

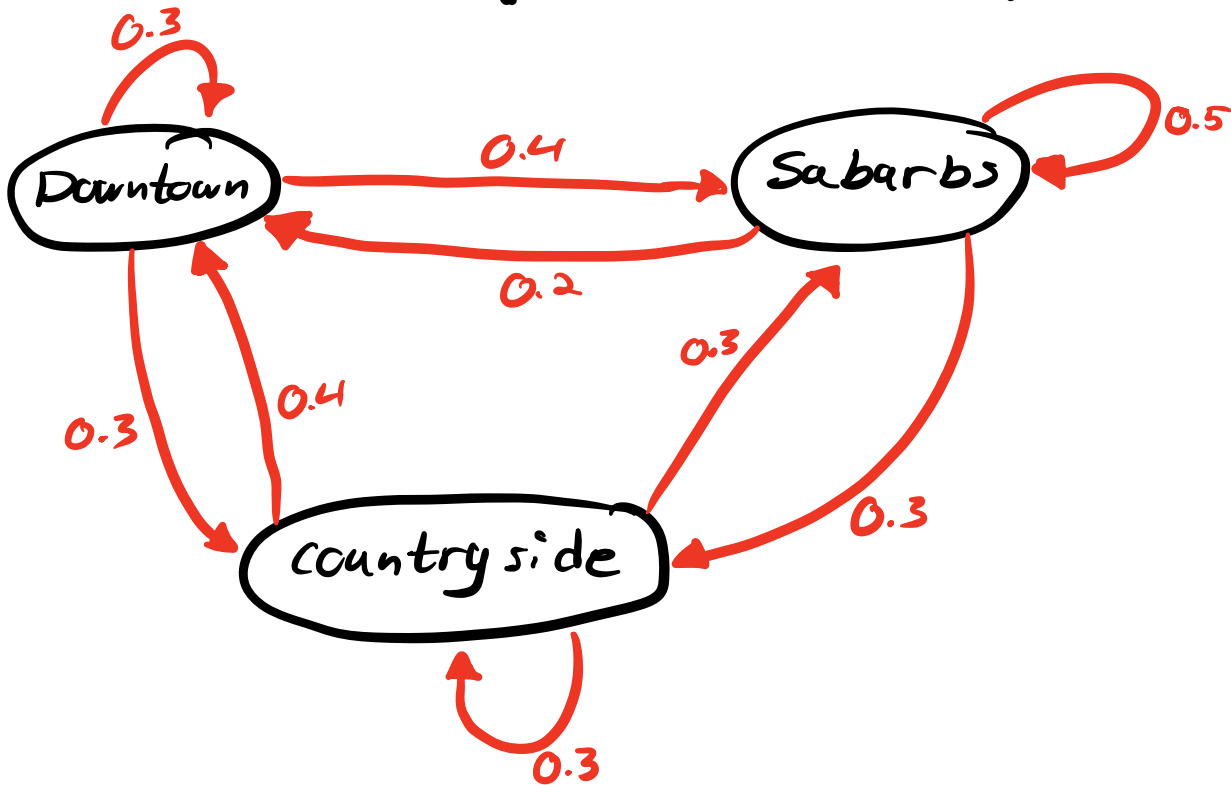
# Problem ④

①

1. If a truck is currently in the suburbs, what is the probability that it will be in the downtown region after two time steps? [1.5p]

Transition Matrix:

$$P = \begin{matrix} & \begin{matrix} \text{Downtown} & \text{Suburbs} & \text{Country side} \end{matrix} \\ \begin{matrix} \text{Downtown} \\ \text{Suburbs} \\ \text{Country side} \end{matrix} & \begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix} \end{matrix}$$



So: The transition matrix is  $P$  above.

We want the probability of starting in suburbs and being in Downtown after two time steps

In Markov chains, two-step transitions are

computed with the matrix:

$$P \cdot P = P^2$$

So: Use matmul to multiply, then look at the element at row Suburbs, column Downtown.

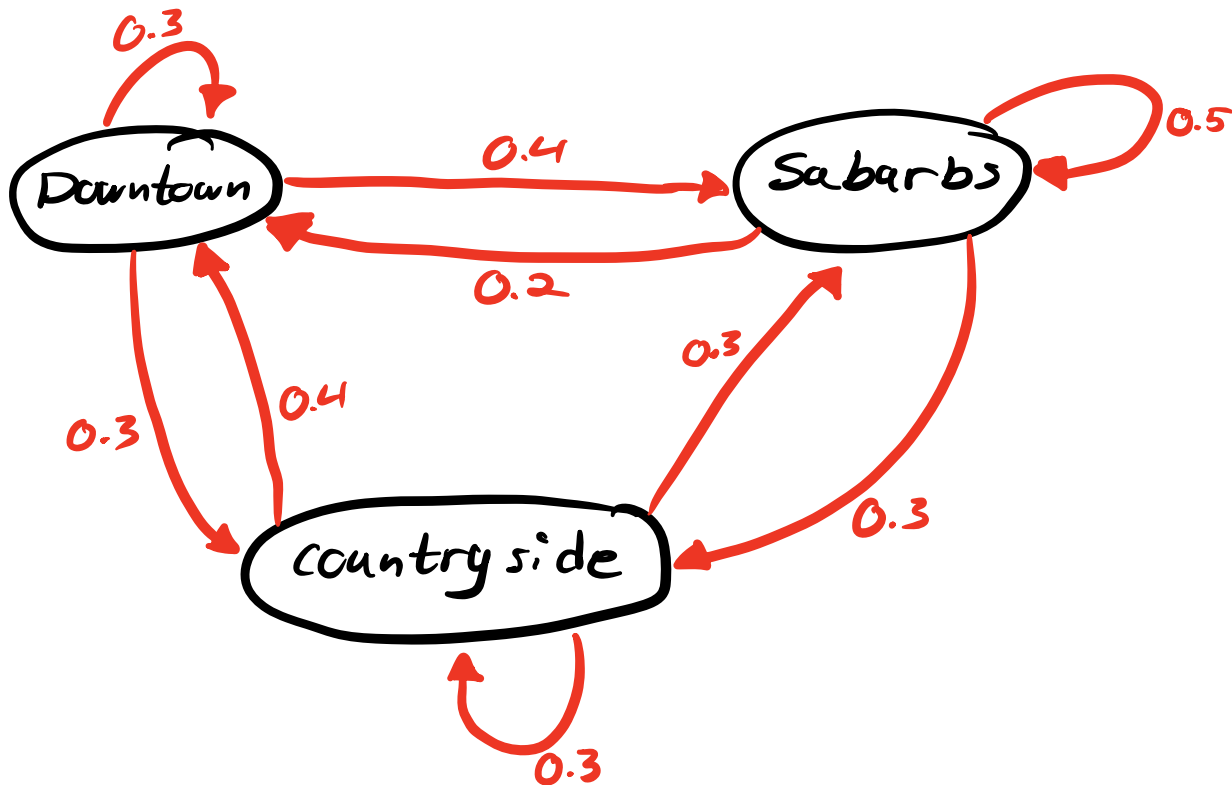
This gives 0.28:

## Problem ①

②

2. If a truck is currently in the suburbs, what is the probability that it will be in the downtown region the first time after two time steps? [1.5p]

This is done by looking at all possible paths, and then sum them up. We look at transition matrix.



Path 1: Suburbs  $\rightarrow$  Suburbs  $\rightarrow$  Downtown

$$\Rightarrow 0.5 \cdot 0.2 = 0.1$$

Path 2: Suburbs  $\rightarrow$  countryside  $\rightarrow$  Downtown

$$\Rightarrow 0.3 \cdot 0.4 = 0.12$$

$$\text{Total: } 0.1 + 0.12 = \underline{\underline{0.22}}$$

Problem ④:

③

- A Markov chain is irreducible if:

- You can get from every state to every other state
- As long as there is some path with positive probability.

\* If matrix has no zeros at all, it is trivially irreducible, because all transitions can occur in 1 step

\* This does not need to necessarily happen in one step

So: No zeros in matrix  $\rightarrow$  irreducible!

Answer:

- In this case it's true since you can go from each state to each other state.  
There's no zeros in markov chain.

## Problem ①

④

4. What is the stationary distribution? [1.5p]

We label the stationary distribution as

$$\pi = (\pi_1, \pi_2, \pi_3) = (\text{Downtown}, \text{Suburbs}, \text{Country side})$$

Transition matrix is still the same:

$$\begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}$$

Stationary distribution satisfies:  $\pi \cdot P = \pi$

$$\Rightarrow \begin{cases} \pi_1 = 0.3\pi_1 + 0.2\pi_2 + 0.4\pi_3 \\ \pi_2 = 0.4\pi_1 + 0.5\pi_2 + 0.3\pi_3 \\ \pi_3 = 0.3\pi_1 + 0.3\pi_2 + 0.3\pi_3 \end{cases}$$

and we also know that:  $\pi_1 + \pi_2 + \pi_3 = 1$

Put these 4 equations into matrixes and solve:

## Problem ①

⑤

5. Advanced question: What is the expected number of steps until the first time one enters the downtown region having started in the suburbs region. Hint: to get within 1 decimal point, it is enough to compute the probabilities for hitting times below 30. [2p]

**Goal:**

Compute the expected number of steps until the chain first enters **Downtown**, starting in **Suburbs**.

We label the states as:

- 0 = Downtown (D)
- 1 = Suburbs (S)
- 2 = Countryside (C)

Transition matrix:

$$P = \begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.2 & 0.5 & 0.3 \\ 0.4 & 0.3 & 0.3 \end{pmatrix}$$


---

**1. Define the hitting time**

Let

$$T = \min\{n \geq 0 : X_n \text{ is in Downtown}\}$$

and define for each state  $i$ :

$$h(i) = \mathbb{E}[T \mid X_0 = i].$$

We want

$$h(S).$$

Since downtown is the target:

$$h(D) = 0.$$

Let:

- $h_S = h(S)$
- $h_C = h(C)$ .

**2. Use first-step decomposition**

For any non-target state  $i$ :

$$h(i) = 1 + \sum_j P_{ij} h(j).$$

**Suburbs row of  $P$**

$$h_S = 1 + 0.2h_D + 0.5h_S + 0.3h_C.$$

Since  $h_D = 0$ :

$$h_S = 1 + 0.5h_S + 0.3h_C.$$

**Countryside row**

$$h_C = 1 + 0.4h_D + 0.3h_S + 0.3h_C.$$

Again  $h_D = 0$ :

$$h_C = 1 + 0.3h_S + 0.3h_C.$$


---

**3. Rearranging the system**

Equation for  $h_S$ :

$$0.5h_S - 0.3h_C = 1.$$

Equation for  $h_C$ :

$$-0.3h_S + 0.7h_C = 1.$$

System:

$$\begin{cases} 0.5h_S - 0.3h_C = 1 \\ -0.3h_S + 0.7h_C = 1 \end{cases}$$

#### 4. Solve the system

Multiply the first equation by 3:

$$1.5h_S - 0.9h_C = 3$$

Multiply the second by 5:

$$-1.5h_S + 3.5h_C = 5$$

Add them:

$$2.6h_C = 8$$

Thus:

$$h_C = \frac{40}{13}.$$

Plug into the first equation:

$$0.5h_S = 1 + 0.3 \cdot \frac{40}{13} = \frac{25}{13}.$$

Therefore:

$$h_S = \frac{50}{13} \approx 3.8461538.$$

#### ✓ FINAL ANSWER

$$\mathbb{E}[T \mid X_0 = \text{Suburbs}] = \frac{50}{13} \approx 3.8 \text{ steps}$$

Problem ③ :

### Assignment 2, PROBLEM 3

Maximum Points = 4

Derive the maximum likelihood estimate for  $n$  IID samples from a random variable with the following probability density function:

$$f(x; \lambda) = \frac{1}{24} \lambda^5 x^4 \exp(-\lambda x), \quad \text{where, } \lambda > 0, x > 0$$

You can solve the MLe by hand (using pencil paper or using key-strokes). Present your solution as the return value of a function called `def MLeForAssignment2Problem3(x)`, where `x` is a list of  $n$  input data points.

$$f(x; \lambda) = \frac{1}{24} \cdot \lambda^5 \cdot x^4 \cdot e^{-\lambda x}$$

$$\log(f(x; \lambda))$$

$$\Rightarrow n \log\left(\frac{1}{24}\right) + n \cdot 5 \log(\lambda) + 4 \log(x) - \lambda x$$

$$\log\left(\frac{1}{24}\right) = \log(24^{-1}) = -\log(24)$$

$$\Rightarrow -n \log(24) + 5n \log(\lambda) + 4 \sum_{i=1}^n \log(x_i) - \lambda \sum_{i=1}^n x_i$$

The maximum likelihood estimate (MLE) is the value of the parameter (here  $\lambda$ ) that maximizes the likelihood of observing our data.

$$\text{MLE: } \hat{\lambda} = \underset{\lambda}{\operatorname{argmax}} \underbrace{\ell(\lambda)}_{\text{loss}}$$

We maximize a function by looking for critical points:

$$\frac{d}{d\lambda} \ell(\lambda) = 0$$

- If slope positive ( $\ell'(\lambda) > 0$ ), function is increasing
- If slope negative ( $\ell'(\lambda) < 0$ ), function is decreasing

• At a peak or valley, the slope is flat:  $\ell'(\lambda) = 0$ .

So, solving  $\ell'(\lambda) = 0$  tells where maxima or minima can occur.

Once we have derivative  $= 0$ , we check second derivative:

•  $\ell''(\lambda) < 0$ : it's concave down  $\rightarrow$  maximum

•  $\ell''(\lambda) > 0$ : it's concave up  $\rightarrow$  minimum.

So:  $\ell'(\lambda) = 0$  gives:  $\frac{d}{d\lambda} \ell = \frac{1}{\lambda}$ .

$$5n(\log \lambda) - \lambda \sum_{i=1}^n x_i$$

$$\Rightarrow 5n \frac{1}{\lambda} - \sum_{i=1}^n x_i = 5n \lambda^{-1} - \sum_{i=1}^n x_i$$

$$\ell''(\lambda) = \lambda^{-2} = 5n \frac{1}{\lambda^2} = -\frac{5n}{\lambda^2} < 0$$

So this is indeed a maximum.

So:  $\ell'(\lambda) = 0 \Rightarrow \frac{5n}{\hat{\lambda}} - \sum_{i=1}^n x_i = 0$

$$\Rightarrow \frac{5n}{\hat{\lambda}} = \sum_{i=1}^n x_i$$

$$\Rightarrow \hat{\lambda} = \frac{5n}{\sum_{i=1}^n x_i}, \quad \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i = n\bar{x}, \text{ since we need mean of } x.$$

$$\Rightarrow \hat{\lambda} = \frac{5n}{n\bar{x}} \Rightarrow \frac{5}{\bar{x}}$$


---

Problem (4)