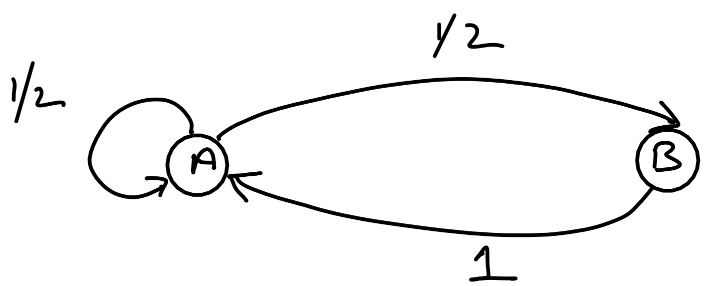
Markov chains: reading course

- Timing (Clash with IT class)
  Fri 2-3:30 pm
- Structure: Intial introductory lectures. [Mid Feb.]
  - Followed by topics (recent papers) (not so recent papers)

What is a Markov chain?



System:
At any given time instance, it is either in state A or state B.

$t \backslash t+1$	A	B
A	1/2	1/2
B	1	0

Transition matrix

- Transition probabilities only depend on current state, and not upon any further past history.

Markov Chain:-

- A collection of states: Ω.
- A $|\Omega| \times |\Omega|$ transition matrix P.
  ↓ 'Stochastic' matrix
  - Each entry is nn.
  - Each row sums to 1.

Interpretation:- For $u, v \in \Omega$.

$P(u,v) :=$ Probability of moving to v at time $t+1$ given that the current state is u.

Why is this a 'chain'?

Let X_0 be a random variable taking values in Ω.

Given X_0, generate a random variable X_1 s.t.

$$\Pr(X_1 = a \mid X_0) = P(X_0, a)$$

$\forall a \in \Omega$.

In general given X_i generate X_{i+1} s.t.

$$P(X_{i+1} = a \mid X_i) = P(X_i, a)$$

$\forall a \in \Omega$
$i \geq 0$.

'Chain':

$$\underbrace{X_0 \quad X_1, \quad X_2, \quad X_3, \quad \ldots \quad X_t, \ldots \ldots}$$

The chain satisfies the 'Markov property'.

$\forall i \geq 0$:

$X_{i+1}, X_{i+2}, \ldots$ is independent of $X_1, \ldots, X_{i-1}$ conditioned on X_i.

Q:- Suppose $X_0, X_1, X_2, \ldots X_t, \ldots$ is a sequence of r.v.'s taking values in Ω and satisfying the Markov property.

Is there necessarily a transition matrix P underlying this sequence?

Not necessarily: the transition matrix may depend upon the index i of X_i "the current time".

- Our definition above is making an extra 'time invariance' assumption in addition to the Markov property.

Consider some initial distribution μ_0 on Ω.

Ω is a finite set.

What is the distribution μ_1 after one step (staring with $X_0 \sim \mu_0$) of the Markov chain with transition matrix P?

$$\mu_1 = \cancel{P\mu_0} \; ?$$

$$\cancel{(P\mu_0)_a = \sum_{b \in \Omega} P_{ab}\, \mu_0(b)}$$

$$\mu_1(a) = \sum_{b \in \Omega} \mu_0(b)\, P_{ba}$$

$$\mu_1 = \mu_0 P \qquad \text{where } \mu_i \in \mathbb{R}^{|\Omega|} \text{ is a}$$

row-vector.

In general the distribution μ_i of X_i satisfies

$$\mu_i = \mu_{i-1} P. \qquad \forall i \geq 1.$$

$$\mu_i(b) = \sum_a \underbrace{\mu_{i-1}(a)}_{\Pr(X_{i-1}=a)} \; \underbrace{P(a,b)}_{\Pr(X_i = b \mid X_{i-1} = a)}$$

by definition of X_i

So,

$$\boxed{\mu_t = \mu_0 P^t \quad \forall \text{ non-negative integer } t.}$$

Q:- Is there some kind of convergence of this?

(Intuition from the power method might lead us to to expect convergence at least in some cases.).

Suppose there is 'convergence':

$$\tau = \lim_{t \uparrow \infty} \mu_0 P^t \quad \text{exists.}$$ Then,

(1) Easy to see then that $\tau \in \Delta_\Omega$

($\tau \geq 0$ & $\tau^T \vec{1} = 1$).

(2) $\boxed{\tau = \tau P}$

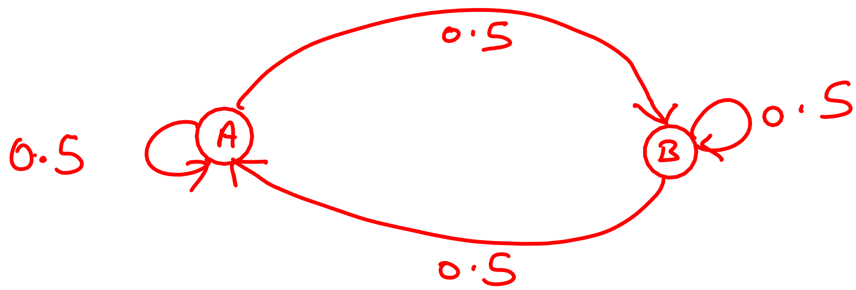
'Stationary distribution'

Example :- $P = \begin{pmatrix} 1/2 & 1/2 \\ 1 & 0 \end{pmatrix}$

$$\pi = \left(2/3, \ \frac{1}{3}\right).$$

$$\pi P = \left(\frac{2}{3}, \ \frac{1}{3}\right) = \pi.$$

Ex 2 :-



$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$

$$\pi = \left(\frac{1}{2}, \ \frac{1}{2}\right)$$

Claim: Take any $\mu \in \Delta_{\{A,B\}}$.

Then $\mu P = \pi$.

The chain converges to a stationary distribution in a single step, (for every starting state)

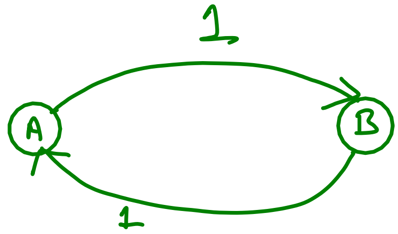
Example :-

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

Every μ a stationary distribution of P.

In the previous two example the chain actually reaches a stationary distribution in finitely many steps.

Example :-



$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$

$\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$ is still a stationary distribution.

But suppose

$\mu_0 = (1, 0)$ $\mu_2 = (1, 0)$ .... not going to converge

$\mu_1 = (0, 1)$ $\mu_3 = (0, 1)$

In this case

$\lim \mu_0 P^t$ does not exist.

Observations :-

(1) No examples so far where a stationary distribution does not exist.

(2) There are examples of μ_0, P s.t. $\lim_{t \to \infty} \mu_0 P^t$ does not exist.

Lemma :- Let P be an $n \times n$ stochastic matrix. Then there exist a $\pi \in \Delta_n$ $\left[\pi \in \mathbb{R}^n, \pi_i \geq 0 \; \forall i \in [n] \text{ and } \sum_{i=1}^{n} \pi_i = 1 \right]$

s.t. $\pi P = \pi$.

Pf :- $P\vec{1} = \vec{1}$ (by stochasticity)

$\Rightarrow$ 1 is an eigenvalue of P.

So there must also exist $u \in \mathbb{R}^n$ (row vector) $uP = u$, $u \neq \vec{0}$. But u possibly has negative entries!

Consider the vector $w \in \mathbb{R}^n$.

$w_i := |u_i| \quad 1 \leq i \leq n.$

<u>Claim:-</u> w is an eigenvector of P with eigenvalue 1.

$$(wP)_j = \sum_{i=1}^{n} w_i P_{ij} = \sum_{i=1}^{n} |u_i| P_{ij}$$

$$= \sum_{i=1}^{n} |u_i P_{ij}| \quad (\because P_{ij} \geq 0).$$

$$\geq \left| \sum_{i=1}^{n} u_i P_{ij} \right|$$

$$= |u_j| = w_j.$$

So, $\forall j \in [n]$ we have

$$(wP)_j \geq w_j$$

<u>Sub-claim:-</u> None of these inequalities is strict.

Suppose not. Then

$$\sum_{j=1}^{n} (wP)_j > \sum_{j=1}^{n} w_j$$

$$\Leftrightarrow \sum_{j=1}^{n} \sum_{i=1}^{n} w_i P_{ij} > \sum_{j=1}^{n} w_j$$

$$\Leftrightarrow \sum_{i=1}^{n} w_i \sum_{j=1}^{n} P_{ij} > \sum_{j=1}^{n} w_j$$

But $\sum_{j=1}^{n} P_{ij} = 1$.

So, we get the contradiction

$$\sum_{i=1}^{n} \omega_i > \sum_{j=1}^{n} \omega_j . \Rightarrow\!|\!\Leftarrow$$

So, we have $\omega \neq 0$ $\omega_i \geq 0$ s.t.

$$\omega P = \omega .$$

Then

$$\pi = \frac{\omega}{\sum_{i=1}^{n} \omega_i}$$ is as in the theorem.

Plausible
<u>Conjecture</u> :- Matrix does not look like.

$$\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \quad ?$$

$\boxed{\text{'Connectivity' + a little bit will imply uniqueness \& convergence.}}$

Will be useful to have a notion of distance on probability distributions.

$$d_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{\|\vec{p} - \vec{q}\|_1}{2} .$$

$p, q \in \Delta_n$

(H.W.) $\max_{S \subseteq [n]} \left(p(S) - q(S) \right)$ $\left[\begin{array}{l} p(S) := \sum_{i \in S} p_i \\ q(S) := \sum_{i \in S} q_i \end{array} \right]$

Lemma :- Suppose P is $n\times n$ stochastic and every entry of P is strictly positive. Then,

P is a strict contraction in d_{TV}.

That is :- $\exists \alpha$, $0 < \alpha < 1$ s.t. for all $\mu \neq \nu$ which are probability distributions on $[n]$, we have

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

Possible slots :- 2:30 – 4:00 Tuesday ✓
5:30 – 7:00 Thursdays ??