SEQUENCE MODELS - AI-3102, IIT-H

SPRING 2023

An Introduction to Markov chain

- A. A. Markov - Russian Mathematician - Early 1900's
- Start with an example:

"Directed weighted graph."

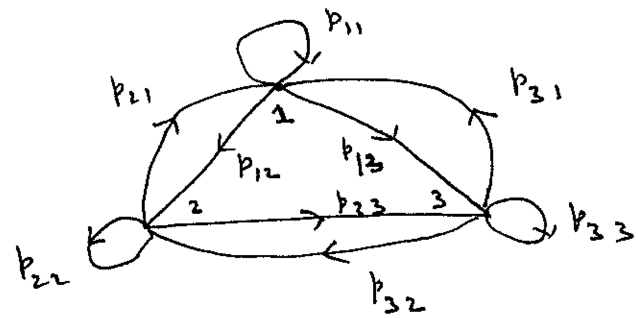


Figure 1

- A directed graph with self-loops. Associated with each edge is a probability denoting the transition probability. (one-step transition prob)

- Consider state i: $1 \le i \le 3$

  p_{ij} = Conditional prob. of staying in state j at time $(k+1)$, if at time the chain is in state i at time k.

- $p_{ij} = \text{Prob}\left[x(k+1) = j \mid x(k) = i\right]$ ⟶ (1)

- $\sum_{j=1}^{3} p_{ij} = p_{i1} + p_{i2} + p_{i3} = 1$ ⟶ (2)

- State transition prob. matrix: Stochastic matrix

$$P = \begin{array}{c|c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \end{array} \longrightarrow (3)$$

- Notice that $0 \le p_{ij} \le 1$ and $\sum_{j=1}^{3} p_{ij} = 1$ for all $i = 1, 2, 3$.

- A matrix P is called a stochastic matrix if each of the elements p_{ij} is such that $0 \le p_{ij} \le 1$ and row sum is unity

- If in addition, if the column sum is also unity:

$$\sum_{i=1}^{3} p_{ij} = 1 \quad \text{for all } j$$

then P is called doubly stochastic matrix

Example 1 Two coins 1 and 2. Coin i falls head with prob. d_i and tail with prob c_i, where $c_i + d_i = 1$. Consider, the chain

$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} d_1 & c_1 \\ c_2 & d_2 \end{bmatrix} \end{array}$$

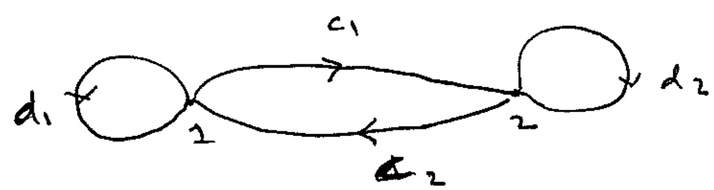


Figure 2

- Assume that if the coin falls head, you get \$1 and you give back \$1 if it falls tail.

- Let $d_1 \neq d_2$ and $d_1 > d_2$. Assume d_i and c_i are not known. The goal is to conduct experiments and decide which coin is better.

X
- This is an unsupervised learning problem, and is called Two-Armed bandit problem.

- Initially at time $k=0$, let $p_i(0)$ be the probability of starting at state i such that

$$p_1(0) + p_2(0) + p_3(0) = 1 \longrightarrow (4)$$

- At time $k > 0$, let $p(k) = (p_1(k), p_2(k), p_3(k))^T$ be a column vector where

$p_i(k)$ = The prob. that the system is at state i in time k

- $p(k)$ for $k \geq 0$ is called the <u>state probability distribution</u>

- What is the <u>dynamics</u> of $p(k)$?

$$p_i(k+1) = \cancel{\text{Prob}[x(k)=j]\, P[x(k+1)]}$$

$$= \sum_{j=1}^{3} P[\,x(k+1) = i \mid x(k) = j\,]\; p_j(k) \longrightarrow (5)$$

$\boxed{\text{Set } i=2}$

$$= P[\,x(k+1) = 2 \mid x(k) = 1\,]\; p_1(k)$$
$$+ P[\,x(k+1) = 2 \mid x(k) = 2\,]\; p_2(k)$$
$$+ P[\,x(k+1) = 2 \mid x(k) = 3\,]\; p_3(k) \longrightarrow (6)$$

$$= p_{12}\, p_1(k) + p_{22}\, p_2(k) + p_{32}\, p_3(k) \longrightarrow (7)$$

$$= \underbrace{[p_{12},\ p_{22},\ p_{32}]}_{\substack{\text{Transpose of}\\ \text{the } 2^{nd} \text{ column}\\ \text{of } P \text{ in } (3)}} \begin{bmatrix} p_1(k) \\ p_2(k) \\ p_3(k) \end{bmatrix}$$

$\therefore$

$$\begin{bmatrix} p_1(k+1) \\ p_2(k+1) \\ p_3(k+1) \end{bmatrix} = \begin{bmatrix} p_{11} & p_{21} & p_{31} \\ p_{12} & p_{22} & p_{23} \\ p_{13} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} p_1(k) \\ p_2(k) \\ p_3(k) \end{bmatrix}$$

(ii) $$\boxed{p(k+1) = P^T p(k)} \longrightarrow (8)$$

- $p(0)$ is the initial condition for this <u>first-order recurrence relation</u> in (8)

<u>Question</u>: Given $p(0)$ and P what is the long term behaviour of $p(k)$? Does $p^* = \lim_{k \to \infty} p(k)$ exist and how characterize this limit.

<u>Note</u>: It is assumed that the matrix P does <u>not</u> change in time. Such an MC is called a homogeneous chain. If P varies in time it is non-homogeneous MC. We only consider homogeneous case.

<u>Conditions for the ~~[illegible]~~ existence of p^* and classification of M.C.</u>

- A MC over a finite state is built on a directed graph.
- MC is called <u>communicating</u> if the underlying directed graph is <u>strongly</u>

Connected, that is, there is a directed path from every state i to every state j. Since each edge is associated with a non-zero probability as its weight, there is a non-zero probability of going from every state to every other state. Hence the name communicating chain.

- Examples of MC in Fig 1 and 2 are communicating MC.

- Absorbing state and absorbing chain:

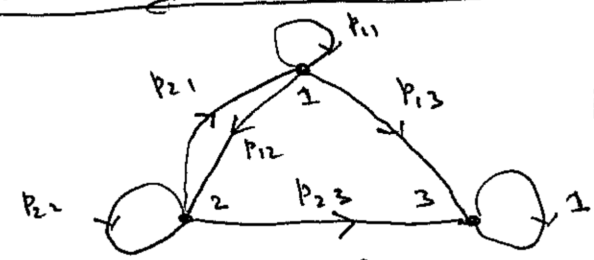


Figure 3

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

- In here, once the state 3 is reached, it stays there forever. In other words, once state 3 is reached, $p_{31} = p_{32} = 0$, no transition is possible. This is called absorbing chain

- This is an example of a non-communicating chain. State 3 is called absorbing state

- In this case in the limit, the system will find itself in state 3 and will stay there for all times.

Periodic chain / cyclic chain

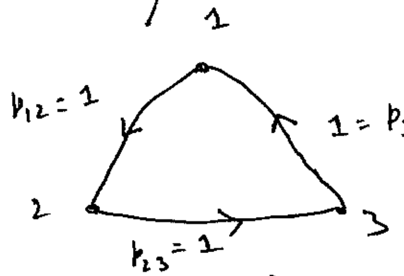


Figure 4

$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{array}$$

· Here state cycles through the three states.

Fundamental result: Theorem

Let X be an n-state M.C. with P as its one-step (homogeneous) transition prob. matrix and let $p(0)$ be the initial state distribution. Let $p(k+1) = P^T p(k)$. Then the limit ~~exist~~ $p^* = \lim_{k \to \infty} p(k)$ exists and is unique if the underchain is a communicating chain. Indeed, this limit is the solution of the linear system $p^* = P^T p^*$.

Two-step transition: $\boxed{\text{Recall } (A^T)^2 = (A^2)^T}$

$$p(k+1) = P^T p(k)$$

$$p(k+2) = P^T p(k+1) = P^T . P^T p(k) = (P^2)^T p(k)$$

$$\Rightarrow \quad p(k) = (P^k)^T p(0) \quad \longrightarrow (9)$$

Examine the entries of P^2

$n=3$

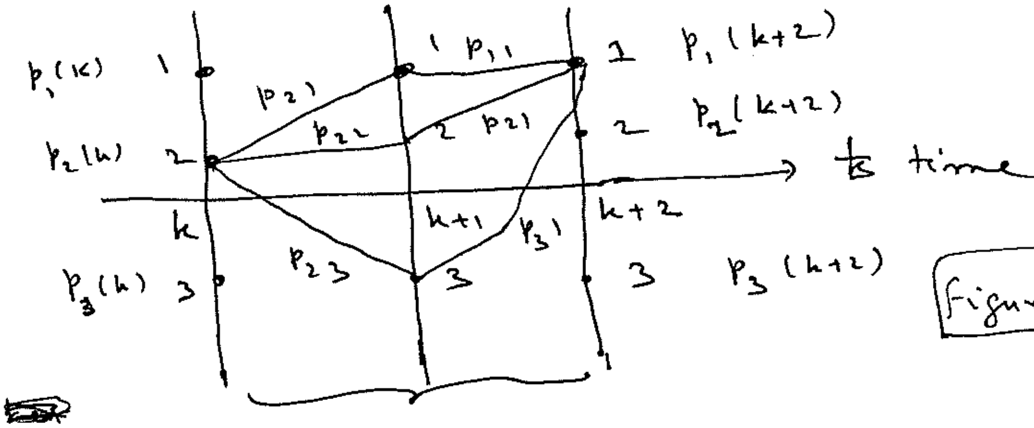


figure 5

~~$p_1(k+2)$~~

$$P^2 = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ \boxed{p_{21} \quad p_{22} \quad p_{23}} & & \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix}$$

$$= \begin{bmatrix} \boxed{}_{11} & \boxed{?}_{12} & \boxed{?}_{13} \\ \boxed{?}_{21} & \boxed{?}_{22} & \boxed{?}_{23} \\ \boxed{?}_{31} & \boxed{?}_{31} & \boxed{?}_{33} \end{bmatrix}$$

(see the illustration

$$\boxed{P^2}_{21} = p_{21} \cdot p_{11} + p_{22} \cdot p_{21} + p_{23} \cdot p_{31} \rightarrow (10)$$

$$\Rightarrow \boxed{P^{k+s} = P^k \cdot P^s} \text{ - Chapman-Kolmogorov equation.}$$

$$\boxed{P^{k+s} = P^s \cdot P^k} \rightarrow (11)$$

- **Definition of discrete time Markov process:**

$$p(x(k+1) \mid \underbrace{x(k), x(k-1), \dots x(2), x(1), x(0)}_{\text{entire past}})$$

$$= p(\underbrace{x(k+1)}_{\text{future}} \mid \underbrace{x(k)}_{\text{present}}) \rightarrow (12)$$

- A communicating chain is called Ergodic chain

Two armed bandit

Consider the chain in fig 2 with $(c_i + d_i = 1)$

$$P = \begin{bmatrix} d_1 & c_1 \\ c_2 & d_2 \end{bmatrix}$$

$\boxed{d_1 \neq d_2}$ $\boxed{d_1 > d_2}$ $\boxed{d\text{'s are not known}}$

- Initial prob. $p(0) = (1/2, 1/2)$

$$p(1) = \begin{bmatrix} d_1 & c_2 \\ c_1 & d_2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = \begin{pmatrix} \frac{d_1 + c_2}{2} \\ \frac{c_1 + d_2}{2} \end{pmatrix} \longrightarrow (13)$$

- Initial reward at time k

$$R(k) = p_1(k)\left[1 \cdot d_1 + (-1)c_1\right] + p_2(k)\left[1 \cdot d_2 + (-1)c_2\right]$$

$$= p_1(k)\left[d_1 - c_1\right] + p_2(k)\left[d_2 - c_2\right]$$

$$= p_1(k)\left[d_1 - (1-d_1)\right] + p_2(k)\left[d_2 - (1-d_2)\right]$$

$$= p_1(k)\left[2d_1 - 1\right] + p_2(k)\left[2d_2 - 1\right]$$

$$= \underbrace{p_1(k)\left[2d_1\right] + p_2(k)\left[2d_2\right]} - \underbrace{\left(p_1(k) + p_2(k)\right)}_{=1}$$

$$= \begin{pmatrix}\text{AVE. Gain} \\ \text{at time } k\end{pmatrix} - 1 \longrightarrow (14)$$

- our goal is to maximize $\boxed{\text{AVE. Gain}}$ as $k \to \infty$.

What is $p^* = \lim_{k \to \infty} p(k)$?

$$p^* = P^T p^* \longrightarrow (15)$$

$$\begin{pmatrix} p_1^* \\ p_2^* \end{pmatrix} = \begin{bmatrix} d_1 & c_2 \\ c_1 & d_2 \end{bmatrix} \begin{bmatrix} p_1^* \\ p_2^* \end{bmatrix}$$

$$\left.\begin{aligned} p_1^* &= d_1 p_1^* + c_2 p_2^* \\ p_2^* &= c_1 p_1^* + d_2 p_2^* \end{aligned}\right\} \qquad p_1^* + p_2^* = 1$$

Then

$$p_1^\alpha = d_1 p_1^\alpha + c_2 p_2^\alpha$$
$$= d_1 p_1^\alpha + c_2 (1 - p_1^\alpha)$$
$$= (d_1 - c_2) p_1^\alpha + c_2$$

$$\therefore p_1^\alpha [\underbrace{1 - d_1}_{= c_1} + c_2] = c_2$$

$$\therefore p_1^\alpha = \frac{c_2}{c_1 + c_2}$$

$$p_2^\alpha = 1 - p_1^\alpha = 1 - \frac{c_2}{c_1+c_2} = \frac{c_1}{c_1+c_2}$$

$$\therefore p^\alpha = \left(\frac{c_2}{c_1+c_2}, \frac{c_1}{c_1+c_2} \right)$$

Given $d_1 > d_2 \Rightarrow 1 - c_1 > 1 - c_2$

$$\Rightarrow c_1 < c_2$$

$$\therefore p_1^\alpha = \frac{c_2}{c_1+c_2} > \frac{c_1}{c_1+c_2} = p_2^\alpha$$

(ii) asymptotically, this method picks the coin with greater reward (d_1) with a larger probability. $\Rightarrow$ Learning has occured

. This is called the play-the-Winner rule

. Verify

$$p_1^\alpha c_1 + p_2^\alpha c_2 = \frac{2c_1c_2}{c_1+c_2} < \frac{c_1+c_2}{2}$$

(ii) $4c_1c_2 < (c_1+c_2)^2 = c_1^2 + 2c_1c_2 + c_2^2$

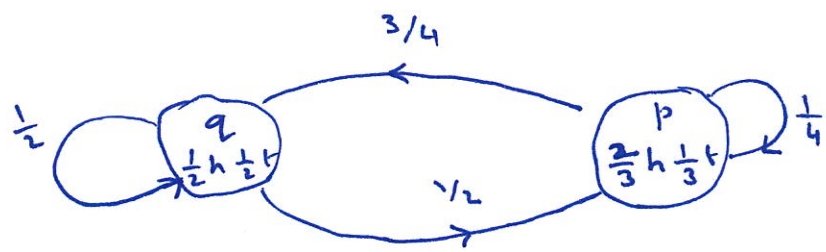
$\Rightarrow \quad 0 < c_1^2 - 2c_1c_2 + c_2^2 = (c_1-c_2)^2$

(ii) asymptotic loss is less than the initial loss

Hidden Markov Model (HMM)

- Finite set of states - with transition probability between states:- a_{ij} is the probability of transition from state i to state j
- Initial probability distribution: $\alpha_i(k)$ = prob. of being in state i at time k. $k=0 \Rightarrow$ initial distribution.
- output probability distribution: $p(o|i)$ = Probability of outputting symbol o in state i.
- State transition to a new state is followed by an output symbol

Example:- HMM: Two states q and p. output symbols are h and t



Initial distribution:

$\alpha(q) = 1$

$\alpha(p) = 0$

Consider three problems:

1) Given an HMM, what is the prob. of an output sequence?
2) Given an HMM and an output sequence, what is the most likely sequence of states?
3) Given that an HMM has n-states and given an output sequence, what is the most likely

HMM?

NOTE · Only the third problem concerns HMM. ~~...~~ In the other two, the model is given and answers are computed in poly. time using Dynamic Programming. There is no known poly. time algorithm for the 3rd problem

1) How probable is an output sequence?

Given an HMM: What is $P[O_0 O_1 O_2 \cdots O_T]$ of output sequence $O = O_0 O_1 \cdots O_T$ of length $(T+1)$.

~~Start at t=0: At t=0 no transition~~.

· For each state i and each initial segment of outputs $O_0 O_1 \cdots O_t$ of length $(t+1)$, the prob. of observing $O_0 O_1 \cdots O_t$ ending in state i is to be computed.

Time $= O(n^2 T)$ · At $t=0$: No transition: Prob. of observing O_0 ending in state i ~~...~~ is the initial prob. of starting in state i time prob. of observing O_0 in state i:

space $= O(n)$

Alg: · $p(O_0, i) = \alpha(i)\, p(O_0 | i) \quad \forall i$

· For $t = 1$ to T

$$P[O_0, O_1 \cdots O_t, i] = \sum_j P[O_0 O_1 \cdots O_{t-1}, j] \cdot a_{ji}\, P[O_t, i]$$

· Prob of observing $O_0 O_1 \cdots O_t$ ending in state i = sum of the prob. over all states j of observing $O_0 O_1 \cdots O_{t-1}$ ending in state j * prob. of transition from j to i and observing O_t

· The time to compute this $O(n^2 T)$. n^2 is due to the fact that we have to examine transition to each of the n states in each time time unit

<u>Example:</u> $P(hhht) = ?$

	q	p
$t=0$	$P[O_0 = h] = 1/2$	—
$t=1$	$a_{qq} = 1/2 \cdot p(h \mid q) = \frac{1}{2}$	$a_{qp} = 1/2 \quad p[h \mid p] = 2/3$
$P[hh,q] =$	$P[O_0 = h]\, a_{qq}\, p[h \mid q]$ $= \frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$	$P[O_0 = h]\, a_{qp}\, p[h \mid p]$ $= 1/2 \; 1/2 \; 2/3 = 1/6 = P[hh,p]$
$t=2$ $P[hhh,q]$	$P[hh,q]\, a_{qq}\, P[h \mid q]$ $+ P[hh,p]\, a_{pq}\, P[h \mid q]$ $= \frac{1}{8} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{\cancel{6}_2} \cdot \frac{3}{4} \cdot \frac{1}{2}$ $\cancel{= \frac{1}{16}\left[\frac{1}{2} + \frac{3}{4}\right] = \frac{1}{16} \times \frac{7}{4}}$ $\cancel{\frac{7}{64}}$ $= \frac{1}{4}\left(\frac{1}{8} + \frac{1}{4}\right) = \frac{1}{4}\,\frac{3}{8}$ $= \frac{3}{32}$	$P[hh,q]\, a_{qp}\, P[h \mid p]$ $+ P[hh,p]\, a_{pp}\, P[h \mid p]$ $\cancel{= \frac{1}{8} \cdot \frac{2}{4_2} \cdot \frac{2}{3} + \frac{1}{8} \cdot \frac{1}{4_2} \cdot \frac{2}{3}}$ $= P[hhh,p]$ $\cancel{= \frac{1}{16} \cdot \frac{1}{3} + \frac{1}{6} \cdot \frac{1}{3} =}$ $= \frac{1}{8} \cdot \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{6} \cdot \frac{1}{\cancel{4}_2} \cdot \frac{2}{3}$ $= \frac{1}{3}\left[\frac{1}{8} + \frac{1}{12}\right] = \frac{1}{3}\,\frac{3+2}{24} = \frac{5}{72}$
$t=3$	$P[hhht,q]$ $= P[hhh,q]\, a_{qq}\, p(t \mid q)$ $+ P[hhh,p]\, a_{pq}\, p(t,q)$ $= \frac{3}{32} \cdot \frac{1}{2} \cdot \frac{1}{2} + \frac{5}{72} \cdot \frac{3}{4} \cdot \frac{1}{2}$ $= \frac{19}{384}$	$P[hhht,p]$ $= P[hhh,q]\, a_{qp}\, p[t \mid p]$ $+ P[hhh,p]\, a_{pp}\, p[t,p]$ $= \frac{3}{32} \cdot \frac{1}{2} \cdot \frac{1}{3}$ $+ \frac{5}{72} \cdot \frac{1}{4} \cdot \frac{1}{3} = \frac{37}{64 \times 27}$ ~~0.0709~~

$$\text{Prob}[hhht] = \frac{19}{384} + \frac{37}{64 \times 27} = 0.0709$$

2. The most likely sequence :- The Viterbi algorithm

- Given HMM and observation $O_0 O_1 \cdots O_T$, what is the most likely sequence of states?
- This is a slight modification of DP algorithm given above
- For $t = 0, 1, 2, \cdots T$, for each state i, compute the prob of the most likely sequence of states producing $O_0\, O_1\, O_2 \cdots O_{t-1}$ ending in state j. Then multiply it by the prob. of going from j to i and producing O_t.
- Then, select the j for which the product is large.
- We record the max and where the max comes from. (Need not be unique)
- The time is $O(n^2 T)$, space $O(nT)$

Instead of sum, take max.

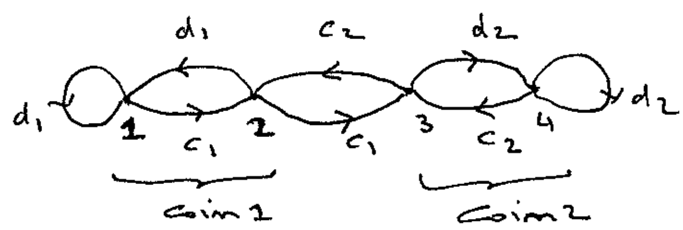
Example :- output is h h h t

	q Prob./state	p Prob./state
$t=0$	$\frac{1}{2}$, q	0, p
$t=1$	$\frac{1}{2}\frac{1}{2}\frac{1}{2} = \frac{1}{8}$, q	$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{2}{3} = \frac{1}{6}$, p
$t=2$	$\max\left\{\frac{1}{8}\cdot\frac{1}{2}\cdot\frac{1}{2}, \frac{1}{6}\cdot\frac{3}{4}\cdot\frac{1}{2}\right\} = \frac{3}{48}$, p	$\max\left\{\frac{1}{8}\cdot\frac{1}{2}\cdot\frac{3}{2}, \frac{1}{6}\cdot\frac{1}{4}\cdot\frac{2}{3}\right\} = \frac{1}{24}$, q
$t=3$	$\max\left\{\frac{3}{48}\cdot\frac{1}{2}\cdot\frac{1}{2}, \frac{1}{24}\cdot\frac{3}{4}\cdot\frac{1}{2}\right\} = \frac{1}{64}$ q or p	$\max\left\{\frac{3}{48}\cdot\frac{1}{2}\cdot\frac{1}{3}, \frac{1}{24}\cdot\frac{1}{4}\cdot\frac{1}{3}\right\} = \frac{1}{96}$ q

State sequence qppq or qpqq - a tie

Problems : MC and HMM

1)



a) Find the transition matrix $P\big|_{4\times 4}$

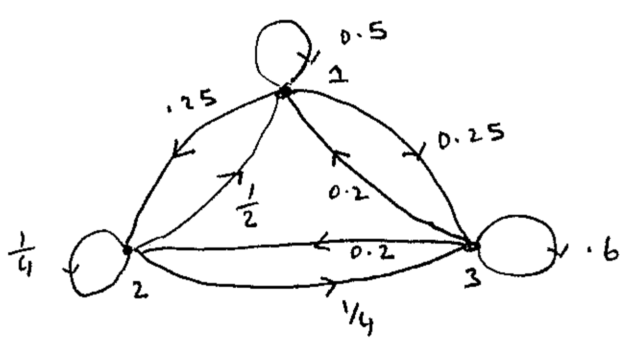
b) Compute p^* where $p^* = P^T p^*$

c) Set $d_1 = 0.8$ and $d_2 = 0.4$. Let $p(0) = (1/4, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})^T$ and ~~iterate~~ iterate $p(k+1) = P^T p(k)$

Plot the components of $p(k)$ vs k for $k = 0$ to $k = 100$

d) Compare the limiting value of $p(k)$ with p^* found in (b)

2)



State	output distribution	
	a	b
1	3/4	1/4
2	1/2	1/2
3	1/4	3/4

From each state there can be two outputs : a and b. Distribution of output as a function of the state is given in Table above.

a) what is the probability output sequence aabb?

b) What is the most likeli sequence of states corresponding to output aabb?

A COLLECTION OF RESULTS ON M.C.

- Let P be the transition matrix of a M.C.
- This M.C. is said to have a regular transition matrix if there is no absorbing state and there is a non-zero probability of going from any state to any other state. That is, the graph is strongly connected.

<u>Theorem 1</u>: Let $P = [p_{ij}]$ and $p_{ij} > 0$. Let ε be the smallest entry in P and $\varepsilon > 0$. Let $x \in \mathbb{R}^n$ and $M_0 = \max_i \{x_i\}$, $m_0 = \min_i \{x_i\}$ and let M_1 and m_1 be the corresponding quantities for Px. Then

$$M_1 < M_0 \quad \text{and} \quad m_1 \geq m_0$$

$$(M_1 - m_1) \leq (1 - 2\varepsilon)(M_0 - m_0)$$

<u>Proof</u>: . Let $\bar{x}$ be obtained from x by replacing all of its elements except m_0 by M_0. Then $x \leq \bar{x}$ (element wise Comparison)

- Let $y = P\bar{x}$ be a vector. Then, each y_i is the average of the elements of $\bar{x}$

$$\text{(ie)} \quad y_i = \sum_{j=1}^{n} p_{ij} \bar{x} \qquad 1 \leq i \leq n.$$

<u>Theorem 2</u>: Let $P = [p_{ij}]$, $p_{ij} > 0$. Then

a) $P^n \to A$

b) Each row of A is the same: $\alpha = (\alpha_1, \alpha_2 \cdots \alpha_n)$

(ii) $A = u\alpha$ where $u = (1, 1, \cdots 1)^T$

c) $\alpha_i > 0$ for all i

<u>Proof</u>: Let $P = [p_{ij}]$ and let $\varepsilon > 0$ be the min. element of P. Let e_j be the j^{th} unit vector.

. Let M_k, m_k be the max and min elements of $P^k e_j$.

. From $P^k e_j = P(P^{k-1} e_j)$ and from Theorem (1) above, it follows that

$$M_1 \geq M_2 \geq M_3 \geq \cdots$$
$$m_0 \leq m_1 \leq m_3 \leq \cdots$$

and
$$\underbrace{M_k - m_k}_{d_k} \leq (1 - 2\varepsilon) \underbrace{(M_{k-1} - m_{k-1})}_{d_{k-1}}$$

(ii) $d_k \leq (1 - 2\varepsilon)\, d_{k-1}$

. Iterating $d_k \leq (1 - 2\varepsilon)^k d_0 \to 0$ as $k \to \infty$

Thus, $P^k e_j \to$ <u>to a Vector with the same element</u>.

. Let α_j be the common value. Then

But $\beta P = \beta$

$\therefore \beta P^2 = (\beta P) P = \beta P = \beta$

$\therefore \beta P^n = \beta \longrightarrow (**)$

Combining $(*)$ and $(**)$ $\Rightarrow$ $\alpha = \beta$ (i.e) unique.

<u>Summary</u>: Let $P = [p_{ij}]$ $\quad p_{ij} > 0$

1) $P^n \longrightarrow A = \begin{bmatrix} \alpha \\ \alpha \\ \vdots \\ \alpha \end{bmatrix}$ $\qquad \alpha = (\alpha_1, \alpha_2 \ldots \alpha_n)$, $\quad \alpha_i > 0, \ \sum_{i=1}^{n} \alpha_i = 1$

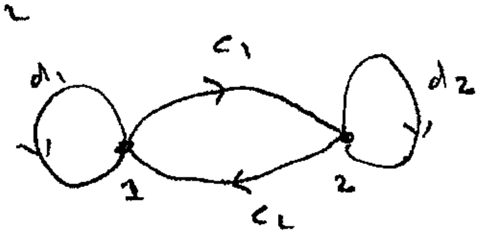
2) $\alpha P = \alpha$ (i.e) α is the left eigenvector corresponding to the e.v $\lambda = 1$.

Thus, given P, we can easily find P^* the limit of P^k as $k \to \infty$, by simply solving for

$$\alpha P = \alpha.$$

3) α is called the stationary distribution.

<u>Example</u>: $n = 2$



$$P = \begin{array}{c c} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} 1 \\ 2 \end{array} & \begin{bmatrix} d_1 & c_1 \\ c_2 & d_2 \end{bmatrix} \end{array}$$