Exercise 7 – *Bayesian inference and Data assimilation*

Due by: Tuesday, 6 June 2023, 23:59 (CEST)

Problem 1 Finally we find a precise weather forecasting model: It models the weather as sunny, overcast or rainy, i.e., $S = \{\text{sunny}, \text{overcast}, \text{rainy}\}$. The probability of observing a certain weather condition in the next day only depends on today's weather, and the probability is given by the following table:

↓Tomorrow \Today→	sunny	overcast	rainy
sunny	1/2	1/4	1/5
overcast	1/3	1/2	2/5
rainy	1/6	1/4	2/5

Then the model is expressed with the following matrix

$$P := \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{2} & \frac{2}{5} \\ \frac{1}{6} & \frac{1}{4} & \frac{2}{5} \end{pmatrix}$$

1. Suppose the weather today is sunny. What is the probability that it will be sunny, overcast or rainy on the day after tomorrow?

2. Find the invariant measure of the Markov process. That is, solve for a probability vector p such that $p = Pp$.

3. Implement the Markov chain. Suppose at day 1, the weather is sunny. Run 190 parallel simulation. Plot the histograms at day 2, 3 and 30 of each weather condition over all samples. Compare them with the invariant measure you have found in step 2.

4. Does it seem to converge to the invariant measure? Analytically show that the chain converges to the invariant measure (hint: find eigenvalues other than 1).

5. Suppose at day 2, your agent tell you that they cannot see the sun, but do not tell you whether it is just cloudy or it rains. Using the simulation result from the previous step and the given information, how would you determine the probability of rain at day 2?

6. Under the same assumption as the previous step, discard all simulated result that says day 2 is sunny, and plot the histogram at day 3 among the remaining. Compare with the result on day 3 from step 4 (without the extra information).

7. It turns out that your agent has very unstable mind, so the information from them is only trustworthy with probability 1/2. Otherwise, the agent is completely trolling and the information has NO USE. Use your simulation result to forecast the weather in day 3, while considering both cases when the information is true and when it should be ignored.

8. **[Optional discusstion]** Note that we had to disregard some of our simulation result that does not align with the extra information. This means that if we gather more information, then more samples would be disregarded, and thus we effectively left with fewer samples. Probably those trustful samples are somewhat "accurate" but we lose flexibility in the future. How would you overcome this paradox?