$X_1, \ldots, X_n$ IID → Statistics

$X_1, X_2, X_3, \ldots$

1st order dependence
markov chain

General theory behind many of the results we've already seen.

0 1 2 3 ...

$X_0 = 2$
$X_1 = 1$
$X_2 = 3$ ...

Chapter 8: Markov Chains

it only matters where you are, not where you've been...

8.1 Introduction

A.A.Markov
1856-1922

So far, we have examined several stochastic processes using transition diagrams and First-Step Analysis.
The processes can be written as $\{X_0, X_1, X_2, \ldots\}$,
where X_t is the state at time t.

On the transition diagram, X_t corresponds to which box we are in at step t.

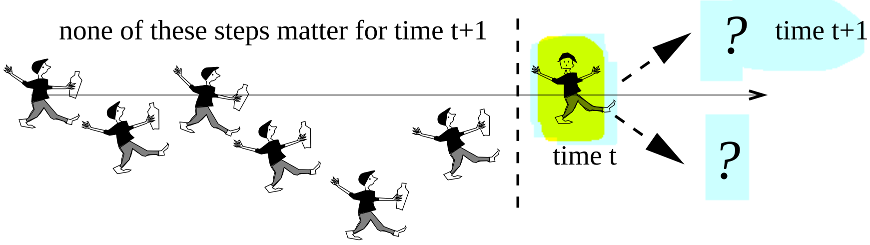
In the Gambler's Ruin (Section 2.7), X_t is the amount of money the gambler possesses after toss t. In the model for gene spread (Section 3.7), X_t is the number of animals possessing the harmful allele A in generation t.

The processes that we have looked at via the transition diagram have a crucial property in common: X_{t+1} depends ONLY on X_t.

It does <u>not</u> depend upon $X_0, X_1, \ldots, X_{t-1}$.

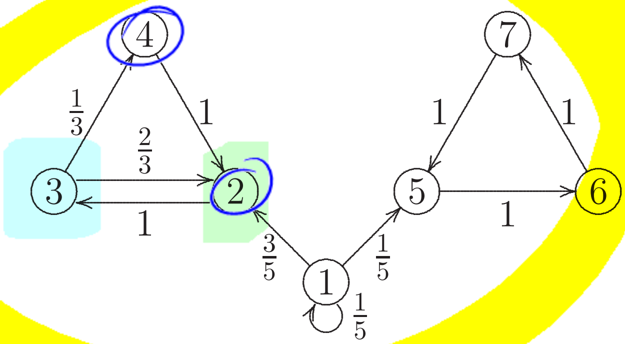
Processes like this are called Markov Chains.

Example: Random Walk (see Chapter 4)



$$\boxed{\text{In a Markov chain, the future depends only upon the present: NOT upon the past.}}$$

The text-book image of a Markov chain has a flea hopping about at random on the vertices of the transition diagram, according to the probabilities shown.



The transition diagram above shows a system with 7 possible states:

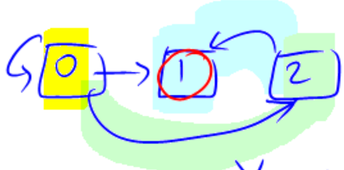
State Space, $S = \{1,2,3,4,5,6,7\}$.

Questions of interest

- Ch2 Starting from state 1, what is the probability of ever reaching state 7? FSA
- Ch3 Starting from state 2, what is the expected time taken to reach state 4? FSA
- Ch9 Starting from state 2, what is the long-run proportion of time spent in state 3? Equilibrium → Ch 9
- Ch8/9 Starting from state 1, what is the probability of being in state 2 at time t? Does the probability converge as $t \to \infty$, and if so, to what?

We have been answering questions like the first two using first-step analysis since the start of STATS 325. In this chapter we develop a unified approach to all these questions using the matrix of transition probabilities, called the transition matrix.

Ass 4 Due Mon 4pm.



$X_1 = 2$ $X_2 = 1$

8.2 Definitions

The Markov chain is the process $X_0, X_1, X_2, \ldots$

Definition: The **state** of a Markov chain at time t is the value of X_t. ie. which box we are in at time t

For example, if $X_t = 6$, we say the process is in state 6 at time t.

Definition: The **state space** of a Markov chain, S, is the set of values that each X_t can take. For example, $S = \{1, 2, 3, 4, 5, 6, 7\}$. ($S$ = {all different boxes on diagram})

Let S have size N (possibly infinite).

Definition: A **trajectory** of a Markov chain is a particular set of values for $X_0, X_1, X_2, \ldots$

For example, if $X_0 = 1$, $X_1 = 5$, and $X_2 = 6$, then the trajectory up to time $t = 2$ is 1, 5, 6.

More generally, if we refer to the trajectory $s_0, s_1, s_2, s_3, \ldots$, we mean that $X_0 = s_0$, $X_1 = s_1$, $X_2 = s_2$, $X_3 = s_3, \ldots$

'Trajectory' is just a word meaning "path".

Markov Property → first-order dependence

The basic property of a Markov chain is that only the most recent point in the trajectory affects what happens next.

This is called the Markov property.
It means that X_{t+1} depends upon X_t, but it does not depend upon $X_{t-1}, X_{t-2}, \ldots, X_1, X_0$.

$X_4 = 2$ $X_5 = ?$ if I know X_4, everything else beforehand is irrelevant.

If I didn't know X_4, it would be the most recent info that I DO HAVE that is the only thing relevant for the distn of X_5.

We formulate the Markov Property in mathematical notation as follows:

$$\boxed{\mathbb{P}(X_{t+1} = s \,|\, X_t = s_t, \cancel{X_{t-1} = s_{t-1}, \ldots}, X_0 = s_0) \;=\; \mathbb{P}(X_{t+1} = s \,|\, X_t = s_t),}$$

for all $t = 1, 2, 3, \ldots$ and for all states $s_0, s_1, \ldots, s_t, s$.

Explanation:

$$\mathbb{P}(X_{t+1} = s \quad | \quad X_t = s_t, \quad \underbrace{\cancel{X_{t-1} = s_{t-1}, X_{t-2} = s_{t-2}, \ldots, X_1 = s_1, X_0 = s_0}})$$

distribution of X_{t+1} ...

... depends upon X_t ...

... but anything that is known relating to earlier times $0, 1, \ldots, t-1$ doesn't matter.

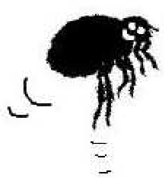
Definition: Let $\{X_0, X_1, X_2, \ldots\}$ be a sequence of discrete random variables. Then $\{X_0, X_1, X_2, \ldots\}$ is a **Markov chain** if it satisfies the Markov property:

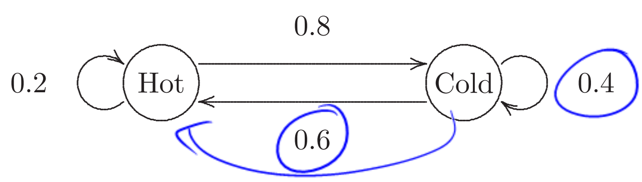
$$\mathbb{P}(X_{t+1} = s \,|\, X_t = s_t, X_{t-1} = s_{t-1}, \ldots, X_0 = s_0) = \mathbb{P}(X_{t+1} = s \,|\, X_t = s_t)$$

for all times $t = 1, 2, 3, \ldots$ and for all states $s_0, s_1, \ldots, s_t, s$.

8.3 The Transition Matrix

We have seen many examples of **transition diagrams** to describe Markov chains. The transition diagram is so-called because it shows the transitions between different states.





We can also summarize the probabilities in a **matrix:**

$$\begin{array}{cc} & \quad X_{t+1} \\ & \begin{array}{cc} \text{Hot} & \text{Cold} \end{array} \\ X_t \begin{array}{c} \text{Hot} \\ \text{Cold} \end{array} & \begin{pmatrix} 0.2 & 0.8 \\ 0.6 & 0.4 \end{pmatrix} \end{array}$$

The matrix describing the Markov chain is called the transition matrix.
It is the most important tool for analysing Markov chains. except your brain.

Transition Matrix

$$
\begin{array}{cc}
 & \overbrace{\longleftarrow \text{ list all states } \longrightarrow}^{X_{t+1}} \quad \text{To} \\
\text{From } X_t \left\{ \begin{array}{c} \text{list} \\ \text{all} \\ \text{states} \end{array} \right. & \left(\begin{array}{c} \texttt{insert} \\ \texttt{probabilities} \\ p_{ij} \end{array} \right) \begin{array}{l} \longleftarrow \text{rows add to 1} \\ \longleftarrow \text{rows add to 1} \end{array}
\end{array}
$$

← going FROM i TO j

because arrows OUT of every state must sum to 1 (got to go somewhere out of the state)

The transition matrix is usually given the symbol P with elements (p_{ij}).

In the transition matrix P:

- the ROWS represent NOW, or FROM (X_t);
- the COLUMNS represent NEXT, or TO (X_{t+1});
- entry (i, j) is the CONDITIONAL probability that NEXT $= j$, GIVEN that NOW $= i$: ie. the probability of going FROM state i TO state j.

$$\boxed{p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i) = \mathbb{P}_{X_t = i}(X_{t+1} = j)}$$

Notes: 1. The transition matrix P must list *all* possible states in the state space S.

2. P is a *square matrix* $(N \times N)$, because X_{t+1} and X_t both take values in the same state space S (of size N).

3. The **rows** of P should each sum to 1:

$$\sum_{j=1}^{N} p_{ij} = \sum_{j=1}^{N} \mathbb{P}(X_{t+1} = j \mid X_t = i) = \sum_{j=1}^{N} \mathbb{P}_{\{X_t = i\}}(X_{t+1} = j) = 1.$$

This simply states that X_{t+1} *must* take one of the listed values.

4. The **columns** of P do **not** in general sum to 1.

Definition: Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space S, where S has size N (possibly infinite). The **transition probabilities** of the Markov chain are

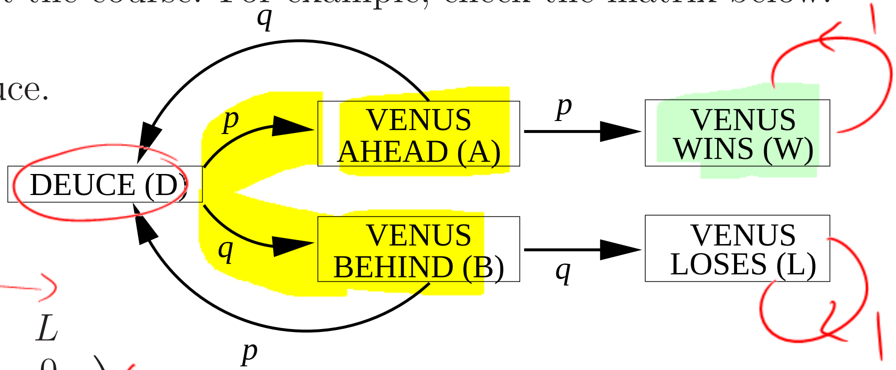
$$p_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i) \quad \text{for } i, j \in S \text{ and } t = 0, 1, 2, \ldots$$

Definition: The **transition matrix** of the Markov chain is $P = (p_{ij})$.

8.4 Example: setting up the transition matrix

We can create a transition matrix for any of the transition diagrams we have seen in problems throughout the course. For example, check the matrix below.

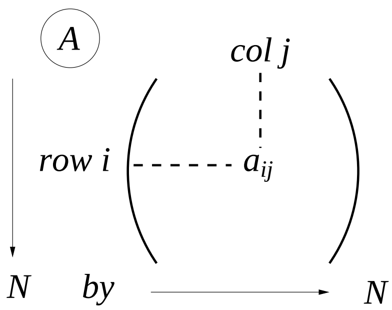
Example: Tennis game at Deuce.



$$\begin{array}{c} \\ D \\ A \\ B \\ W \\ L \end{array} \begin{array}{c} \begin{array}{ccccc} D & A & B & W & L \end{array} \\ \begin{pmatrix} 0 & p & q & 0 & 0 \\ q & 0 & 0 & p & 0 \\ p & 0 & 0 & 0 & q \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

sum = 1 for each row.

8.5 Matrix Revision



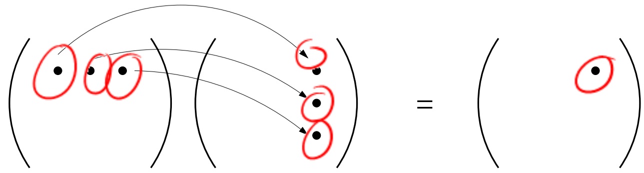
Notation

Let A be an $N \times N$ matrix.

We write $A = (a_{ij})$,
i.e. A comprises elements a_{ij}.

The (i, j) element of A is written both as a_{ij} and $(A)_{ij}$:
e.g. for matrix A^2 we might write $(A^2)_{ij}$.

Matrix multiplication



Let $A = (a_{ij})$ and $B = (b_{ij})$ be $N \times N$ matrices.

The product matrix is $A \times B = AB$, with elements $(AB)_{ij} = \sum_{k=1}^{N} a_{ik} b_{kj}$.

Summation notation for a matrix squared

Let A be an $N \times N$ matrix. Then

$$(A^2)_{ij} = \sum_{k=1}^{N} (A)_{ik}(A)_{kj} = \sum_{k=1}^{N} a_{ik} a_{kj}.$$

Pre-multiplication of a matrix by a vector

Let A be an $N \times N$ matrix, and let $\boldsymbol{\pi}$ be an $N \times 1$ column vector: $\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_N \end{pmatrix}$.

We can pre-multiply A by $\boldsymbol{\pi}^T$ to get a $1 \times N$ row vector, $\boldsymbol{\pi}^T A = \left((\boldsymbol{\pi}^T A)_1, \ldots, (\boldsymbol{\pi}^T A)_N\right)$, with elements

$$(\boldsymbol{\pi}^T A)_j = \sum_{i=1}^{N} \pi_i a_{ij}.$$

π^T A N N N $= \pi^T A$

8.6 The *t*-step transition probabilities

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $S = \{1, 2, \ldots, N\}$.

Recall that the elements of the transition matrix P are defined as:

$$(P)_{ij} = p_{ij} = \mathbb{P}(X_1 = j \mid X_0 = i) = \mathbb{P}(X_{n+1} = j \mid X_n = i) \quad \text{for any } n.$$

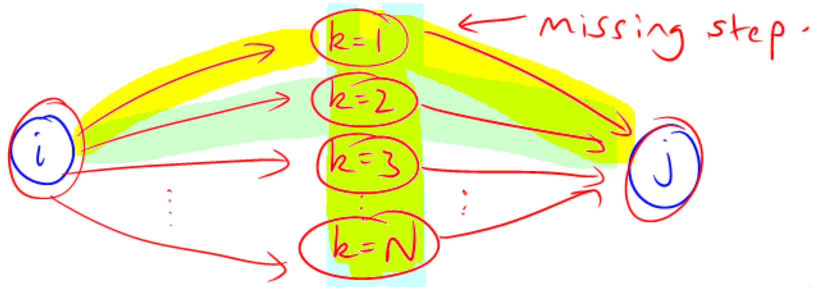
p_{ij} is the probability of making a transition FROM state i TO state j in a SINGLE step.

Question: what is the probability of making a transition from state i to state j over **two** steps?

ie. what is $\mathbb{P}(X_2 = j \mid X_0 = i)$?

$i \xrightarrow{\text{1 step}} j$ Prob $= p_{ij}$

$i \to ? \to j$ Prob $= ?$



We are seeking $\mathbb{P}(X_2 = j \mid X_0 = i)$. Use the Partition Theorem:

$$\mathbb{P}(X_2=j \mid X_0=i) = \mathbb{P}_i(X_2=j) \quad \text{using subscript notation from Ch2}$$

$$= \sum_{k=1}^{N} \mathbb{P}_i(X_2=j \mid X_1=k)\, \mathbb{P}_i(X_1=k) \quad \text{Partition Thm.}$$

$$= \sum_{k=1}^{N} \mathbb{P}(X_2=j \mid X_1=k, \cancel{X_0=i})\, \mathbb{P}(X_1=k \mid X_0=i)$$

$$= \sum_{k=1}^{N} \mathbb{P}(X_2=j \mid X_1=k)\, \mathbb{P}(X_1=k \mid X_0=i) \quad \text{by Markov Property}$$

$$= \sum_{k=1}^{N} p_{kj}\, p_{ik} \quad \text{by definitions}$$

$$= \sum_{k=1}^{N} p_{ik}\, p_{kj}$$

$$= (P^2)_{ij} \quad \text{(see Matrix Revision)}$$

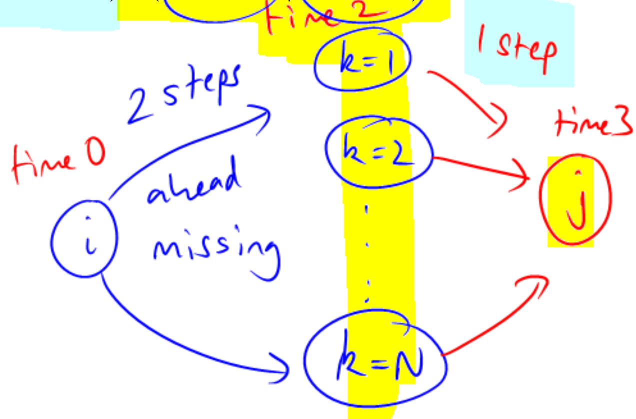
NOTE: $(P^2)_{ij}$ NOT $(p_{ij})^2$

The two-step transition probabilities are therefore given by the matrix P^2:

$$\boxed{\mathbb{P}(X_2=j \mid X_0=i) = \mathbb{P}(X_{n+2}=j \mid X_n=i) = (P^2)_{ij} \quad \text{for any } n.}$$

3-step transitions: We can find $\mathbb{P}(X_3 = j \mid X_0 = i)$ similarly, but conditioning on the state at time 2:

$$\begin{aligned} \mathbb{P}(X_3 = j \mid X_0 = i) &= \sum_{k=1}^{N} \mathbb{P}(X_3 = j \mid X_2 = k)\mathbb{P}(X_2 = k \mid X_0 = i) \\ &= \sum_{k=1}^{N} p_{kj} (P^2)_{ik} \\ &= (P^3)_{ij}. \end{aligned}$$



The three-step transition probabilities are therefore given by the matrix P^3:

$$\boxed{\mathbb{P}(X_3 = j \,|\, X_0 = i) = \mathbb{P}(X_{n+3} = j \,|\, X_n = i) = \left(P^3\right)_{ij} \quad \text{for any } n.}$$

General case: t-step transitions

The above working extends to show that the t-step transition probabilities are given by the matrix P^t for any t: (doing an induction in our heads)

$$\boxed{\mathbb{P}(X_t = j \,|\, X_0 = i) = \mathbb{P}(X_{n+t} = j \,|\, X_n = i) = (P^t)_{ij} \text{ for any } n.}$$

We have proved the following Theorem. $(P^t)_{ij}$ is NOT the same as $(P_{ij})^t$.

Theorem 8.6: Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with $N \times N$ transition matrix P. Then the t-step transition probabilities are given by the matrix P^t. That is,

$$\mathbb{P}(X_t = j \,|\, X_0 = i) = \left(P^t\right)_{ij}.$$

It also follows that

$$\mathbb{P}(X_{n+t} = j \,|\, X_n = i) = \left(P^t\right)_{ij} \quad \text{for any } n. \qquad \square$$

8.7 Distribution of X_t

X_t = which box out of $\boxed{1}$ $\boxed{2}$ ... $\boxed{N}$ we are in at time t.

Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with state space $S = \{1, 2, \ldots, N\}$.

Now each X_t is a random variable, so it has a probability distribution.

We can write the probability distribution of X_t as an $N \times 1$ vector.

For example, consider X_0. Let $\boldsymbol{\pi}$ be an $N \times 1$ vector denoting the probability distribution of X_0:

$$\underset{\sim}{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{pmatrix} = N \begin{pmatrix} \mathbb{P}(X_0 = 1) \\ \mathbb{P}(X_0 = 2) \\ \vdots \\ \mathbb{P}(X_0 = N) \end{pmatrix} \quad N \times 1 \text{ vector.}$$

In the flea model, this corresponds to the flea choosing at random which vertex (box) it starts off from at time 0, such that

$$\mathbb{P}(\text{flea chooses vertex } i \text{ to start}) = \pi_i .$$

Notation: we will write $X_0 \sim \underset{\sim}{\pi}^T$ to denote that the row vector of probabilities is given by the row vector $\boldsymbol{\pi}^T$.

Probability distribution of X_1

$X_0 \sim \underset{\sim}{\pi}^T$

Use the Partition Rule, conditioning on X_0:

$$\mathbb{P}(X_1 = j) = \sum_{i=1}^{N} \underbrace{\mathbb{P}(X_1 = j \mid X_0 = i)}_{p_{ij}} \underbrace{\mathbb{P}(X_0 = i)}_{\pi_i}$$

$$= \sum_{i=1}^{N} p_{ij} \pi_i \quad \text{by definitions}$$

$$= \sum_{i=1}^{N} \pi_i p_{ij} \quad \text{just swapping round}$$

$$\mathbb{P}(X_1 = j) = (\underset{\sim}{\pi}^T P)_j$$

pre-multiplication of matrix P by row vector $\underset{\sim}{\pi}^T$: see §8.5.

This shows that $\mathbb{P}(X_1 = j) = (\underset{\sim}{\pi}^T P)_j$

The row vector $\boldsymbol{\pi}^T P$ is therefore the probability distribution of X_1:

$$\boxed{\begin{array}{l} X_0 \sim \underset{\sim}{\pi}^T \\ \Rightarrow X_1 \sim \underset{\sim}{\pi}^T P \end{array}}$$

Probability distribution of X_2

Using the Partition Rule as before, conditioning again on X_0:

$$\mathbb{P}(X_2 = j) = \sum_{i=1}^{N} \mathbb{P}(X_2 = j \mid X_0 = i)\mathbb{P}(X_0 = i) = \sum_{i=1}^{N} \left(P^2\right)_{ij} \pi_i = \left(\boldsymbol{\pi}^T P^2\right)_j .$$

The row vector $\boldsymbol{\pi}^T P^2$ is therefore the probability distribution of X_2:

$$\boxed{\begin{array}{lcl} X_0 & \sim & \boldsymbol{\pi}^T \\ X_1 & \sim & \boldsymbol{\pi}^T P \\ X_2 & \sim & \boldsymbol{\pi}^T P^2 \\ & \vdots & \\ X_t & \sim & \boldsymbol{\pi}^T P^t. \end{array}}$$

These results are summarized in the following Theorem.

Theorem 8.7: Let $\{X_0, X_1, X_2, \ldots\}$ be a Markov chain with $N \times N$ transition matrix P. If the probability distribution of X_0 is given by the $1 \times N$ row vector $\boldsymbol{\pi}^T$, then the probability distribution of X_t is given by the $1 \times N$ row vector $\boldsymbol{\pi}^T P^t$. That is,

$$X_0 \sim \underset{\sim}{\pi}^T \quad \Rightarrow \quad X_t \sim \underset{\sim}{\pi}^T P^t$$

$\underset{\sim}{\pi}^T$ etc $1 \times N$
all vectors

Note: The distribution of X_t is $X_t \sim \underset{\sim}{\pi}^T P^t$

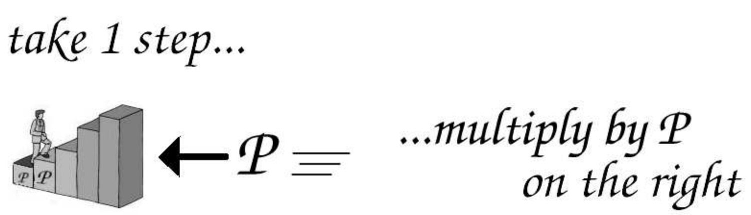
The distribution of X_{t+1} is $X_{t+1} \sim \underset{\sim}{\pi}^T P^{t+1} = \underset{\sim}{\pi}^T P^t * P$

Taking one step in the Markov chain corresponds to multiplying by P on the right.

Note: The t-step transition matrix is P^t

The $(t + 1)$-step transition matrix is $P^{t+1} = P^t * P$

Again, taking one step in the Markov chain corresponds to multiplying by P on the right.

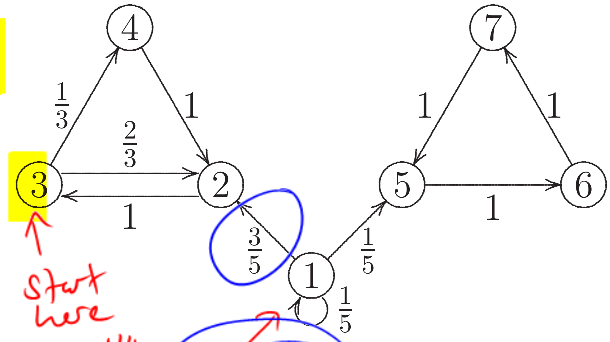


Grumpy → Sleepy → - - - → Dopey

8.8 Trajectory Probability

Recall that a trajectory is a sequence of values for $X_0, X_1, \ldots, X_t$.

Because of the Markov Property, we can find the probability of any trajectory by multiplying together the starting probability and all subsequent single-step probabilities.



Start here w.p. 1/4

Start here w.p. 3/4

Example: Let $X_0 \sim (\frac{3}{4}, 0, \frac{1}{4}, 0, 0, 0, 0)$. What is the probability of the trajectory 1, 2, 3, 2, 3, 4?

$$\mathbb{P}(1,2,3,2,3,4) = \mathbb{P}(X_0 = 1) * p_{12} * p_{23} * p_{32} * p_{23} * p_{34}$$

$$= \frac{3}{4} * \frac{3}{5} * 1 * \frac{2}{3} * 1 * \frac{1}{3}$$

$$= \frac{1}{10}$$

Proof in formal notation using the Markov Property:

Let $X_0 \sim \boldsymbol{\pi}^T$. We wish to find the probability of the trajectory $s_0, s_1, s_2, \ldots, s_t$. (1 2 3 2 3 4)

$\mathbb{P}(X_0 = s_0, \underbrace{X_1 = s_1, \ldots}_{B}, \underbrace{X_t = s_t}_{A})$ ← intersection prob: $\mathbb{P}(X_0 = s_0 \cap X_1 = s_1 \cap X_2 = s_2 \cap \ldots)$

$= \mathbb{P}(X_t = s_t \,|\, X_{t-1} = s_{t-1}, \ldots, \cancel{X_0 = s_0}) \times \mathbb{P}(X_{t-1} = s_{t-1}, \ldots, X_0 = s_0)$ — $\mathbb{P}(A \,|\, B) * \mathbb{P}(B)$; irrelevant

$= \mathbb{P}(X_t = s_t \,|\, X_{t-1} = s_{t-1}) \times \mathbb{P}(X_{t-1} = s_{t-1}, \ldots, X_0 = s_0)$ (Markov Property)

$= p_{s_{t-1}, s_t} \mathbb{P}(X_{t-1} = s_{t-1} \,|\, X_{t-2} = s_{t-2}, \ldots \cancel{X_0 = s_0}) \times \mathbb{P}(X_{t-2} = s_{t-2}, \ldots, X_0 = s_0)$

$\vdots$

$= p_{s_{t-1}, s_t} \times p_{s_{t-2}, s_{t-1}} \times \ldots \times p_{s_0, s_1} \times \mathbb{P}(X_0 = s_0)$

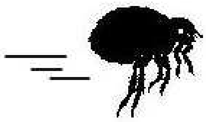
$= p_{s_{t-1}, s_t} \times p_{s_{t-2}, s_{t-1}} \times \ldots \times p_{s_0, s_1} \times \pi_{s_0}.$

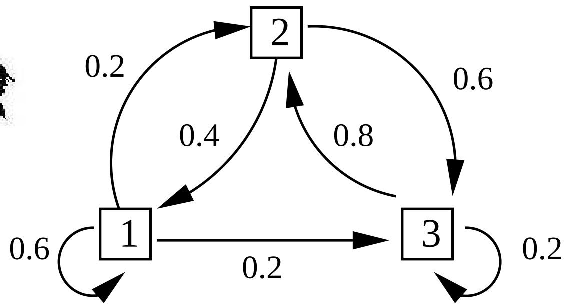
$\mathbb{P}(A \cap B) = \mathbb{P}(A \,|\, B)\, \mathbb{P}(B)$

$\mathbb{P}(X_t = s_t \cap \text{all previous history}) = \mathbb{P}(X_t = s_t \,|\, \text{all prev hist})\, \mathbb{P}(\text{all prev history})$

8.9 Worked Example: distribution of X_t and trajectory probabilities

Purpose-flea zooms around the vertices of the transition diagram opposite. Let X_t be Purpose-flea's state at time t $(t = 0, 1, \ldots)$.





(a) Find the transition matrix, P.

Answer: $P = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix}$

(b) Find $\mathbb{P}(X_2 = 3 \mid X_0 = 1)$. $\Rightarrow$ need P^2 but ONLY ONE elt of P^2 is being asked for.

$$\mathbb{P}(X_2 = 3 \mid X_0 = 1) = \left(P^2\right)_{13} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & 0.2 \\ \cdot & \cdot & 0.6 \\ \cdot & \cdot & 0.2 \end{pmatrix}$$

$$= 0.6 \times 0.2 + 0.2 \times 0.6 + 0.2 \times 0.2$$

$$= 0.28.$$

Note: we only need one element of the matrix P^2, so don't lose exam time by finding the whole matrix.

(c) Suppose that Purpose-flea is equally likely to start on any vertex at time 0. Find the probability distribution of X_1. $\Rightarrow$ $X_0 \sim \underline{\pi}^T = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

From this info, the distribution of X_0 is $\boldsymbol{\pi}^T = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. We need $X_1 \sim \boldsymbol{\pi}^T P$.

$$\boldsymbol{\pi}^T P = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

Thus $X_1 \sim \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$ and therefore X_1 is also equally likely to be 1, 2, or 3.

$X_0 \sim \underset{\sim}{\pi}^T = (1, 0, 0)$

(d) Suppose that Purpose-flea begins at vertex 1 at time 0. Find the probability distribution of X_2.

The distribution of X_0 *is now* $\boldsymbol{\pi}^T = (1, 0, 0)$. *We need* $X_2 \sim \boldsymbol{\pi}^T P^2$.

$$
\begin{aligned}
\boldsymbol{\pi}^T P^2 &= \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix} \\
&= \begin{pmatrix} 0.6 & 0.2 & 0.2 \end{pmatrix} \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix} \\
&= \begin{pmatrix} 0.44 & 0.28 & 0.28 \end{pmatrix}.
\end{aligned}
$$

Thus $\quad \mathbb{P}(X_2 = 1) = 0.44,\ \mathbb{P}(X_2 = 2) = 0.28,\ \mathbb{P}(X_2 = 3) = 0.28.$

Note that it is quickest to multiply the vector by the matrix first: we don't need to compute P^2 *in entirety.*

$X_0 \sim \underset{\sim}{\pi}^T = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$

(e) Suppose that Purpose-flea is equally likely to start on any vertex at time 0. Find the probability of obtaining the trajectory (3, 2, 1, 1, 3).

$$
\begin{aligned}
\mathbb{P}(3, 2, 1, 1, 3) &= \mathbb{P}(X_0 = 3) \times p_{32} \times p_{21} \times p_{11} \times p_{13} \qquad \textit{(Section 8.8)} \\
&= \tfrac{1}{3} \times 0.8 \times 0.4 \times 0.6 \times 0.2 \\
&= 0.0128.
\end{aligned}
$$

8.10 Class Structure

ie. the boxes

The state space of a Markov chain can be partitioned into a set of non-overlapping communicating classes.

States i and j are in the same communicating class if there is some way of getting from state i to state j, AND there is some way of getting from state j to state i. It needn't be possible to get between i and j in a ***single*** step, but it must be possible over some number of steps to travel between them both ways.

We write $i \leftrightarrow j$ if states i and j communicate.

Definition: Consider a Markov chain with state space S and transition matrix P, and consider states $i, j \in S$. Then **state *i* communicates with state *j*** if:

1. there exists some t such that $(P^t)_{ij} > 0$, AND — ie. can get $i \to j$ in some #steps with prob > 0
2. there exists some u such that $(P^u)_{ji} > 0$. — can get $j \to i$ in some #steps with prob > 0.

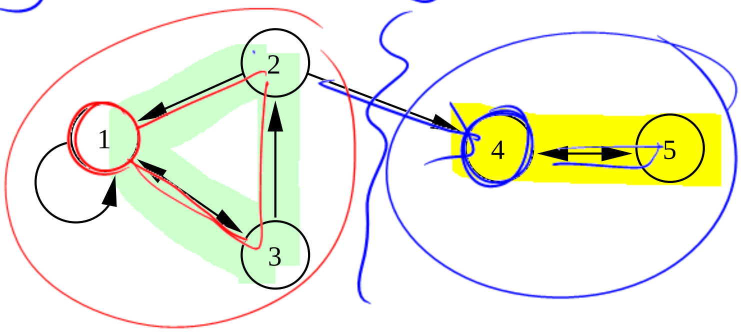
$t \geq 0, u \geq 0$

Mathematically, it is easy to show that the communicating relation $\leftrightarrow$ is an **equivalence relation,** which means that it partitions the ~~sample~~ state space S into non-overlapping equivalence classes.

Definition: States i and j are in the same **communicating class** if $i \leftrightarrow j$ ie. if each state is accessible from the other.

Every state is a member of exactly one communicating class.

Example: Find the communicating classes associated with the transition diagram shown.



Solution:

$\{1, 2, 3\}$ $\{4, 5\}$

State 2 leads to state 4, but 4 does NOT lead back to 2, so they are in different classes.

$-1 \xleftarrow{1/2} 0 \xrightarrow{1/2} 1$ LET $=\infty$

Definition: A communicating class of states is **closed** if it is not possible to LEAVE that class.

That is, the communicating class C is closed if $p_{ij} = 0$ whenever $i \in C$ and $j \notin C$.

Example: In the transition diagram above:

- Class $\{1, 2, 3\}$ is NOT closed (we CAN escape to class $\{4, 5\}$)
- Class $\{4, 5\}$ is CLOSED: it is not possible to escape.

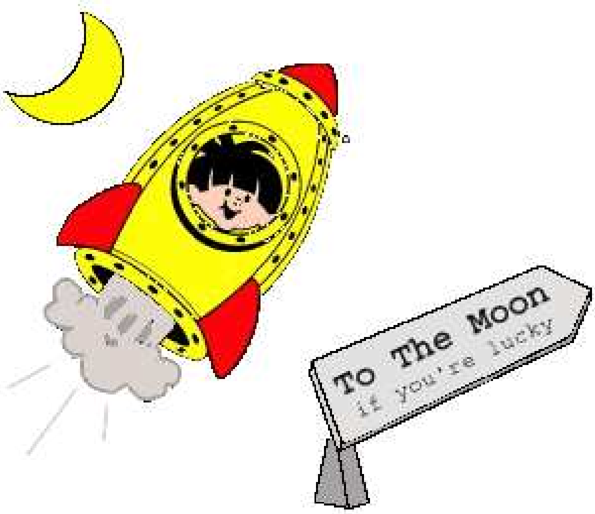
Definition: A state i is said to be **absorbing** if the set $\{i\}$ is a closed class.

$\rightarrow \boxed{i} \circlearrowleft 1$

Definition: A Markov chain or transition matrix P is said to be **irreducible** if $i \leftrightarrow j$ for all $i, j \in S$. That is, the chain is irreducible if the state space S is a single communicating class (can't be reduced or split up into smaller classes). Real point: we can get to everywhere starting from anywhere.

8.11 Hitting Probabilities

We have been calculating hitting probabilities for Markov chains since Chapter 2, using First-Step Analysis. The hitting probability describes the probability that the Markov chain will *ever* reach some state or set of states.



In this section we show how hitting probabilities can be written in a single vector. We also see a general formula for calculating the hitting probabilities. In general it is easier to continue using our own common sense, but occasionally the formula becomes more necessary.

$D \to A$, $D \to B$; W, L

$h_D = \mathbb{P}(\text{Venus wins} \mid \text{start@}D)$

$h_A = \mathbb{P}(\text{---} \mid \text{---} A)$

$h_B = \mathbb{P}(\text{---} \mid \text{---} B) \ldots$

$\mathbb{P}(\text{Venus Wins})$ = hitting probability for state W.

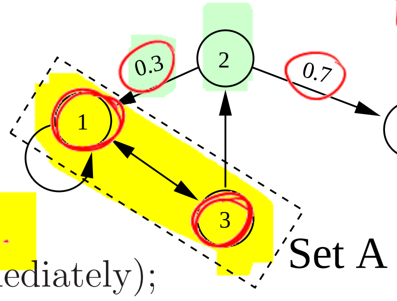
Vector of hitting probabilities

Let A be some subset of the state space S. (A need not be a communicating class: it can be any subset required, including the subset consisting of a single state: e.g. $A = \{4\}$.)

The **hitting probability** from state i to set A is the probability of ever reaching the set A, starting from initial state i. We write this probability as h_{iA}. Thus

$$h_{iA} = \mathbb{P}(X_t \in A \text{ for some } t \geq 0 \mid X_0 = i).$$

Example: Let set $A = \{1, 3\}$ as shown.



$h_{iA} = ?$ for $i = 1, \ldots, 5$

The hitting probability for set A is:

- 1 starting from states 1 or 3.
  (We are starting in set A, so we hit it immediately);
- 0 starting from states 4 or 5.
  (The set $\{4, 5\}$ is a closed class, so we can never escape out to set A);
- 0.3 starting from state 2.
  (We could hit A at the first step (probability 0.3), but otherwise we move to state 4 and get stuck in the closed class $\{4, 5\}$ (probability 0.7).)

We can summarize all the information from the example above in a vector of hitting probabilities:

$$\underset{\sim}{h}_A = \begin{pmatrix} h_{1A} \\ h_{2A} \\ h_{3A} \\ h_{4A} \\ h_{5A} \end{pmatrix} = \begin{pmatrix} 1 \\ 0.3 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

Vector format is just organisational for hitting probs.

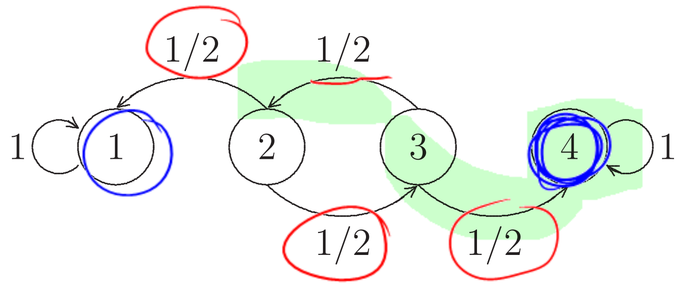
Note: When A is a closed class, the hitting probability h_{iA} is called the absorption probability.

In general, if there are N possible states, the vector of hitting probabilities is

$$\underset{\sim}{h}_A = \begin{pmatrix} h_{1A} \\ h_{2A} \\ \vdots \\ h_{NA} \end{pmatrix} = \begin{pmatrix} \mathbb{P}(\text{hit } A \mid \text{start at state } 1) \\ \mathbb{P}(\ \cdots \mid \cdots\cdots\ \ 2) \\ \vdots \\ \mathbb{P}(\ \cdots \mid \cdots\cdots\ \ N) \end{pmatrix}.$$

Example: finding the hitting probability vector using First-Step Analysis

Suppose $\{X_t : t \geq 0\}$ has the following transition diagram:



Find the vector of hitting probabilities for state 4. $A = \{4\}$.

Solution: Let $h_{i4} = \mathbb{P}(\text{hit state } 4 \mid \text{start at state } i)$ for $i = 1, 2, 3, 4$.

= human brain → Clearly

$$h_{14} = 0$$
$$h_{44} = 1$$

FSA:

$$h_{24} = \tfrac{1}{2} h_{14} + \tfrac{1}{2} h_{34} = \tfrac{1}{2} h_{34} \qquad ①$$

$$h_{34} = \tfrac{1}{2} * 1 + \tfrac{1}{2} h_{24} = \tfrac{1}{2} + \tfrac{1}{2} h_{24} \qquad ②$$

Subst ① in ② $\Rightarrow$ $h_{34} = \frac{1}{2} + \frac{1}{2}\left(\frac{1}{2} h_{34}\right)$

$\Rightarrow$ $h_{34} = \frac{2}{3}$ so also $h_{24} = \frac{1}{2} h_{34} = \frac{1}{3}$

So the vector of hitting probs is

$$\underset{\sim}{h}_4 = \left(0, \tfrac{1}{3}, \tfrac{2}{3}, 1\right).$$

Branching Proc $\boxed{0}$ $\boxed{1}$

extinction start $Z_0 = 1$

$\gamma = \mathbb{P}(\text{ever hit state } 0 \mid \text{start at state } 1)$

FSA for BP

$\Rightarrow \gamma = G(\gamma)$

Formula for hitting probabilities

In the previous example, we used our common sense to state that $h_{14} = 0$. While this is easy for a human brain, it is harder to explain a general rule that would describe this 'common sense' mathematically, or that could be used to write computer code that will work for all problems.

Although it is usually best to continue to use common sense when solving problems, this section provides a general formula that will *always* work to find a vector of hitting probabilities $\boldsymbol{h_A}$.

$h_1 = h_1$

FSA

Theorem 8.11: The vector of hitting probabilities $\boldsymbol{h_A} = (h_{iA} : i \in S)$ is the minimal non-negative solution to the following equations:

$$h_{iA} = \begin{cases} 1 & \text{for } i \in A, \\ \sum_{j \in S} p_{ij} h_{jA} & \text{for } i \notin A. \end{cases}$$

definitely reach A, starting from A.

FSA eqns

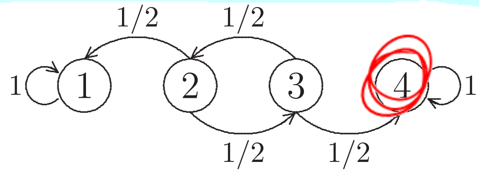
all 1st step probs

prob reach A after that 1st step.

The 'minimal non-negative solution' means that:

1. the values $\{h_{iA}\}$ collectively satisfy the equations above;
2. each value h_{iA} is ≥ 0 (non-negative);
3. given any other non-negative solution to the equations above, say $\{g_{iA}\}$ where $g_{iA} \geq 0$ for all i, then $h_{iA} \leq g_{iA}$ for all i (minimal solution).

Example: How would this formula be used to substitute for 'common sense' in the previous example?



The equations give:

$$h_{i4} = \begin{cases} 1 & \textit{if} \quad i = 4, \\ \sum_{j \in S} p_{ij} h_{j4} & \textit{if} \quad i \neq 4. \end{cases}$$

FSA

Thus,

$$\begin{aligned} h_{44} &= 1 \\ h_{14} &= h_{14} \quad \textit{unspecified! Could be anything!} \\ h_{24} &= \tfrac{1}{2} h_{14} + \tfrac{1}{2} h_{34} \\ h_{34} &= \tfrac{1}{2} h_{24} + \tfrac{1}{2} h_{44} = \tfrac{1}{2} h_{24} + \tfrac{1}{2} \end{aligned}$$

this is where we need the generality of the Thm.

Because h_{14} could be anything, we have to use the minimal non-negative value, which is $h_{14} = 0$.
(Need to check $h_{14} = 0$ does not force $h_{i4} < 0$ for any other i: OK.)

The other equations can then be solved to give the same answers as before. □

Proof of Theorem 8.11 (non-examinable):

Consider the equations $h_{iA} = \begin{cases} 1 & \text{for } i \in A, \\ \sum_{j \in S} p_{ij} h_{jA} & \text{for } i \notin A. \end{cases}$ $(\star)$

We need to show that:

(i) the hitting probabilities $\{h_{iA}\}$ collectively satisfy the equations $(\star)$;

(ii) if $\{g_{iA}\}$ is any other non-negative solution to $(\star)$, then the hitting probabilities $\{h_{iA}\}$ satisfy $h_{iA} \le g_{iA}$ for all i (minimal solution).

Proof of (i): Clearly, $h_{iA} = 1$ if $i \in A$ (as the chain hits A immediately).

Suppose that $i \notin A$. Then

FSA = Partition Thm.

$$\begin{aligned} h_{iA} &= \mathbb{P}(X_t \in A \text{ for some } t \ge 1 \mid X_0 = i) \\ &= \sum_{j \in S} \mathbb{P}(X_t \in A \text{ for some } t \ge 1 \mid X_1 = j)\mathbb{P}(X_1 = j \mid X_0 = i) \\ &\qquad\qquad \text{(Partition Rule)} \\ &= \sum_{j \in S} h_{jA}\, p_{ij} \qquad \text{(by definitions).} \end{aligned}$$

Thus the hitting probabilities $\{h_{iA}\}$ must satisfy the equations $(\star)$.

Proof of (ii): Let $h_{iA}^{(t)} = \mathbb{P}(\text{hit } A \text{ at or before time } t \mid X_0 = i)$.

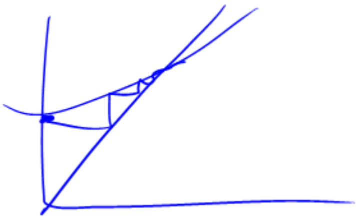
g_{iA} = "imposter" solution

We use mathematical induction to show that $h_{iA}^{(t)} \le g_{iA}$ for all t, and therefore $h_{iA} = \lim_{t\to\infty} h_{iA}^{(t)}$ must also be $\le g_{iA}$.

set up an induction.

$\underbrace{h_{iA}^{(0)}}_{\text{put time 0 here}} \le \underbrace{g_{iA}}_{\text{no time here}}$ for all i

t



<u>Time $t = 0$:</u> $\quad h_{iA}^{(0)} = \begin{cases} 1 & \text{if} \quad i \in A, \\ 0 & \text{if} \quad i \notin A. \end{cases}$

But because g_{iA} is non-negative and satisfies $(\star)$, $\begin{cases} g_{iA} = 1 & \text{if } i \in A, \\ g_{iA} \geq 0 & \text{for all } i. \end{cases}$

So $g_{iA} \geq h_{iA}^{(0)}$ for all i.

The inductive hypothesis is true for time $t = 0$.

<u>Time t:</u> Suppose the inductive hypothesis holds for time t, i.e.

$$h_{jA}^{(t)} \leq g_{jA} \quad \text{for all } j.$$

Consider

$$\begin{aligned} h_{iA}^{(t+1)} &= \mathbb{P}(\text{hit } A \text{ by time } t+1 \,|\, X_0 = i) \\ &= \sum_{j \in S} \mathbb{P}(\text{hit } A \text{ by time } t+1 \,|\, X_1 = j)\mathbb{P}(X_1 = j \,|\, X_0 = i) \\ & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(Partition Rule)} \\ &= \sum_{j \in S} h_{jA}^{(t)} p_{ij} \qquad \text{by definitions} \\ &\leq \sum_{j \in S} g_{jA}\, p_{ij} \qquad \text{by inductive hypothesis} \\ &= g_{iA} \qquad \text{because } \{g_{iA}\} \text{ satisfies } (\star). \end{aligned}$$

Thus $h_{iA}^{(t+1)} \leq g_{iA}$ for all i, so the inductive hypothesis is proved.

By the Continuity Theorem (Chapter 2), $h_{iA} = \lim_{t \to \infty} h_{iA}^{(t)}$.

So $h_{iA} \leq g_{iA}$ as required. $\square$

---

hitting probs: Ch 2 FSA

$\mathbb{E}$(reaching times): Ch 3 FSA

8.12 Expected hitting times

In the previous section we found the **probability** of hitting set A, starting at state i. Now we study **how long** it takes to get from i to A. As before, it is best to solve problems using first-step analysis and common sense. However, a general formula is also available.

Definition: Let A be a subset of the state space S. The **hitting time** of A is the random variable T_A, where

$$T_A = \min\{t \geq 0 : X_t \in A\}$$

$T_A = 0$ if we start in set A

T_A is the time taken before hitting set A for the first time.

The hitting time T_A can take values $0, 1, 2, \ldots,$ and ∞.

If the chain *never* hits set A, then $T_A = \infty$.

$T_A = \infty$ if we never get to set A.

Note: The hitting time is also called the **reaching time**. If A is a closed class, it is also called the absorption time.

Definition: The **mean hitting time** for A, starting from state i, is

$$m_{iA} = \mathbb{E}(T_A \mid X_0 = i).$$

Note: If there is any possibility that the chain *never* reaches A, starting from i,

ie. if the hitting probability $h_{iA} < 1$, then $\mathbb{E}(T_A \mid X_0 = i) = \infty$

also $\mathbb{P}(T_A = \infty) > 0$, T_A is defective.

Calculating the mean hitting times

Theorem 8.12: The vector of expected hitting times $\boldsymbol{m_A} = (m_{iA} : i \in S)$ is the minimal non-negative solution to the FSA equations:

$$m_{iA} = \begin{cases} 0 & \text{for } i \in A \\ 1 + \sum_{j \notin A} p_{ij} m_{jA} & \text{for } i \notin A. \end{cases}$$

Proof (sketch):

Consider the equations $m_{iA} = \begin{cases} 0 & \text{for} \quad i \in A, \\ 1 + \sum_{j \notin A} p_{ij} m_{jA} & \text{for} \quad i \notin A. \end{cases}$ $(\star)$.

We need to show that:

(i) the mean hitting times $\{m_{iA}\}$ collectively satisfy the equations $(\star)$; LOTE

(ii) if $\{u_{iA}\}$ is any other non-negative solution to $(\star)$, then the mean hitting times $\{m_{iA}\}$ satisfy $m_{iA} \leq u_{iA}$ for all i (minimal solution).

We will prove point (i) only. A proof of (ii) can be found online at: `http://www.statslab.cam.ac.uk/~james/Markov/` , Section 1.3.

Proof of (i): Clearly, $m_{iA} = 0$ if $i \in A$ (as the chain hits A immediately).

Suppose that $i \notin A$. Then

$$\begin{aligned} m_{iA} &= \mathbb{E}(T_A \,|\, X_0 = i) \\ &= 1 + \sum_{j \in S} \mathbb{E}(T_A \,|\, X_1 = j)\mathbb{P}(X_1 = j \,|\, X_0 = i) \end{aligned}$$

(conditional expectation: take 1 step to get to state j at time 1, then find $\mathbb{E}(T_A)$ from there)

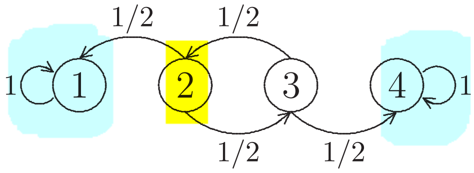
$$\begin{aligned} &= 1 + \sum_{j \in S} m_{jA}\, p_{ij} \qquad \text{(by definitions)} \\ &= 1 + \sum_{j \notin A} p_{ij}\, m_{jA}\,, \qquad \text{because } m_{jA} = 0 \text{ for } j \in A. \end{aligned}$$

Thus the mean hitting times $\{m_{iA}\}$ must satisfy the equations $(\star)$.

$\square$

Example: Let $\{X_t : t \geq 0\}$ have the same transition diagram as before:

Starting from state 2, find the expected time to absorption.

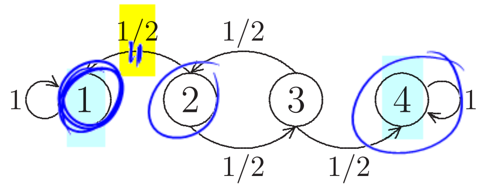


Solution: Notation: let

$m_{iA} = \mathbb{E}$ (#steps to reach set $A = \{1, 4\}$ | start at state i) for $i = 1, 2, 3, 4$.

= #arrows crossed

We require m_{2A}.



FSA: $m_{1A} = 0$, $m_{4A} = 0$ } because $1 \in A$ and $4 \in A$

Also $m_{2A} = 1 + \frac{1}{2} m_{3A}$ ①

$m_{3A} = 1 + \frac{1}{2} m_{2A}$ ②

Subst ② into ①:

$$m_{2A} = 1 + \frac{1}{2}\left\{1 + \frac{1}{2} m_{2A}\right\}$$

$$\Rightarrow \frac{3}{4} m_{2A} = \frac{3}{2}$$

$$\Rightarrow \underline{\underline{m_{2A} = 2.}}$$

Also, $\underline{m_{3A} = 1 + \frac{1}{2} m_{2A} = 1 + \frac{2}{2} = 2.}$

Required answer is $\underline{\underline{m_{2A} = 2}}$

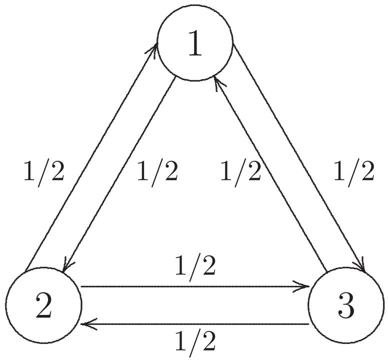
$\underline{\underline{\mathbb{E}(T \mid \text{starting at } 2) = 2 \text{ steps.}}}$

$\underset{\sim}{h} = (h_1, h_2, \dots, h_N)$ NEVER ADD.

$\underset{\sim}{m} = (m_1, m_2, \dots, m_N)$ NEVER ADD like $m_1 + m_2 + \dots + m_N$, will never be the right answer.

Example: Glee-flea hops around on a triangle. At each step he moves to one of the other two vertices at random. What is the expected time taken for Glee-flea to get from vertex 1 to vertex 2?

Solution:



transition matrix, $P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$.

We wish to find m_{12}.

Now $$m_{i2} = \begin{cases} 0 & \textit{if} \quad i = 2, \\ 1 + \sum_{j \neq 2} p_{ij} m_{j2} & \textit{if} \quad i \neq 2. \end{cases}$$

Thus

$$\begin{aligned} m_{22} &= 0 \\ m_{12} &= 1 + \tfrac{1}{2} m_{22} + \tfrac{1}{2} m_{32} \;=\; 1 + \tfrac{1}{2} m_{32}. \\ m_{32} &= 1 + \tfrac{1}{2} m_{22} + \tfrac{1}{2} m_{12} \\ &= 1 + \tfrac{1}{2} m_{12} \\ &= 1 + \tfrac{1}{2}\left(1 + \tfrac{1}{2} m_{32}\right) \\ \Rightarrow \quad m_{32} &= 2. \end{aligned}$$

Thus $m_{12} = 1 + \frac{1}{2} m_{32} = 2$ *steps.*