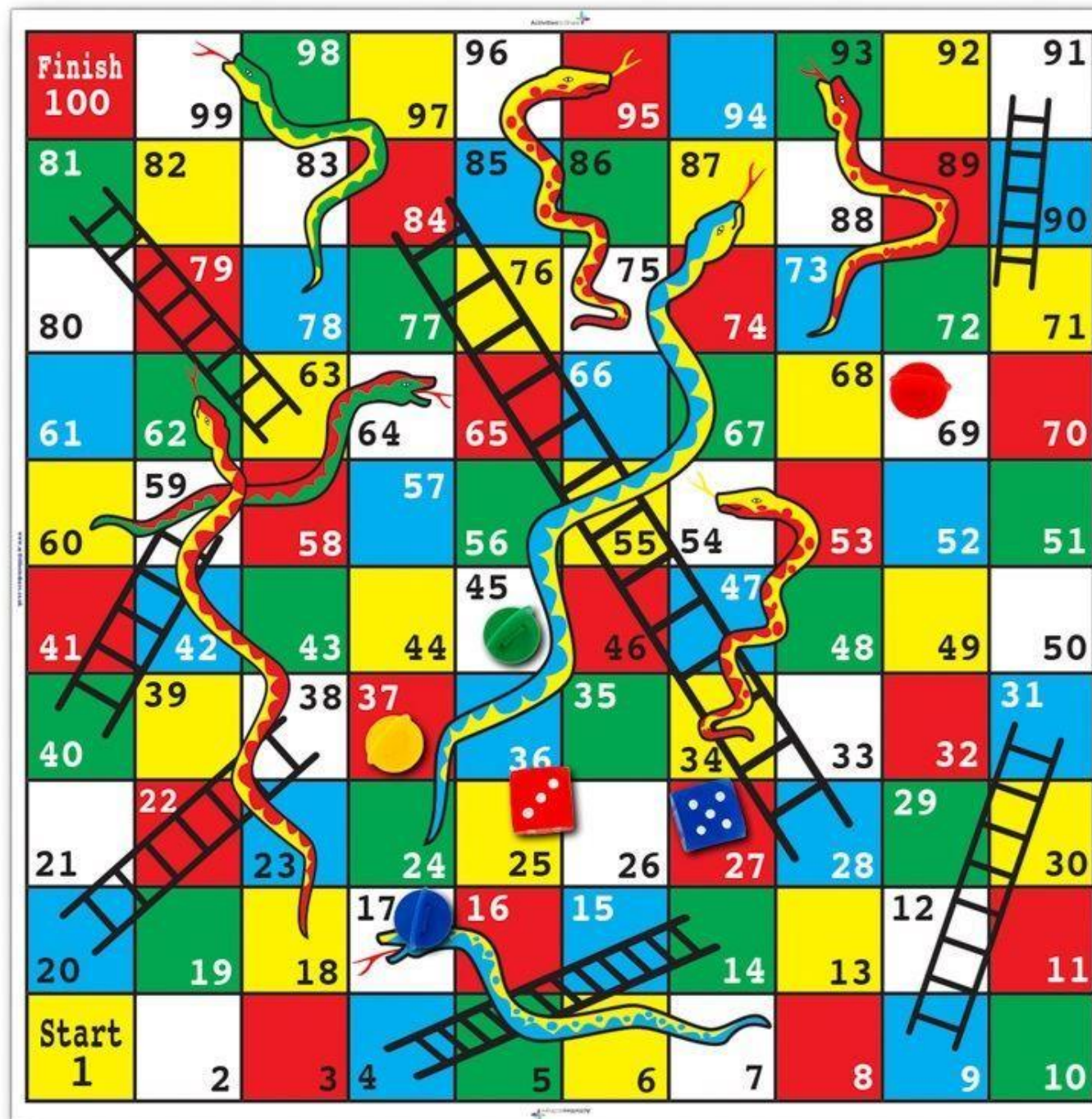


Workshop 16: Markov Chains

FIT 3139 Computational Modelling and Simulation



<https://www.activitiestoshare.co.uk/snakes-and-ladders-floor-mat>

DISCRETE

Markov chain

- Set of states $S = \{s_1, s_2, s_3, \dots, s_r\}$
- Probability of going from s_i to s_j is given by p_{ij} (transition probability)

$$\sum_j p_{ij} = 1$$

- Transition matrix, $A = \{p_{ij}\}$

Conditional probability distribution of future states depends only on the present state

“Memoryless” process.

Conditional probability distribution of future states depends only on the present state

“Markovian”

Early 20th century...



A. A. Markov (1886).

https://en.wikipedia.org/wiki/Andrey_Markov

*The Land of Oz is blessed by many things,
but not by good weather.*

*They never have two nice days in a row.
If they have a nice day, they are just as likely
to have snow as rain the next day.*

*If they have snow or rain, they have an even
chance of having the same the next day.*

*If there is change from snow or rain, only half
of the time is this a change to a nice day.*

Transition Matrix

$$\begin{array}{c} \text{R} \\ \text{N} \\ \text{S} \end{array} \begin{array}{ccc} \text{R} & \text{N} & \text{S} \\ \left(\begin{array}{ccc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{array} \right) \end{array}$$

(stochastic matrix, because rows are probability distributions)

Simulation

- At every time step there is a discrete random variable that needs to be sampled.
- While stopping criteria is not met, decide what is the next state.

Simulation

If transitions are explicit....

Each state has an associated distribution, given by the corresponding row of the transition matrix.

$$X = \begin{cases} x_0 & \text{if } U < p_0 \\ x_1 & \text{if } p_0 \leq U < p_0 + p_1 \\ \vdots & \\ x_j & \text{if } \sum_{i=0}^{j-1} p_i \leq U < \sum_{i=0}^j p_i \\ \vdots & \end{cases}$$

with U , uniformly distributed on $(0,1)$

$$\begin{array}{c}
 \\
 \\
 \begin{array}{c} \mathbf{R} \\ \mathbf{N} \\ \mathbf{S} \end{array}
 \end{array}
 \begin{array}{ccc}
 \mathbf{R} & \mathbf{N} & \mathbf{S} \\
 \left(\begin{array}{ccc}
 \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\
 \frac{1}{2} & \mathbf{0} & \frac{1}{2} \\
 \frac{1}{4} & \frac{1}{4} & \frac{1}{2}
 \end{array} \right)
 \end{array}$$

If it is raining today, what is the probability that it will be snowing in two days?

	R	N	S
R	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
N	$\frac{1}{2}$	0	$\frac{1}{2}$
S	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

today tomorrow day after

$$\begin{aligned}
 & R \longrightarrow \boxed{} \longrightarrow S \\
 & R \longrightarrow R \longrightarrow S \quad \frac{1}{2} \cdot \frac{1}{4} \\
 & R \longrightarrow N \longrightarrow S \quad \frac{1}{4} \cdot \frac{1}{2} \\
 & R \longrightarrow S \longrightarrow S \quad \frac{1}{4} \cdot \frac{1}{2}
 \end{aligned}
 + + = \frac{3}{8}$$

If it is raining today, what is the probability that it will be snowing in two days?

	R	N	S
R	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
N	$\frac{1}{2}$	0	$\frac{1}{2}$
S	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

$$= P^2 = \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix}$$

If it is raining today, what is the probability that it will be snowing in two days?

$$\begin{array}{c}
 \text{R} \quad \text{N} \quad \text{S} \\
 \begin{array}{c} \text{R} \\ \text{N} \\ \text{S} \end{array}
 \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}
 \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}
 = \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix} = P^2
 \end{array}$$

If it is nice today, what is the probability that it is raining in two days?

If it is raining today, what is the probability that it will be snowing in two days?

$$\begin{array}{c}
 \\
 \\
 \text{R} \\
 \text{N} \\
 \text{S}
 \end{array}
 \begin{array}{c}
 \text{R} \quad \text{N} \quad \text{S} \\
 \left(\begin{array}{ccc}
 \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\
 \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\
 \frac{3}{8} & \frac{3}{16} & \frac{7}{16}
 \end{array} \right)
 \end{array}
 = P^2$$

Weather in two days...

given weather today
(pick a row)

$$P^2 \cdot P = \begin{pmatrix} \frac{7}{16} & \frac{3}{16} & \frac{3}{8} \\ \frac{3}{8} & \frac{1}{4} & \frac{3}{8} \\ \frac{3}{8} & \frac{3}{16} & \frac{7}{16} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{11}{64} & \frac{5}{64} & \frac{1}{8} \\ \frac{3}{16} & \frac{1}{8} & \frac{3}{16} \\ \frac{1}{8} & \frac{5}{64} & \frac{11}{64} \end{pmatrix} \begin{matrix} R \\ N \\ S \end{matrix} = P^3$$

Weather in three days...

For a Markov chain given by transition matrix \mathbf{P} the probability to go from state s_i to state s_j in n steps, is given by the *ij-th entry of \mathbf{P}^n*
Entries of this matrix are denoted as: $p_{ij}^{(n)}$

Long-term behaviour:

$$\mathbf{P}^* = \lim_{k \rightarrow \infty} \mathbf{P}^k$$

$$\begin{array}{c} \text{R} \\ \text{N} \\ \text{S} \end{array} \begin{array}{ccc} \text{R} & \text{N} & \text{S} \\ \left(\begin{array}{ccc} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{array} \right) \end{array} = P$$

$$\begin{array}{c} \text{R} \\ \text{N} \\ \text{S} \end{array} \begin{array}{ccc} \text{R} & \text{N} & \text{S} \\ \left(\begin{array}{ccc} \frac{205}{512} & \frac{205}{1024} & \frac{409}{1024} \\ \frac{205}{512} & \frac{51}{256} & \frac{205}{512} \\ \frac{409}{1024} & \frac{205}{1024} & \frac{205}{512} \end{array} \right) \end{array} = P^5$$

$$(1,0,0) \cdot P^5 = \left(\frac{205}{512} \quad \frac{205}{1024} \quad \frac{409}{1024} \right)$$

$$(0,1,0) \cdot P^5 = \left(\frac{205}{512} \quad \frac{51}{256} \quad \frac{205}{512} \right)$$

$$(0,0,1) \cdot P^5 = \left(\frac{409}{1024} \quad \frac{205}{1024} \quad \frac{205}{512} \right)$$

Initial
state

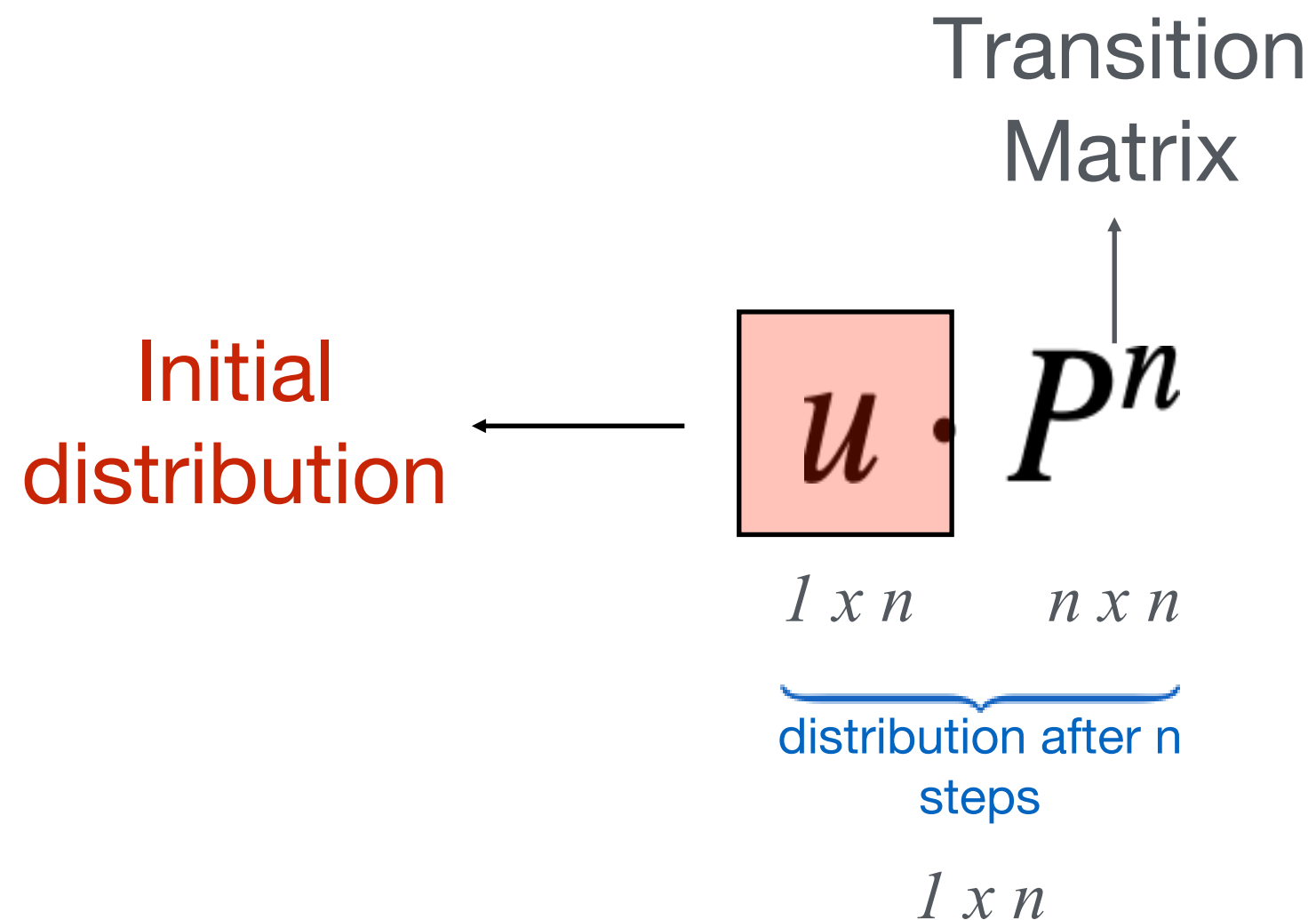
$$\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3} \right)$$

$\cdot P^5 =$

$$\left(\frac{205}{512} \quad \frac{205}{1024} \quad \frac{409}{1024} \right)$$

Probability
of being in a state
after 5 steps
...given initial state

Initial distribution



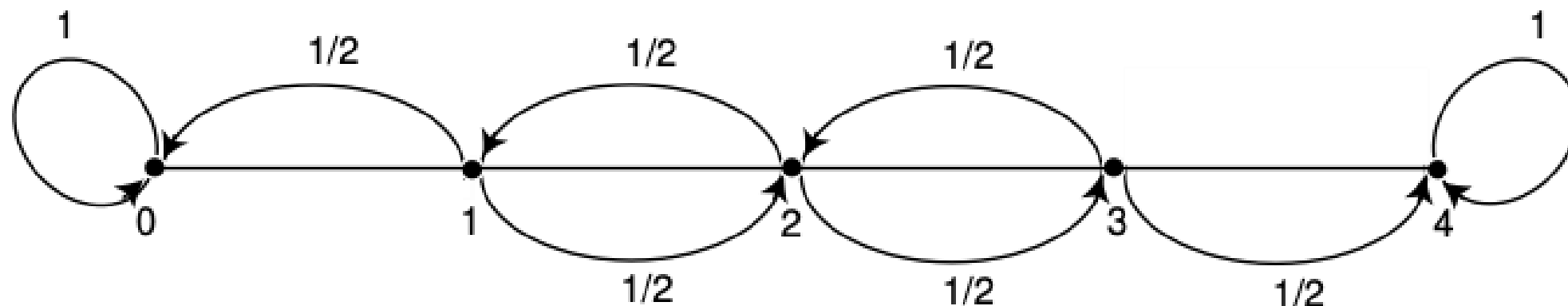
Drunkard's walk

A man walks across a four block stretch of Sydney road.

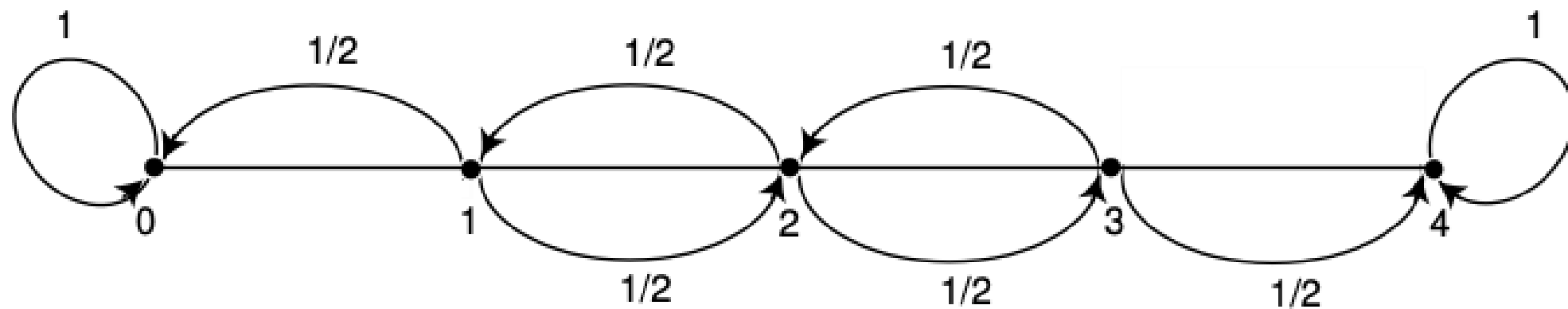
If he's at corner 1, 2, or 3 he walks to the left or right with equal probability.

He continues until he reaches corner 4, a hip bar... or corner 0, which is home.

If he reaches either home, or the bar he stays there.



$$\begin{array}{c}
 0 \\
 1 \\
 2 \\
 3 \\
 4
 \end{array}
 \begin{array}{c}
 0 \quad 1 \quad 2 \quad 3 \quad 4 \\
 \left(\begin{array}{ccccc}
 1 & 0 & 0 & 0 & 0 \\
 0.5 & 0 & 0.5 & 0 & 0 \\
 0 & 0.5 & 0 & 0.5 & 0 \\
 0 & 0 & 0.5 & 0 & 0.5 \\
 0 & 0 & 0 & 0 & 1
 \end{array} \right)
 \end{array}$$



What happens in the long term...

Absorbing vs Ergodic

Absorbing chains.

- A state is absorbing if it is impossible to leave (i.e., $p_{ii} = 1$).
- A chain is absorbing, if it has at least one absorbing state, and if from every state it is possible to go to an absorbing state (possibly in several steps)
- A state that is non-absorbing is also called transient.

Which ones are absorbing?

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0.4 & 0.6 \\ 0 & 0.2 & 0.8 \end{pmatrix}$$

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.1 & 0.7 \end{pmatrix}$$

Absorbing

Transient

	0	1	2	3	4
0	1	0	0	0	0
1	0.5	0	0.5	0	0
2	0	0.5	0	0.5	0
3	0	0	0.5	0	0.5
4	0	0	0	0	1

	1	2	3	0	4
1	0	0.5	0	0.5	0
2	0.5	0	0.5	0	0
3	0	0.5	0	0	0.5
0	0	0	0	1	0
4	0	0	0	0	1

Canonical form

$$\mathbf{P} = \begin{array}{c} \text{TR.} \\ \text{ABS.} \end{array} \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & \mathbf{I} \end{array} \right)$$

Re-number states so that **transient states come first**

Probability of being in a
transient state after n
steps

$$\mathbf{P}^n = \begin{array}{cc} & \begin{array}{cc} \text{TR.} & \text{ABS.} \end{array} \\ \begin{array}{c} \text{TR.} \\ \text{ABS.} \end{array} & \left(\begin{array}{c|c} \mathbf{Q}^n & * \\ \hline \mathbf{0} & \mathbf{I} \end{array} \right) \end{array}$$

In an absorbing Markov chain the probability that the process will be absorbed is 1

$$\lim_{n \rightarrow \infty} \mathbf{Q}^n = \mathbf{0}$$

```

1  # Define M
2  M = np.array([[0, 0.5, 0, 0.5, 0 ],
3                [0.5, 0, 0.5, 0, 0 ],
4                [0, 0.5, 0, 0, 0.5],
5                [0, 0, 0, 1, 0 ],
6                [0, 0, 0, 0, 1 ]])
7  print(M)

```

```

[[0.  0.5 0.  0.5 0. ]
 [0.5 0.  0.5 0.  0. ]
 [0.  0.5 0.  0.  0.5]
 [0.  0.  0.  1.  0. ]
 [0.  0.  0.  0.  1. ]]

```

```

1  # M to some power
2  print(np.linalg.matrix_power(M, 8))

```

```

[[0.03125 0. 0.03125 0.71875 0.21875]
 [0. 0.0625 0. 0.46875 0.46875]
 [0.03125 0. 0.03125 0.21875 0.71875]
 [0. 0. 0. 1. 0. ]
 [0. 0. 0. 0. 1. ]]

```

```

1  # Q to some power
2  print(np.linalg.matrix_power(M[0:3, 0:3], 8))

```

```

[[0.03125 0. 0.03125]
 [0. 0.0625 0. ]
 [0.03125 0. 0.03125]]

```

In an absorbing Markov chain the probability that the process will be absorbed is 1

$$\lim_{n \rightarrow \infty} \mathbf{Q}^n = \mathbf{0}$$

- Let m_j be the minimum number of steps required to reach an absorbing state, starting in transient state j
- Let p_j be the probability that starting in j , the process will not reach an absorbing state in m_j steps. Thus, $p_j < 1$.
- Let m be the largest m_j , and p be the largest p_j
- The probability of not being absorbed in m steps is at most p
- The probability of not being absorbed in mk steps is at most p^k
- The probability of not being absorbed goes to 0 as k increases.



```

1 # Define M
2 M = np.array([[0, 0.5, 0, 0.5, 0 ],
3               [0.5, 0, 0.5, 0, 0 ],
4               [0, 0.5, 0, 0, 0.5],
5               [0, 0, 0, 1, 0 ],
6               [0, 0, 0, 0, 1 ]])
7 Q = M[0:3, 0:3]
8 print(Q)

```

```

[[0.  0.5 0. ]
 [0.5 0.  0.5]
 [0.  0.5 0. ]]

```

```

1 print(np.linalg.matrix_power(Q, 15))

```

```

[[0.          0.00390625 0.          ]
 [0.00390625 0.          0.00390625]
 [0.          0.00390625 0.          ]]

```

```

1 print(np.linalg.matrix_power(Q, 150))

```

```

[[1.32348898e-23 0.00000000e+00 1.32348898e-23]
 [0.00000000e+00 2.64697796e-23 0.00000000e+00]
 [1.32348898e-23 0.00000000e+00 1.32348898e-23]]

```

	1	2	3	0	4
1	0	0.5	0	0.5	0
2	0.5	0	0.5	0	0
3	0	0.5	0	0	0.5
0	0	0	0	1	0
4	0	0	0	0	1

Remaining interesting questions?

- How many times do I visit a particular transient state before absorption?
- How long until absorption?
- Which of the absorbing states am I more likely to end up in?

Fundamental matrix

$$N = (I - Q)^{-1} = I + Q^1 + Q^2 + Q^3 + \dots$$

↓
fundamental
matrix

The entry ij of N , is the expected number of times the chain visits state j , starting in state i ... before absorption.

(you will prove this in the tutorial next week)

The entry ij of N , is the expected number of times the chain visits state j , starting in state i ... before absorption.

Absorption time

$$t = N \cdot \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}$$

vector with as many
components as transient
states

t_i is the **expected time before absorption** from transient state i

(intuition: sum the times I visit all transient states before absorption)


```

1  # Define M
2  M = np.array([[0, 0.5, 0, 0.5, 0 ],
3                [0.5, 0, 0.5, 0, 0 ],
4                [0, 0.5, 0, 0, 0.5],
5                [0, 0, 0, 1, 0 ],
6                [0, 0, 0, 0, 1 ]])
7  Q = M[0:3, 0:3]
8  print(Q)

```

```

[[0.  0.5  0. ]
 [0.5  0.  0.5]
 [0.  0.5  0. ]]

```

```

1  # Find fundamental matrix
2  N = np.linalg.inv((np.eye(3) - Q))
3  # Calculate Absorption times
4  t = N.dot(np.array([1, 1, 1]))
5  print(t)

```

```

[3.  4.  3.]

```

Absorption probability

$$\mathbf{P} = \begin{array}{c} \text{TR.} \\ \text{ABS.} \end{array} \left(\begin{array}{c|c} \mathbf{Q} & \mathbf{R} \\ \hline \mathbf{0} & \mathbf{I} \end{array} \right)$$

Remember R

Absorption probability

Let b_{ij} be the probability that an absorbing chain will be absorbed in the absorbing state j if it starts in the transient state i .

Let B be the matrix with entries b_{ij} . Then B is a t-by-r matrix and:

$$B = NR$$

$$B_{ij} = \sum_n \sum_k q_{ik}^{(n)} r_{kj} \quad q_{ik}^{(n)} = \begin{matrix} i & k \text{ entry} \\ \text{of} & \end{matrix} Q^n$$

$$= \sum_k \sum_n q_{ik}^{(n)} r_{kj} \quad (\text{regroup})$$

$$= \sum_k n_{ik} r_{kj}$$

$$= [NR]_{ij}$$

```

1  # Define M
2  M = np.array([[0, 0.5, 0, 0.5, 0 ],
3                [0.5, 0, 0.5, 0, 0 ],
4                [0, 0.5, 0, 0, 0.5],
5                [0, 0, 0, 1, 0 ],
6                [0, 0, 0, 0, 1 ]])
7  Q = M[0:3, 0:3]
8  print(Q)
9  R = M[0:3, 3:]
10 print(R)

```

```

[[0.  0.5 0. ]
 [0.5 0.  0.5]
 [0.  0.5 0. ]]
[[0.5 0. ]
 [0.  0. ]
 [0.  0.5]]

```

```

1  # Find fundamental matrix
2  N = np.linalg.inv((np.eye(3) - Q))
3  # Calculate Absorption probabilities
4  B = N.dot(R)
5  print(B)

```

```

[[0.75 0.25]
 [0.5  0.5 ]
 [0.25 0.75]]

```

Absorbing vs Ergodic

Ergodic chains.

- A Markov chain is called ergodic if it is possible to go from every state to every state (not necessarily in one step).
- Ergodic chains have a unique long-run equilibrium that does not depend on initial conditions.
- Intuitively, given enough time, the process *forgets* about the initial configuration.

Long-term behaviour:

$$\mathbf{P}^* = \lim_{k \rightarrow \infty} \mathbf{P}^k$$

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \mathbf{0} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix}$$

$$\begin{pmatrix} 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \\ 0.4 & 0.2 & 0.4 \end{pmatrix}$$

$$\mathbf{P}^* = \{p_{ij}^*\} \quad \text{if history does not matter} \quad p_{ij}^* = p_{i'j}^*$$

i.e. all rows of \mathbf{P}^* must be the same

$$\mathbf{P}\mathbf{P}^* = \mathbf{P} \lim_{k \rightarrow \infty} \mathbf{P}^k = \lim_{k \rightarrow \infty} \mathbf{P}^{k+1} = \mathbf{P}^*$$

$\mathbf{u} = (u_1, u_2, \dots, u_n)$ is the common row of \mathbf{P}^*

$$u_j = \lim_{m \rightarrow \infty} p_{ij}^{(m)}$$

stationary distribution


```
1  # Define M
2  M = np.array([[0.5, 0.25, 0.25],
3                [0.5, 0, 0.5],
4                [0.25, 0.25, 0.5]])
```

```
1  print(np.linalg.matrix_power(M, 5))
```

```
[[0.40039062 0.20019531 0.39941406]
 [0.40039062 0.19921875 0.40039062]
 [0.39941406 0.20019531 0.40039062]]
```

```
1  print(np.linalg.matrix_power(M, 20))
```

```
[[0.4 0.2 0.4]
 [0.4 0.2 0.4]
 [0.4 0.2 0.4]]
```

Problem

Consider: $M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$

Is M ergodic?

Yes, it is possible to go from every state to every other state

What happens when we raise M to some power?

$$M^{2k} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad M^{2k+1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

The rows don't converge –
what's the stationary distribution?

This method only works for Regular Markov Chains

A Markov Chain is called a *regular* chain if some power of the transition matrix has only positive elements.

$$M = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \text{ is not regular}$$

So how do we find the stationary distribution for M ?

$$\mathbf{u} = \mathbf{uP}$$

\mathbf{u} is the left eigenvector for eigenvalue 1 of \mathbf{P}

stationary distribution

What problem do we have with \mathbf{u} when calculated this way?

```
1 # Define M
2 M = np.array([[0.5, 0.25, 0.25],
3               [0.5, 0, 0.5],
4               [0.25, 0.25, 0.5]])
```

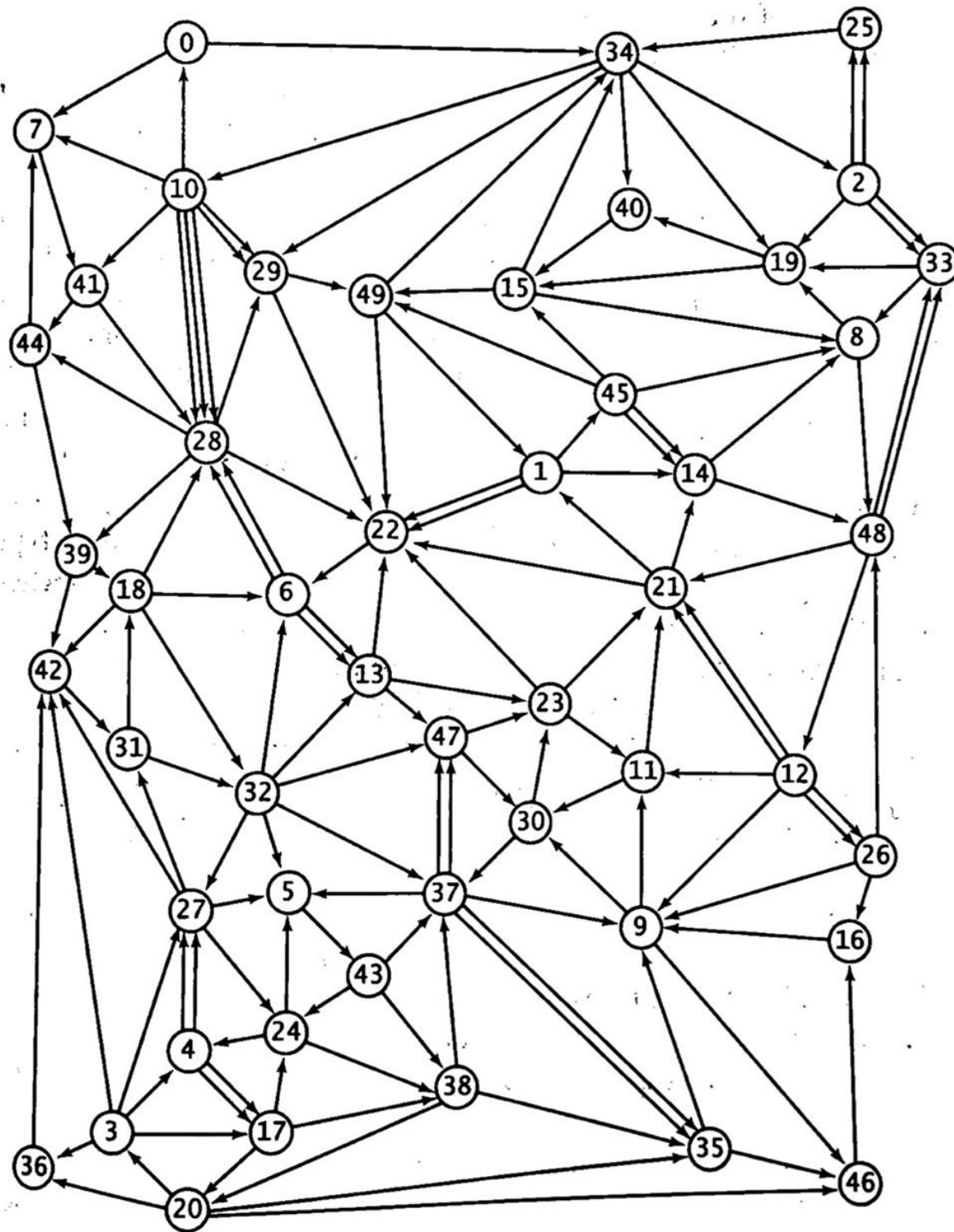
```
1 lambda_, v = np.linalg.eig(M.T)
2 print(lambda_)
```

```
[ 1.    0.25 -0.25]
```

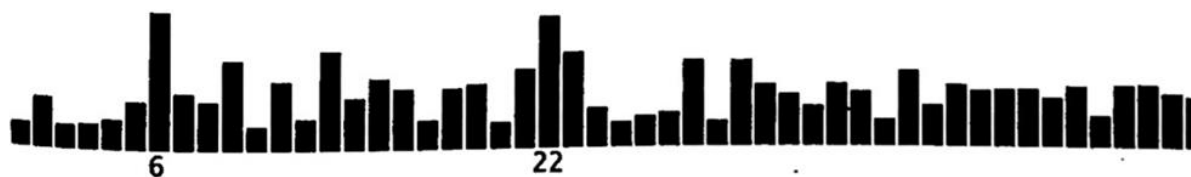
```
1 print(v[:,0]/sum(v[:,0]))
```

```
[0.4 0.2 0.4]
```

The ergodic chain of a random web surfer



0	.002
1	.017
2	.009
3	.003
4	.006
5	.016
6	.066
7	.021
8	.017
9	.040
10	.002
11	.028
12	.006
13	.045
14	.018
15	.026
16	.023
17	.005
18	.023
19	.026
20	.004
21	.034
22	.063
23	.043
24	.011
25	.005
26	.006
27	.008
28	.037
29	.003
30	.037
31	.023
32	.018
33	.013
34	.024
35	.019
36	.003
37	.031
38	.012
39	.023
40	.017
41	.021
42	.021
43	.016
44	.023
45	.006
46	.023
47	.024
48	.019
49	.016



Page ranks with histogram for a larger example