1

Lesson 18

Markov Chains

The Markovian Property

Recall if a sequence of events $\{A_k\}$ are independent

$$\mathbb{P}\left(\bigcap_{k=1}^{n} A_k\right) = \prod_{k=1}^{n} \mathbb{P}(A_i)$$

In general, the multiplication rule applies to $\{A_i\}$:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\,\mathbb{P}(A_2|A_1)\,\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}\left(A_n \Big| \bigcap_{j=1}^{n-1} A_j\right)$$

There is an interesting case which falls in between:

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2)\ldots\mathbb{P}(A_n|A_{n-1})$$

Events that satisfy the above property are said to have the Markovian Property.

If the indices are interpreted as "time," the Markovian property implies that the future ~~probability~~ (A_{i+1}) depends only on the present (A_i), or the present (A_i) only depends on the immediate past (A_{i-1}). The Markovian property is also called the "memoryless property" as well

This idea ~~property~~ Can be applied

to random variables

Def: (Markov Chain)

Assume that $\{X_i\}_{i\in I}$ is a sequence of r.v.'s on $(\Omega, \mathcal{F}, \mathbb{P})$, where I is an index set $(I \subseteq \mathbb{N})$

$\{X_i\}_{i\in I}$ is Called a

4

Markov Chain iff

$$\mathbb{P}\left(\{X_{n+1} = x_{n+1}\} \,\middle|\, \bigcap_{i=1}^{n} \{X_i = x_i\}\right)$$

$$= \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

In other words:

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \ldots, X_1 = x_1)$$

$$= \mathbb{P}(X_{n+1} = x_{n+1} \mid X_n = x_n)$$

Assume that R_i is the range of X_i. $S = \bigcup_{i \in I} R_i$ is the set of all possible values X_i's can take and is called the State Space of the Markov Chain

Example: Show that S_n is a Markov Chain, if $\{X_i\}_{i \in \mathbb{N}}$ is a sequence of i.i.d. r.v.'s and

$$S_n = \sum_{i=1}^{n} X_i, \quad \forall n \in \mathbb{N}.$$

S_n is called a random walk, especially when $X_i \sim Ber(p)$.

6

7

Def: Transition Probability

The probability that the chain moves from state i to state j at time n is called the transition probability p_{ij} from state i to state j.

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad \forall i, j \in S$$

One can assume, without loss of generality, that $S = \{1, 2, \ldots, m\}$.

Therefore, transition probabilities can be shown in the following transition matrix

$$P = [p_{ij}]$$

Note that p_{ij} can be dependent on n in general. If p_{ij} does not depend on n, the Markov Chain (MC) is called homogeneous.

Clearly, $\sum_j p_{ij} = \sum_j \mathbb{P}(X_{n+1}=j \mid X_n = i) = 1$

i.e. the sum of each row is 1.

Example: A rabbit moves along a straight line, taking unit steps. At each time period, it moves one unit to the right with probability 0.4, one

unit to the left with probability 0.4, and remains in the same place with probability 0.2, independent of its previous moves. Two rabbit holes are at

10

positions 1 and m. If the rabbit reaches each of the rabbit hole, its search ends. Construct a Markov Chain model for this process,

assuming that the rabbit starts in a position between 1 and m. ~~[illegible]~~

11

12

Transition Graph

The Transition Graph of a MC is a weighted directional graph associated with the transition matrix

Example: Draw the transition graph for the previous example.

13

n-Step Transition Probabilities

We usually require the probability ~~~~ of some state at some future step be calculated conditioned

upon the current state.

This gives rise to the definition of m-step transition probabilities:

$$P_{ij}(m) = \mathbb{P}(X_m = j \mid X_0 = i)$$

Obviously: $P_{ij}(1) = P_{ij}$

and $p_{ij}(0) = \mathbb{P}(X_0 = j \mid X_0 = i) = \delta_{ij}$

where $\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

is the Kronecker Delta.

The Chapman-Kolmogorov Equation

$$p_{ij}(n+m) = \sum_k p_{ik}(n)\, p_{kj}(m)$$

Proof:
To see a proof, see Gubner p. 484.

In general, the matrix with entries $p_{ij}(n)$ is given by P^n for a homogeneous Markov Chain, according to Chapman-Kolmogorov. Also

$$P(n+m) = P(n)P(m) = P^n P^m$$

16

Steady State Behavior of Markov Chains

The transient behavior of a MC is completely determined by $\mathbb{P}$. However, if we need

the unconditional probabilities at step n, ~~e.g.~~ i.e $\mathbb{P}(X_n = i)$ for $i \in \{1, 2, \dots, m\} = S$, we need an initial pdf (row) vector

$$\underline{\pi}_0 = \left[\mathbb{P}(X_0 = 1) \quad \dots \quad \mathbb{P}(X_0 = m)\right]$$

17

The pdf vector for X_n, $\underline{\pi}_n$ is

$\underline{\pi}_n = \underline{\pi}_0 P(n)$

or when the MC is homogeneous,

$\underline{\pi}_n = \underline{\pi}_0 P^n$

We are interested in the behavior of $\underline{\pi}_n$, the pdf vector of X_n, as $n \longrightarrow \infty$.

A "steady state" distribution vector $\underline{\pi}_\infty$ needs to satisfy the following property

$$\underline{\pi}_\infty = \underline{\pi}_\infty P$$

This is obviously an eigenvector problem:

$$P^T \pi_\infty^T = \pi_\infty^T$$

π_∞^T is the eigenvector associated with the matrix P^T, corresponding to the eigenvalue $\lambda = 1$, that satisfies $\sum \pi_i = 1$.

Every Markov chain has at least one $\underline{\pi}_\infty$ satisfying

$$\underline{\pi}_\infty P = \underline{\pi}_\infty$$

(This is due to the Brouwer's fixed-point Theorem)

However, there is no guarantee that

$$\lim_{n \to \infty} \pi_n = \underline{\pi}_\infty$$

Categories of States in a MC

Def:
State j is reachable from State i iff there exists some n such that $P_{ij}(n) \neq 0$.

Def:
States i and j communicate iff i is reachable from j and j is reachable from i.

21

It can be shown that the relation $R_C : S \to S$

$R_C = \{(i,j) \mid i \text{ communicates with } j\}$

is an <u>equivalence</u> relation on the State Space S

i.e. it is reflexive, symmetric, and transitive

$i R_c i \quad \forall i \in S$

$i R_c j \iff j R_c i \quad \forall i,j \in S$

$i R_c j,\ j R_c k \Rightarrow i R_c k \quad \forall i,j,k \in S$

An equivalence relation on S partitions it into equivalence classes. Therefore, if

$C_i = \{j \mid j \text{ communicates with } i\}$

then

$\bigcup_i C_i = S$ &

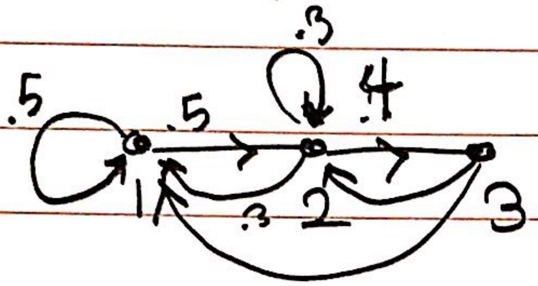
$C_i \cap C_j = \emptyset$ when $C_i \neq C_j$

Note that some C_i's may be the same, e.g. all the states that communicate with

23

$i=1$, communicate with $j=2$

as well, so $C_1 = C_2$



Therefore we have the

condition "when $C_i \neq C_j$"

i.e. "distinct" equivalence

classes are ~~[illegible]~~ non-overlaping

Def: (The period of state i)

$T(i) = \gcd\{n \mid P_{ii}(n) \neq 0\}$

where gcd is the greatest common divisor and

$P_{ii}(n) = \mathbb{P}(X_n = i \mid X_0 = i)$

$T(i)$ is the period of all possible numbers of steps the chain needs to start from i and return to it.

- If $T(i) = 1$, i is called an aperiodic state
- If $T(i) > 1$, the state is called periodic

Def (Ergodic MC)

A Markov chain is called Ergodic if all of its states are aperiodic and all of its states communicate

26

Theorem: If a Markov Chain is Ergodic, $\underline{\pi}_\infty$ is unique and $\lim_{n\to\infty} \underline{\pi}_n = \underline{\pi}_\infty$

Note: $\lim_{n \to \infty} \underline{\pi}_n = \underline{\pi}_\infty$ means that the Markov Chain (X_n) Converges in distribution, i.e.

$$X_n \xrightarrow{d} X_\infty$$

where the pmf of X_∞ is $\underline{\pi}_\infty$.

Speed of Convergence

Recall that $\underline{r}_\infty^T$ is the eigenvector of P^T that is associated with $\lambda = 1$.

Theorem (Perron - Frobenius):

Assume $A \in \mathbb{R}^{n \times n}$, $\forall i,j \; a_{ij} > 0$

If $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$ is the set of all eigenvalues of A, then there exists a positive real number $r > 0$ called the

Perron root such that

$|\lambda_i| \leq r$, where r is a simple eigenvalue of A, associated with an eigenvector with positive elements.

For a Markov chain that is Ergodic

$\forall i, \ |\lambda_i| \leq |\lambda_{max}| = 1$

Denote λ_2 as the eigenvalue

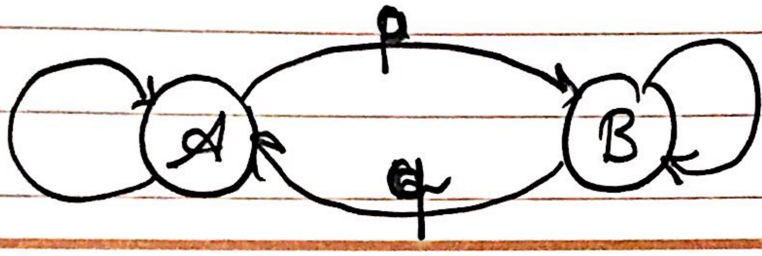
with the second largest modulus then

Markov Chain's rate of convergence is proportional to $1/|\lambda_2|$

So larger $|\lambda_2|$ results in slower

Convergence rates.

Example: Find the steady state distribution for the following Markov Chain



$p, q \in (0,1)$

$$\underline{P} = \begin{bmatrix} & \\ & \end{bmatrix}$$

Transience vs. Recurrence

- We can define a transient state as a state ~~that~~ for which the probability of never returning to that state is non-zero.

- We can define a recurrent state as a state that is revisited infinitely often (i.e., the probability of never returning to it is zero)

Def: In a Markov chain,

state i is called transient iff

$$\sum_{n=1}^{\infty} p_{ii}^{n} < \infty$$

and i is called recurrent

iff $$\sum_{n=1}^{\infty} p_{ii}^{n} = \infty$$

We can use another concept, <u>first passage time</u>, to define transition & recurrence.

First passage time of state i is the first time (k) that

Def: In a Markov chain, state i is called transient iff

$$\sum_{n=1}^{\infty} p_{ii}^{n} < \infty$$

and i is called recurrent iff

$$\sum_{n=1}^{\infty} p_{ii}^{n} = \infty$$

We can use another concept, first passage time, to define transition & recurrence.

First passage time of state i is the first time (k) that the Markov Chain visits state i. We denote it as $T_1(i)$

Def) The first passage time of State i is:

$$T_1(i) = \min \{ k \geq 1 \mid X_k = i \}$$

Def) Probability of Return to i in finite time f_i is

$$f_i = \mathbb{P}(T_1(i) < \infty \mid X_0 = i)$$

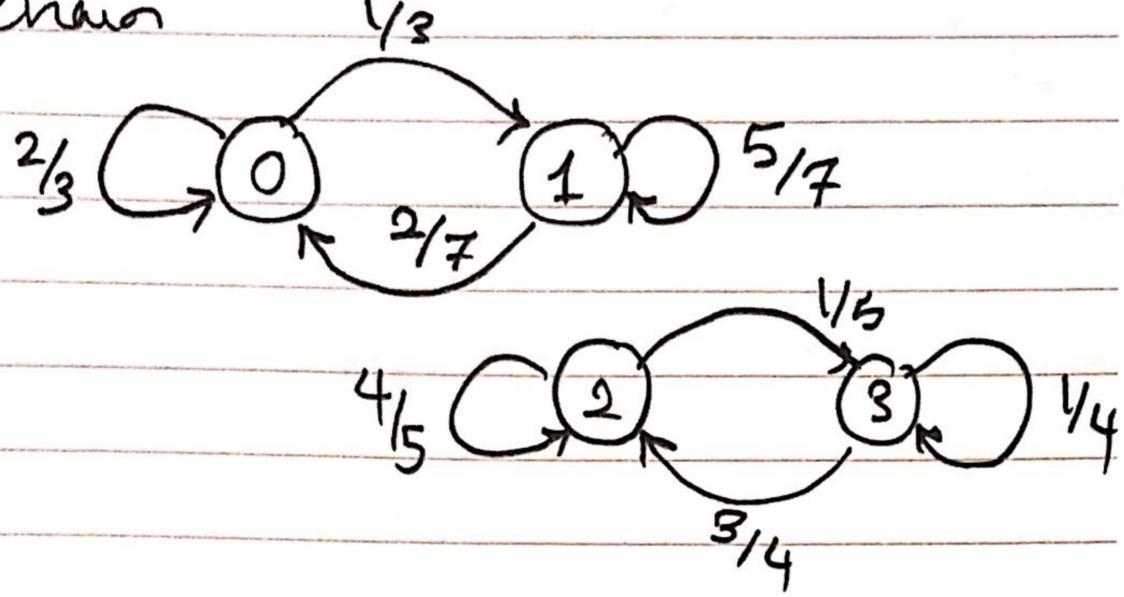
Then one can see that

$f_i < 1 \iff i$ is a transient state

$f_i = 1 \iff i$ is a recurrent state

Theorem: ~~Iff~~ States i and j communicate in a Markov Chain iff both i, j are recurrent or both i, j are transient.

Example: Consider the following chain



$P =$

Check that for both

$\underline{\pi}_{\infty_1} = [6/13 \quad 7/13 \quad 0 \quad 0]$ and

$\underline{\pi}_{\infty_2} = [0 \quad 0 \quad 15/19 \quad 4/19]$

$\underline{\pi}_{\infty_1} P = \underline{\pi}_{\infty_1}$ and $\underline{\pi}_{\infty_2} = \underline{\pi}_{\infty_2}$

i.e. the Markov Chain has more than one stationary distributions

State 0 is recurrent, because:

State 1 is ~~recurrent~~ transient because:

Limit Theorems for Markov Chains

Theorem: Assume that $\{X_n\}_{n\geq 0}$ is an ergodic Markov chain with stationary pdf $\underline{\pi}_\infty$ and $f: R \to R$ is a function such that

$$\mathbb{E}_{\underline{\pi}_\infty}[|f|] = \sum_i \pi_\infty(i) f(i) < \infty$$

Then: Law of Large Numbers for Markov Chains

$$\overline{f}_n(X) = \frac{1}{n} \sum_{k=1}^{n} f(X_k) \xrightarrow{a.s.} \mathbb{E}_{\pi_\infty}[f(X)]$$

and

$$\sqrt{n}\,(\overline{f}_n(X) - \mathbb{E}_{\pi_\infty}[f(X)] \xrightarrow{d} \mathcal{N}(0,\sigma^2)$$

where $\sigma^2 = \mathrm{Var}_{\pi_\infty}[f(X_1)]$

$$+ \mathrm{Cor}_{\pi_\infty}[f(X_1), f(X_k)]$$

Towards Continuous Time and Continuous State Space

So far we considered $|S| < \infty$

Extensions to $|S| = \mathbb{N}$ do not seem to be hard. But what

if S is not countable?

In this case, pmfs become pdfs

$|S| < \infty$

$P = [p_{ij}]$

$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i)$

or briefly

$p_{i \to j}$

$|S| = |\mathbb{R}|$

$\in \mathbb{R}$ $\quad \in \mathcal{F}$

$\underbrace{P(x, A)}_{\text{Kernel}} = \mathbb{P}(X_{n+1} \in A \mid X_n = x)$

or briefly

$p_{x \to A}$

$\sum_j p_{ij} = 1$

$\forall j \in S$

$\int \underbrace{p(x,y)}_{\text{Kernel}} dy$

$= \int p(y \mid x) dy = 1$

Kolmogorov – Chapman:

$P(n+m) = P(n)\,P(m)$

So $P(n+1) = P(n)\,P(1)$

$$P^{n+1}(x,A) = \int p^{n}(x,y)\,P(y,A)\,dy$$

Steady State equation

$$\underline{\pi}_{\infty} = \underline{\pi}_{\infty} P$$

$$f(y) = \int p(y,x)\,f(x)\,dx = \int p(y|x)\,f(x)\,dx$$

Continuous Time Markov Chains

(For simplicity, Assume that $|S| = m$)

Def) A family of integer-valued random variables $\{X_t, t \geq 0\}$ is called a Markov Chain iff

$$\mathbb{P}(X_t = j \mid X_s = i, X_{s_{n-1}} = i_{n-1}, \ldots, X_{s_0} = i_0)$$
$$= \mathbb{P}(X_t = j \mid X_s = i)$$

This means that given the sequence of values $i_0, \ldots, i_{n-1}, i$, the conditional probability of what X_t will be depends only on the condition $X_s = i$.

The quantity $\mathbb{P}(X_t = j \mid X_s = i)$ is called the transition probability.

Time Homogeneity:

If the transition probability $\mathbb{P}(X_t=j \mid X_s=i)$ depends only on $t-s$, we call the chain homogeneous. This means if

$$p_{ij}(t) = \mathbb{P}(X_t=j \mid X_0=i)$$

Then

$$\mathbb{P}(X_t=j \mid X_s=i)$$
$$= \mathbb{P}(X_{t-s}=j \mid X_0=i)$$
$$= p_{ij}(t-s)$$

Note that then

$$p_{ij}(0) = \delta_{ij}$$

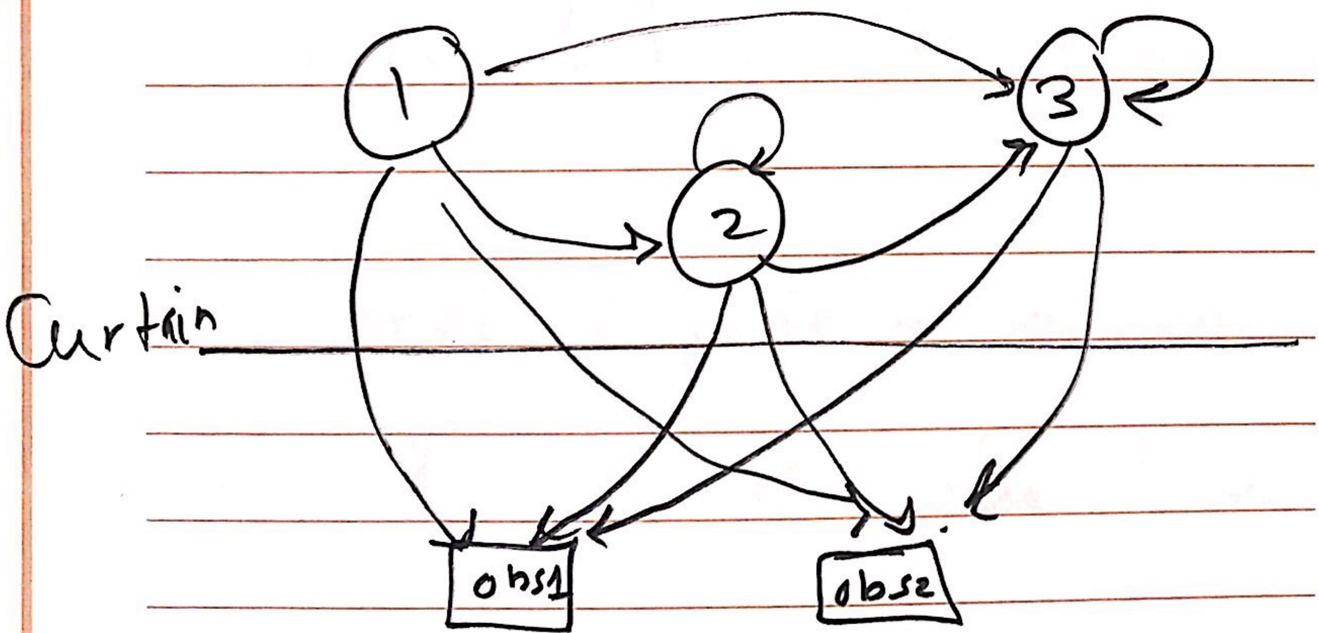
Continuous Time Kolmogorov Chapman:

Theorem: For a homogeneous continuous-time Markov Chain

$$p_{ij}(t+s) = \sum_k p_{ik}(t)\, p_{kj}(s)$$

An Application: Hidden Markov Models

Used for Modeling non-stationary sequential data (random samples whose distributions change)



Each of the edges

show the probability that observation $O_j \in \{O_1, \dots, O_M\}$ came from state i

$\cancel{B_{ij} = \mathbb{P}(O = O_j)}$

$b_j(k) = \mathbb{P}$(Observation k at time t | state j at t)

- Viterbi Algorithm is used to estimate the hidden states given a sequence of observations

– The Baum–Welch Algorithm is used to fit a model $(P, b, \underline{\pi}_0)$ to a set of observations.