq D(s).$$

$\mathrm{TV}(X_s \mid X_0 = x', X_s \mid X_0 = y')$

Here, the second equality uses the Markov property. Then, set
+ time homogeneity

$$(\widehat{X}_{t+s}, \widehat{Y}_{t+s}) = (\widehat{U}_s, \widehat{W}_s).$$

Analysis of the coupling

- By construction, it is clear that $\widehat{X}_{t+s} \sim X_{t+s} \mid X_0 = x$ and $\widehat{Y}_{t+s} \sim X_{t+s} \mid X_0 = y$.

Analysis of the coupling

- By construction, it is clear that $\widehat{X}_{t+s} \sim X_{t+s} \mid X_0 = x$ and $\widehat{Y}_{t+s} \sim X_{t+s} \mid X_0 = y$.
- Therefore, by the coupling lemma,

$$\begin{aligned} D_{x,y}(t+s) &\le \mathbb{P}[\widehat{X}_{t+s} \neq \widehat{Y}_{t+s}] \\ &= \mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t, \widehat{U}_s \neq \widehat{W_s}] \end{aligned}$$

Analysis of the coupling

- By construction, it is clear that $\widehat{X}_{t+s} \sim X_{t+s} \mid X_0 = x$ and $\widehat{Y}_{t+s} \sim X_{t+s} \mid X_0 = y$.
- Therefore, by the coupling lemma,

$$
\begin{aligned}
D_{x,y}(t+s) &\le \mathbb{P}[\widehat{X}_{t+s} \neq \widehat{Y}_{t+s}] \\
&= \mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t, \widehat{U}_s \neq \widehat{W}_s] \\
&= \underbrace{\mathbb{P}[\widehat{U}_s \neq \widehat{W}_s \mid \widehat{X}_t \neq \widehat{Y}_t]}\,\underbrace{\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t]} \\
&\le D(s) \cdot D(t)
\end{aligned}
$$

and now take max over x, y

Analysis of the coupling

- By construction, it is clear that $\widehat{X}_{t+s} \sim X_{t+s} \mid X_0 = x$ and $\widehat{Y}_{t+s} \sim X_{t+s} \mid X_0 = y$.
- Therefore, by the coupling lemma,

$$\begin{aligned} D_{x,y}(t+s) &\leq \mathbb{P}[\widehat{X}_{t+s} \neq \widehat{Y}_{t+s}] \\ &= \mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t, \widehat{U}_s \neq \widehat{W}_s] \\ &= \mathbb{P}[\widehat{U}_s \neq \widehat{W}_s \mid \widehat{X}_t \neq \widehat{Y}_t]\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] \\ &\leq D(s) D_{x,y}(t). \end{aligned}$$

Analysis of the coupling

- By construction, it is clear that $\widehat{X}_{t+s} \sim X_{t+s} \mid X_0 = x$ and $\widehat{Y}_{t+s} \sim X_{t+s} \mid X_0 = y$.
- Therefore, by the coupling lemma,

$$\begin{aligned} D_{x,y}(t+s) &\le \mathbb{P}[\widehat{X}_{t+s} \neq \widehat{Y}_{t+s}] \\ &= \mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t, \widehat{U}_s \neq \widehat{W}_s] \\ &= \mathbb{P}[\widehat{U}_s \neq \widehat{W}_s \mid \widehat{X}_t \neq \widehat{Y}_t]\mathbb{P}[\widehat{X}_t \neq \widehat{Y}_t] \\ &\le D(s)D_{x,y}(t). \end{aligned}$$

- Taking the maximum over all $x, y \in S$, we get that

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

Bounding $D(r)$

- We claim that $D(r_0) < 1$.
- On the homework, you will show that if P is irreducible and aperiodic, then there exists some r_0 such that $P^{r_0}_{x,y} > 0$ for all $x, y \in S$.

Bounding $D(r)$

- We claim that $D(r_0) < 1$.
- On the homework, you will show that if P is irreducible and aperiodic, then there exists some r_0 such that $P^{r_0}_{x,y} > 0$ for all $x, y \in S$.
- In particular, for any $x \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have $P^{r_0}(x, A) > 0$.

Bounding $D(r)$

- We claim that $D(r_0) < 1$.
- On the homework, you will show that if P is irreducible and aperiodic, then there exists some r_0 such that $P^{r_0}_{x,y} > 0$ for all $x, y \in S$.
- In particular, for any $x \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have $\boxed{P^{r_0}(x, A) > 0.}$
- For such an r_0, for any $x, y \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have

$$\max_{A \subseteq \Omega} \; |P^{r_0}(x, A) - P^{r_0}(y, A)| \leq |1 - \min\{\underbrace{P^{r_0}(x, A), P^{r_0}(y, A)}_{>0}\}| < 1.$$

$$\Vert$$

$$TV(X_{r_0} \mid X_0 = x, X_{r_0} \mid X_0 = y)$$

$$\Vert$$

$$D_{xy}(r_0)$$

Bounding $D(r)$

- We claim that $D(r_0) < 1$.
- On the homework, you will show that if P is irreducible and aperiodic, then there exists some r_0 such that $P^{r_0}_{x,y} > 0$ for all $x, y \in S$.
- In particular, for any $x \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have $P^{r_0}(x, A) > 0$.
- For such an r_0, for any $x, y \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have

$$|P^{r_0}(x, A) - P^{r_0}(y, A)| \leq |1 - \min\{P^{r_0}(x, A), P^{r_0}(y, A)\}| < 1.$$

- Taking the maximum over all $A \neq \emptyset$ shows that

$$D_{x,y}(r_0) = \mathrm{TV}(X_{r_0} \mid X_0 = x, X_{r_0} \mid X_0 = y) < 1.$$

Bounding $D(r)$

- We claim that $D(r_0) < 1$.
- On the homework, you will show that if P is irreducible and aperiodic, then there exists some r_0 such that $P^{r_0}_{x,y} > 0$ for all $x, y \in S$.
- In particular, for any $x \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have $P^{r_0}(x, A) > 0$.
- For such an r_0, for any $x, y \in S$ and for any $A \subseteq S$, $A \neq \emptyset$, we have

$$|P^{r_0}(x, A) - P^{r_0}(y, A)| \leq |1 - \min\{P^{r_0}(x, A), P^{r_0}(y, A)\}| < 1.$$

- Taking the maximum over all $A \neq \emptyset$ shows that

$$D_{x,y}(r_0) = \mathrm{TV}(X_{r_0} \mid X_0 = x, X_{r_0} \mid X_0 = y) < 1.$$

- Finally, taking the maximum over all $x, y \in S$, we have $D(r_0) < 1$.