EL2805, REINFORCEMENT LEARNING

Robert Mattila

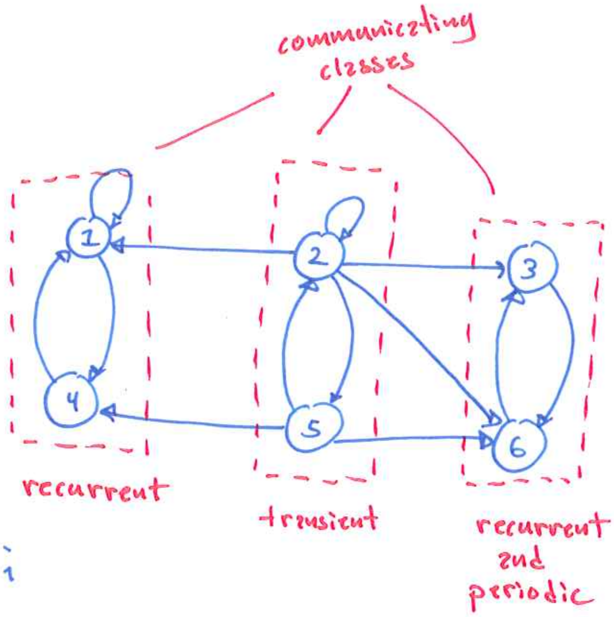
EXERCISE 1: MARKOV CHAINS

Today's central concepts:

- Markov chains
  - State space
  - Transition probabilities (matrix for discrete state space)

    $$P_{ij} = \Pr\{\underline{X}_{k+1} = j \mid \underline{X}_k = i\}$$
  - The Markov property
- transient state:

  from at least one state which may be eventually reached from i, the system can never return to i



- recurrent state:

  from every state which may be reached eventually from i, the system can eventually return to i
- periodic state:

  a recurrent state for which $P_{ii}^{(k)}$ may be non-zero only for $k = d, 2d, 3d, \ldots$ where $d > 1$ is integer
- accessibility:

  $i \rightarrow j$ means that it is possible to go from i to j
- communicating:

  $i \leftrightarrow j$ means that $i \rightarrow j$ and $j \rightarrow i$.

  Note: $i \leftrightarrow i$
- communicating class:

  a maximal set of states C such that for all $i, j \in C$, we have $i \leftrightarrow j$.
- irreducible chain:

  there is only one communicating class (the whole state space)

Ex 1.1 Fair die tossed. X_n = maximum of the first n throws.

a, Show that $(X_n)_{n\geq 0}$ is a Markov chain

b, Compute P

c, Specify the classes of the chain

Solution:

Intuition: How the maximum was achieved is irrelevant (for the future), everything is contained in the maximum. Should be a M.C.

a, Let the outcomes of the die be denoted $\{Z_n\}$. These are i.i.d. with a uniform distribution over $\{1,..,6\}$.

Our state is

$$X_n = \max\{Z_1, ..., Z_n\}.$$

Note: we define $X_0 = 1$, so this is actually $\max\{Z_1, 1\} = Z_1$

We verify the Markov property:

$$\Pr\{X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1}, ..., X_1 = i_1\} =$$

$$\Pr\{\max\{Z_1, ..., Z_n, Z_{n+1}\} = j \mid \max\{Z_1, .., Z_n\} = i_n, ..., \max\{Z_1\} = i_1\} =$$

$$= \left\{ \begin{array}{l} \text{Note that:} \\ \max\{Z_1, ..., Z_n, Z_{n+1}\} = \max\{\max\{Z_1, .., Z_n\}, Z_{n+1}\} \\ \left(\max\{3, 1, 5, 7\} = \max\{\max\{3, 1, 5\}, 7\} = \max\{5, 7\} = 7\right) \end{array} \right\} =$$

$$\Pr\{\max\{\underbrace{\max\{Z_1, .., Z_n\}}_{= X_n}, Z_{n+1}\} = i_{n+1} \mid \underbrace{\max\{Z_1, .., Z_n\}}_{= X_n} = i_n, ..., \max\{Z_1\} = i_1\}$$

$$\Pr\{\max\{X_n, Z_{n+1}\} = j \mid X_n = i_n, \ldots, X_1 = i_1\} =$$

The next throw (Z_{n+1}) is by assumption an independent r.v., so with respect to it, we can remove everything we condition on without loss of information.
However, we also have the r.v. X_n there, so we have to keep that conditioning.

$=$

$$\Pr\{\underbrace{\max\{X_n, Z_{n+1}\}}_{= X_{n+1}} = j \mid X_n = i_n\} =$$

$$\Pr\{X_{n+1} = j \mid X_n = i_n\}.$$

In summary,

$$\Pr\{X_{n+1} = j \mid X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1\} =$$
$$\Pr\{X_{n+1} = j \mid X_n = i_n\}$$

which verifies the Markov property.

b, what is the statespace of X_n? $\{1,2,3,4,5,6\}$

The transition matrix

$$P = \begin{array}{c|cccccc} & & & & \text{to } j & & \\ & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline \text{from } i \;\; 1 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 2 & 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 3 & & & & & & \\ 4 & & & & & & \\ 5 & & & & & & \\ 6 & & & & & & \end{array}$$

Recall:

$$P_{ij} = \Pr\{X_{n+1}=j \mid X_n = i\} = \Pr\{\max\{X_n, Z_{n+1}\} = j \mid X_n = i\}$$
$$= \Pr\{\max\{i, Z_{n+1}\} = j\}$$

<u>First row:</u>

First element $i=1$ $j=1$:

$$\Pr\{\max\{1, Z_{n+1}\} = 1\} = \Pr\{Z_{n+1} = 1\} = 1/6$$

Second element $i=1$ $j=2$:

$$\Pr\{\max\{1, Z_{n+1}\} = 2\} = \Pr\{Z_{n+1} = 2\} = 1/6$$

Third element $i=1$ $j=3$:

$$\Pr\{\max\{1, Z_{n+1}\} = 3\} = \Pr\{Z_{n+1} = 3\} = 1/6$$

etc.

<u>Second row:</u>

First element $i=2$ $j=1$:

$$\Pr\{\max\{2, Z_{n+1}\} = 1\} = 0$$

Second element $i=2$ $j=2$:

$$\Pr\{\max\{2, Z_{n+1}\} = 2\} = \Pr\{Z_{n+1} \le 2\} = \Pr\{Z_{n+1} = 1\} + \Pr\{Z_{n+1} = 2\} = \frac{2}{6}$$

Third element $i=2$ $j=3$:

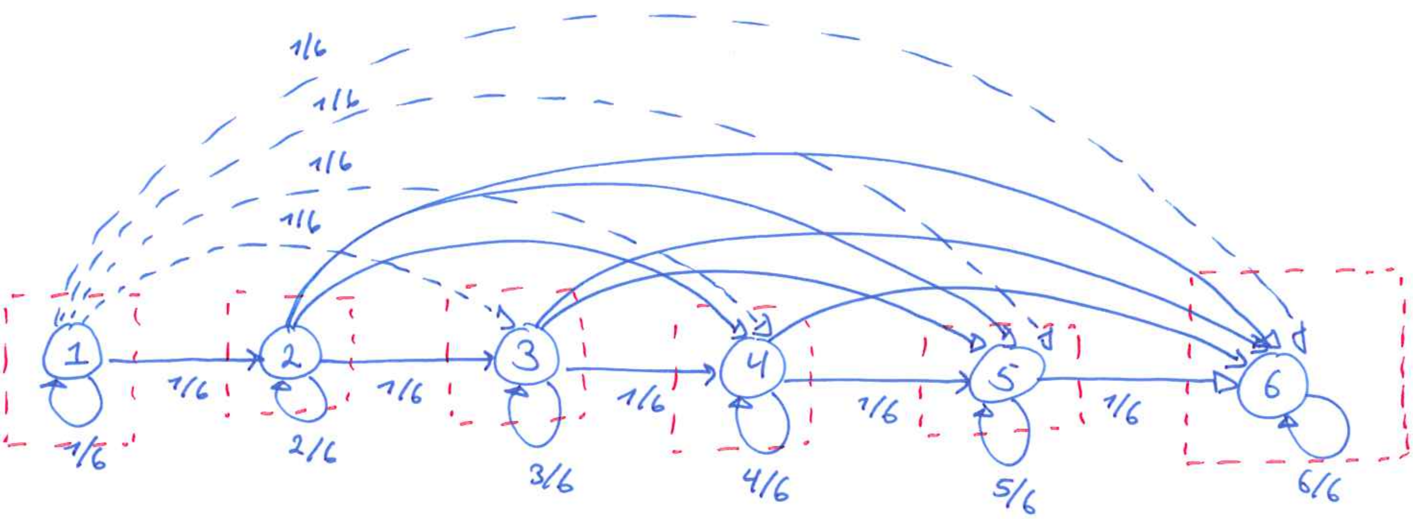
$$\Pr\{\max\{2, Z_{n+1}\} = 3\} = \Pr\{Z_{n+1} = 3\} = 1/6$$

That was the formal way, the intuition/induction should be clear:

$$P = \begin{array}{c|cccccc} & 1 & 2 & 3 & 4 & 5 & 6 \\ \hline 1 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 2 & 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 3 & 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 4 & 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 5 & 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 6 & 0 & 0 & 0 & 0 & 0 & 6/6 \end{array}$$

"If the old maximum was 4, then all outcomes below 5 will make us stay (4/6). The other will make us move higher."

c) To identify classes, it's easiest to draw the chain:



Communicating classes:
{1}, {2}, {3}, {4}, {5}, {6}

Communicating class:
"every $i,j \in C$ has $i \leftrightarrow j$"

Recurrent classes:
{6}

(absorbing state)

Recurrent:
"for every state that can eventually be reached from i, it is possible to eventually return to i"

Transient classes:
{1}, {2}, {3}, {4}, {5}

Ex 1.3 Assume $(X_n)_{n\geq 0}$ is Markov.

Show that for $r < n$, we have

$$p_{ij}^{(n)} = \sum_x{}' \, p_{ix}^{(r)} \, p_{xj}^{(n-r)} .$$

Solution:

$$p_{ij}^{(n)} \overset{\text{def.}}{=} \Pr\{X_{k+n} = j \mid X_k = i\} = \left\{\begin{array}{l} \text{we want to get } k+r \\ \text{in there somehow} \end{array}\right\}$$

$$= \sum_x{}' \Pr\{X_{k+n} = j, X_{k+r} = x \mid X_k = i\} =$$

$$= \sum_x{}' \Pr\{X_{k+n} = j \mid X_{k+r} = x, X_k = i\} \Pr\{X_{k+r} = x \mid X_k = i\}$$

$$= \left\{ \cdots \to x_{k-1} \to x_k \to x_{k+1} \to \cdots \to x_{k+r} \to \cdots \to x_{k+n} \to \cdots \right\}$$

seen from x_{k+n}, this node contains all information (including x_k)

$$= \sum_x{}' \Pr\{X_{k+n} = j \mid X_{k+r} = x\} \underbrace{\Pr\{X_{k+r} = x \mid X_k = i\}}_{\overset{\text{def.}}{=} p_{ix}^{(r)}}$$

$$= \left\{\begin{array}{l} \text{Let } k' = k+r \iff k = k' - r. \\ \text{Then: } k+n = k' - r + n = k' + (n-r) \end{array}\right\}$$

$$= \sum_x{}' \underbrace{\Pr\{X_{k'+(n-r)} = j \mid X_{k'} = x\}}_{\overset{\text{def.}}{=} p_{xj}^{(n-r)}} p_{ix}^{(r)}$$

$$= \sum_x{}' p_{ix}^{(r)} p_{xj}^{(n-r)}$$

$\not\!\!\chi$

Ex 1.4 Consider $P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}$, $p+q \neq 0$ and $0 \leq p, q \leq 1$.

Compute P^n.

(Note: If $p+q=0$, then $P=I$ and $P^n = I$, trivially.)

Solution:

A well-known trick from linear algebra is to compute matrix powers via diagonalized forms:

with $P = QDQ^{-1}$, we have that

$$P^n = \underbrace{(QDQ^{-1})(QDQ^{-1}) \cdots (QDQ^{-1})}_{n \text{ times}} =$$

cancelation

$$= QD^nQ^{-1}.$$

We can accomplish this using an eigen/spectral factorization, where D is a diagonal matrix of eigenvalues and Q has the corresponding eigenvectors as columns.

Recall that a (sufficient) condition for being able to diagonalize a matrix is that all its eigenvalues are distinct.

(In general, the algebraic and geometric multiplicities of each eigenvalue have to be equal.)

Step 1: Compute eigenvalues

Remark: Recall that P is a stochastic matrix. This means that the elements on each row sum to one. In math:

$$P\mathbb{1} = \mathbb{1}$$

where $\mathbb{1}^T = [1, \dots, 1]$. But this is an eigenvector equation!

$$P\mathbb{1} = 1 \cdot \mathbb{1}$$

eigenvector eigenvalue

$\Longrightarrow$ All stochastic matrices have an eigenvalue at 1 and a (right) eigenvector $\mathbb{1}$.

The characteristic equation:

$$\det(P - \lambda I) = 0 \implies$$

$$0 \stackrel{!}{=} \det\begin{pmatrix} 1-p-\lambda & p \\ q & 1-q-\lambda \end{pmatrix} = (1-p-\lambda)(1-q-\lambda) - pq =$$

$$= \left[(1-p)-\lambda\right]\left[(1-q)-\lambda\right] - pq =$$

$$= (1-p)(1-q) - \lambda(1-p) - \lambda(1-q) + \lambda^2 - pq =$$

$$= 1 - q - p + \cancel{pq} - \lambda(1-p-q+1) + \lambda^2 - \cancel{pq} =$$

$$= 1-(p+q) - \lambda(1-(p+q)) - \lambda + \lambda^2 =$$

$$= \left[1-(p+q)\right]\left[1-\lambda\right] - \lambda(1-\lambda) =$$

$$= \left[1-(p+q)-\lambda\right]\left[1-\lambda\right]$$

which has solutions

$$\lambda_1 = 1 \qquad \lambda_2 = 1-(p+q).$$

(which we knew already)

Step 2: Compute eigenvectors.

$\lambda_1 = 1$: From $P\mathbb{1} = \mathbb{1}$, we know that $\mathbb{1}$ is an eigenvector

$\lambda_2 = 1-(p+q)$:

Let $v = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

Then

$$Pv = \lambda_2 v \implies$$

$$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = (1-(p+q)) \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \implies$$

The equations are (by definition) singular.
The first row gives:

$$\alpha(1-p) + \beta p = \alpha - (p+q)\alpha \implies$$

$$\alpha(1-p+p+q-1) = -\beta p \implies$$

$$\alpha q = -\beta p \implies$$

$$\alpha = -p/q \cdot \beta$$

Hence,

$$v = \begin{bmatrix} -p/q \cdot \beta \\ \beta \end{bmatrix} = \begin{bmatrix} p \\ -q \end{bmatrix} \cdot \frac{-\beta}{q} \in \operatorname{span}\left\{ \begin{bmatrix} p \\ -q \end{bmatrix} \right\}$$

Step 3: Compute the diagonalization

with

$$D = \begin{bmatrix} 1 & 0 \\ 0 & 1-(p+q) \end{bmatrix}$$

$$Q = \begin{bmatrix} 1 & p \\ 1 & -q \end{bmatrix}$$

we have that

$$P^n = QD^nQ^{-1} = \begin{bmatrix}1 & p\\ 1 & -q\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & 1-(p+q)\end{bmatrix}^n\begin{bmatrix}1 & p\\ 1 & -q\end{bmatrix}^{-1}$$

$$= \frac{1}{-q-p}\begin{bmatrix}1 & p\\ 1 & -q\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & [1-(p+q)]^n\end{bmatrix}\begin{bmatrix}-q & -p\\ -1 & 1\end{bmatrix} = \left\{\begin{matrix}\text{Let}\\ \alpha = 1-(p+q)\end{matrix}\right\} =$$

$$= \frac{1}{p+q}\begin{bmatrix}1 & p\\ 1 & -q\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & \alpha^n\end{bmatrix}\begin{bmatrix}q & p\\ 1 & -1\end{bmatrix} =$$

$$= \frac{1}{p+q}\begin{bmatrix}1 & p\alpha^n\\ 1 & -q\alpha^n\end{bmatrix}\begin{bmatrix}q & p\\ 1 & -1\end{bmatrix} =$$

$$= \frac{1}{p+q}\begin{bmatrix}q+p\alpha^n & p-p\alpha^n\\ q-q\alpha^n & p+q\alpha^n\end{bmatrix} =$$

$$= \frac{1}{p+q}\begin{bmatrix}q+p\alpha^n & p(1-\alpha^n)\\ q(1-\alpha^n) & p+q\alpha^n\end{bmatrix}.$$

<u>Remark:</u>

The second eigenvalue $\lambda_2 \overset{\text{here}}{=} 1-(p+q) \overset{\text{def.}}{=} \alpha$ determines how fast the chain converges to its stationary distribution/forgets its initial conditions.

Consider one column of $(P^T)^n$:

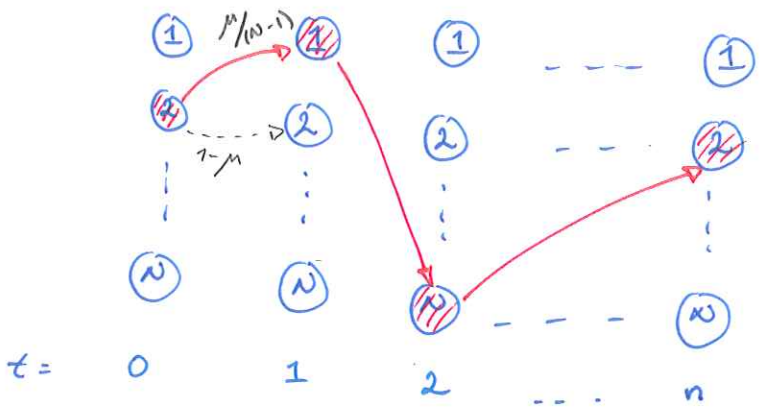
$$\frac{1}{p+q}\begin{bmatrix}q+p\alpha^n\\ p-p\alpha^n\end{bmatrix} = \frac{1}{p+q}\begin{bmatrix}q\\ p\end{bmatrix} + \frac{p}{p+q}\begin{bmatrix}1\\ -1\end{bmatrix}\alpha^n \longrightarrow \underbrace{\frac{1}{p+q}\begin{bmatrix}q\\ p\end{bmatrix}}_{\pi_\infty}$$

at a rate determined by α^n.

Ex 1.11 A virus can be in N strains. Changes strain w.p. μ in every generation.
What is the probability that the virus is of the same strain in the n^{th} generation, as initially?

Solution:

There are N different strains.



The probability that it stays the same is $1-\mu$.

The probability that it mutates to any other particular strain is $\frac{\mu}{N-1}$.

This corresponds to the following transition matrix:

$$P = \begin{array}{c} \\ 1 \\ 2 \\ \vdots \\ N \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & \cdots & N \end{array} \\ \begin{bmatrix} 1-\mu & \frac{\mu}{N-1} & \cdots & \frac{\mu}{N-1} \\ \frac{\mu}{N-1} & 1-\mu & & \frac{\mu}{N-1} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mu}{N-1} & \frac{\mu}{N-1} & \cdots & 1-\mu \end{bmatrix} \end{array}$$

Given an initial strain s_0, we need to compute

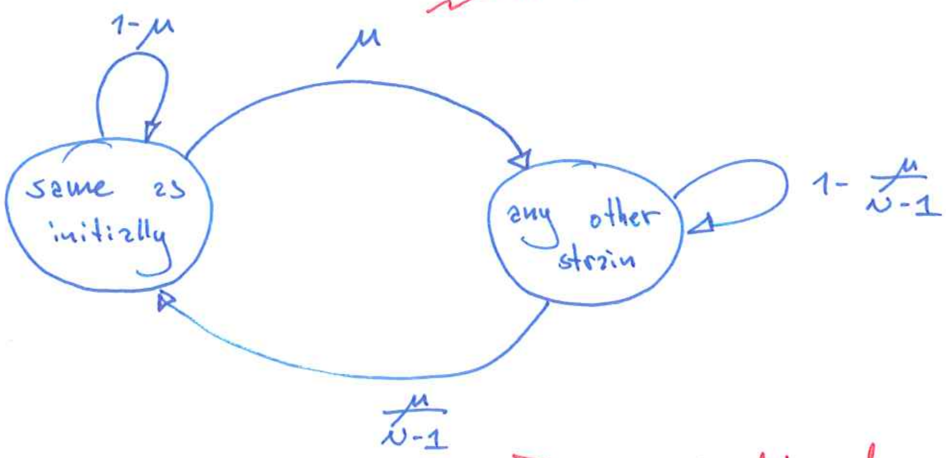
$$\Pr\{X_n = s_0 \mid X_0 = s_0\} = P^n_{s_0,s_0}.$$

(Note that, by the symmetry, this is equal for any $s_0 \in \{1, \ldots, N\}$.)

One way to solve the problem is to check if P is diagonalizable and then compute $P^n = QD^nQ^{-1}$ as before. However, finding the eigenvectors (for Q) gets tedious fast if N is large.
Can we remodel the problem?

Consider instead the following model:

it mutates to another strain



probability of mutating (μ) and ending up in the initial strain ($\frac{1}{N-1}$)

It has transition matrix:

$$P = \begin{array}{c} \\ \text{s.a.i.} \\ \text{a.o.s.} \end{array} \begin{array}{cc} \text{s.a.i.} & \text{a.o.s.} \\ \multicolumn{2}{c}{\begin{bmatrix} 1-\mu & \mu \\ \frac{\mu}{N-1} & 1-\frac{\mu}{N-1} \end{bmatrix}} \end{array}$$

We want to compute

$$\Pr\{X_n = \text{s.a.i.} \mid X_0 = \text{s.a.i.}\} = P^n_{\text{s.a.i.},\text{s.a.i.}} = P^n_{1,1}.$$

From exercise 1.4, we know that for a transition matrix

$$P = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix} \sim \begin{bmatrix} 1-\mu & \mu \\ \frac{\mu}{N-1} & 1-\frac{\mu}{N-1} \end{bmatrix},$$

it holds that

$$P^n_{1,1} = \frac{1}{p+q}\left(q + p\alpha^n\right) = \{\alpha = 1-(p+q)\} =$$

$$= \frac{1}{p+q}\left(q + p(1-p-q)^n\right) = \left\{ \begin{array}{l} \text{Our variables:} \\ p = \mu, \; q = \frac{\mu}{N-1} \end{array} \right\}$$

$$= \frac{1}{\mu + \frac{\mu}{N-1}}\left(\frac{\mu}{N-1} + \mu\left(1 - \mu - \frac{\mu}{N-1}\right)^n\right) =$$

$$= \frac{N-1}{N-1+1}\left(\frac{1}{N-1} + \left(1 - \mu\left[\frac{N-1+1}{N-1}\right]\right)^n\right) =$$

$$= \frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(1 - \frac{\mu N}{N-1}\right)^n$$

The symmetry of the problem allowed us to lump states. Not always possible.

Take home:

How you model the problem is important, it can make the solution easy or difficult!

Ex 1.6 | Consider a doubly stochastic matrix ($\mathbb{1}^T P = \mathbb{1}$, $P\mathbb{1} = \mathbb{1}$, $P \geq 0$) that is ergodic (aperiodic + irreducible). Show that its stationary distribution is the uniform.

Solution:

If P is ergodic, then the limiting distribution is equal to the unique stationary distribution defined by

$$\begin{cases} P^T \pi_\infty = \pi_\infty, & \text{(stationary)} \\ \pi_\infty^T \mathbb{1} = 1. & \text{(sum-to-one)} \end{cases}$$

(i.e., for an ergodic MC, these equations have a unique solution)

Let's check if $\frac{1}{N}\mathbb{1}$ solves these equations:

i) $P^T\left(\frac{1}{N}\mathbb{1}\right) = \frac{1}{N} P^T \mathbb{1} = \frac{1}{N}\left(\mathbb{1}^T P\right)^T$

$= \left\{ \begin{array}{c} \text{By doubly stochastic assumption} \\ \mathbb{1}^T P = \mathbb{1}^T \end{array} \right\}$

$= \frac{1}{N}(\mathbb{1}^T)^T = \frac{1}{N}\mathbb{1}$ ok!

ii) $\left(\frac{1}{N}\mathbb{1}\right)^T \mathbb{1} = \frac{1}{N}\mathbb{1}^T\mathbb{1} = \frac{1}{N}\cdot N = 1$ ok!

We conclude that $\pi_\infty = \frac{1}{N}\mathbb{1}$ since it is a solution, and for ergodic M.C. it is unique.

Ex 1.2

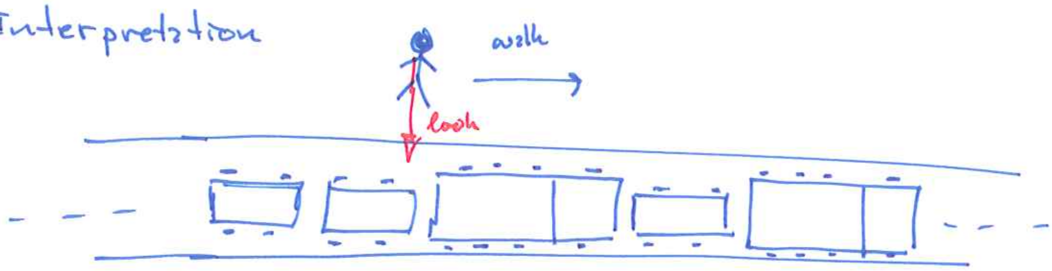
On a road,
the probability that a truck is followed by a car is $3/4$ and the probability that a car is followed by a truck is $1/5$.

a, what is the proportion between the vehicle types?

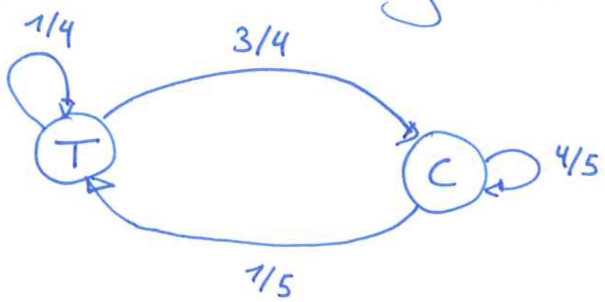
b, If I see a truck pass, on average how many vehicles will pass before the next truck?

Solution:

Interpretation



Let the state be the currently observed vehicle:



The corresponding transition matrix:

$$P = \begin{array}{c c} & \begin{array}{cc} T & C \end{array} \\ \begin{array}{c} T \\ C \end{array} & \begin{bmatrix} 1/4 & 3/4 \\ 1/5 & 4/5 \end{bmatrix} \end{array}$$

2, The stationary distribution describes the proportions in which the chain occupies its different states in the long run.

For us, it gives the proportion between cars and trucks on the road.

The chain is ergodic (aperiodic & irreducible), so we can solve

$$\begin{cases} P^T \pi_\infty = \pi_\infty \\ \mathbb{1}^T \pi_\infty = 1 \end{cases}$$

for a unique solution. Let $\pi_\infty = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$.

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

$$\begin{bmatrix} 1/4 & 1/5 \\ 3/4 & 4/5 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \implies \text{(First row)}$$

Remark: $P^T \pi_\infty = \pi_\infty$ is an eigenvector equation, so the rows are linearly dependent.

$$\frac{\alpha}{4} + \frac{\beta}{5} = \alpha \implies$$

$$4\beta = 20\alpha - 5\alpha \implies$$

$$\beta = \frac{15}{4}\alpha$$

Hence,

$$\pi_\infty = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \alpha \\ \frac{15}{4}\alpha \end{bmatrix} = \begin{bmatrix} 1 \\ 15/4 \end{bmatrix}\alpha.$$

Then we normalize to get the unique solution

$$1 \hat{=} \mathbb{1}^T \pi_\infty = [1 \quad 1] \begin{bmatrix} 1 \\ 15/4 \end{bmatrix}\alpha = \left(1 + \frac{15}{4}\right)\alpha = \frac{19}{4}\alpha \implies$$

$$\alpha = \frac{4}{19} \Rightarrow \beta = \frac{15}{4} \cdot \alpha = \frac{15}{4} \frac{4}{19} = \frac{15}{19}$$

The stationary distribution is

$$\pi_\infty = \begin{bmatrix} 4/19 \\ 15/19 \end{bmatrix}, \quad \begin{array}{l} \text{——— } \pi_\infty(T) \\ \text{——— } \pi_\infty(C) \end{array}$$

and the proportion of cars on the road is $\pi_\infty(C) = 15/19$.

b, The mean return time to a state j is given by $1/\pi_\infty(j)$. In this example, note that "time" is indexed by "number of seen vehicles".

The mean time (number of vehicles) between two trucks is:

$$r_{T,T} = \frac{1}{\pi_\infty(T)} = \frac{19}{4}.$$