M1 SCA Université de Lorraine Enseignant : Laure Buhry

[UE 805: Memory and Machine Learning
Practical 2 – Discrete-Time Markov Chains

Exercise 1

Let us consider a TV game show where a contestant has to answer questions that are more and more difficult.
If the players correctly answer n questions, they go to state $n+1$, and have a probability p_{n+1} of correctly answering the following question, which bring them to state $n+2$. Their probability to fail is thus $1-p_{n+1}$, then they fall into state 0 and loose.

a) Let $(X^{(n)})_{,n\geq 0}$ be a sequence giving the state of a player at time n. Is $(X^{(n)})_{,n\geq 0}$ a Markov chain?

b) Represent the different states and their transition probabilities as a diagram.

c) Give the associated transition matrix $P = P_{ij}$.

d) Give the mathematical expression of P^n. *Direction: distinguish 3 cases.*

Exercise 2

Let $(X_n)_{N\geq 0}$ be a Markov chain defined for states 1 to 7 by the transition matrix P:

$$P = \begin{pmatrix} 1/2 & 1/4 & 0 & 1/4 & 0 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 & 0 & 1/2 \\ 0 & 0 & 1/8 & 0 & 7/8 & 0 & 0 \\ 1/4 & 0 & 0 & 0 & 0 & 0 & 3/4 \\ 0 & 1/9 & 7/9 & 0 & 0 & 1/9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

a) Draw the diagram associated with this Markov chain.

b) Determine the recurrent and transient classes. Specify whether the chain is irreducible or not.

c) Calculate $\mathbf{P}(X_2 = 6|X_0 = 3)$ and $\mathbf{P}(X_2 = 4|X_0 = 2)$.

The goal of the following practical aims at numerically handling Markov chains so that one can identify specific asymptotic behaviors only by simulations.
You will make use of *Python* and its associated scientific, *Scipy*, and plotting, *Matplotlib*, libraries.
Figure examples and associated scripts are available at:
http://matplotlib.org/users/screenshots.html

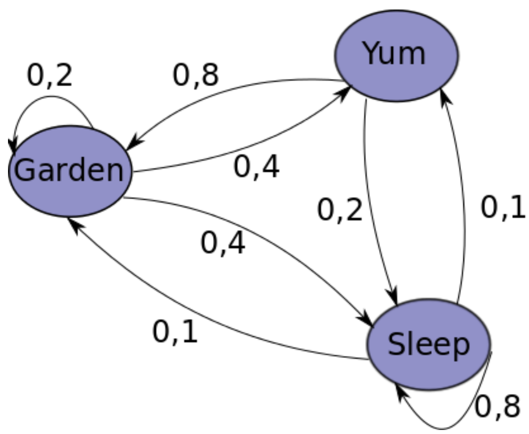
A summary of function available in *Matplotlib* is accessible here:
http://matplotlib.org/api/pyplot_summary.html

Remark: Do not forget to import *numpy*. You will then have to import the following modules:
scipy.linalg (to handle linear algebra tools; from scipy import scipy.linalg)
numpy.random (to handle statistical tools, generate pseudo-random numbers, etc)
matplotlib.pyplot (for plots)

Exercise 3 – Computer lab

Let us go back to the example of "Leo Cat" seen during the last lecture:
Leo can essentially be in one of the three following states: sleeping in bed on the couch (C), eating cat nuggets (M), walking in the garden (J).

- After Leo falls asleep, there are 8 chances out of 10 that it does not wake up in less than a minute
- After he wakes up, there is one chance out of 2 that he eats and one chance out of 2 that he goes to the garden
- After he eats, there are 2 chances out of 10 that he goes back to sleep and 8 chances out of 10 that he goes to the garden
- After arriving in the garden, there are 2 chances out of 10 that he stays outside, 4 chances out of 10 that he goes back to sleep, and 4 chances out of 10 that he eats.



The transition matrix P associated with the corresponding Markov chain is defined by:

$$\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix} \qquad \text{With 1: sleep, 2: yum, 3: garden}$$

a) How would you try to verify, by simulations, what are the recurrent and transient states? Based on the definition of stationary distribution provided below, how would you numerically (by simulations only) check that the system converges toward a stationary distribution?

Implement the random walk of Leo cat for 50, 100, and 1000 iterations and for each, run 3 simulations, initializing the walk with each possible state.

b) Plot the normalized histogram of the number of passages by the different states after 50, 100 and 1000 steps with different initializations for each. For this purpose, you will use the Matplotlib library. What do you observe?

c) Using the mathematical proposition about P, now **calculate** the probabilities of being in each state after 100 and 1000 steps with different initializations, one in each state. Compare the results with b). What do you notice?

Vocabulary : Let P be a transition matrix. There can exist one or several measures $\pi = (\pi_i)_{i \in E}$ on the state space E, so that: $\pi = \pi P$ or $\forall j \in E, \qquad \pi_j = \sum_{i \in E} \pi_i \, p_{i,j}$.
Such a measure π is called ***invariant measure***. It is also called ***stationary dustribution*** or ***equilibrium distribution*** if the following condition holds:

$$\forall i \in E, \qquad \pi_i \ \geq \ 0, \sum_{i \in E} \pi_i \ = \ 1.$$

In the present exercise, the cat's behavior converges toward a ***a stationary distribution***: after a certain amount of time, the probability distribution of its behavior is independent from the initial distribution.
Let q be its behavior at infinite time: $\mathbf{q} = \lim_{n\to\infty} \mathbf{x}^{(n)}$.
There is convergence because the Markov chain is irreducible and aperiodic.
One can then search q by solving the following equation:

$$\begin{aligned}
\mathbf{q}P &= \mathbf{q} \qquad (\mathbf{q} \text{ est la loi invariante par rapport à } P.) && (1)\\
&= \mathbf{q}I && (2)\\
\Leftrightarrow \mathbf{q}(I-P) &= \mathbf{0} && (3)\\
\mathbf{q}(I-P) &= \mathbf{q}\left(\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}\right) && (4)\\
&= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} 0,2 & -0,1 & -0,1 \\ -0,2 & 1 & -0,8 \\ -0,4 & -0,4 & 0,8 \end{bmatrix} && (5)\\
&= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} && (6)
\end{aligned}$$

Knowing that $q_1 + q_2 + q_3 = 1$, one can obtain $\begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix}$

d) Calculate q_1, q_2 et q_3.
Remark: One should normally find again the values obtained in the histogram by numerical simulations (b) after a large ammount of time steps.

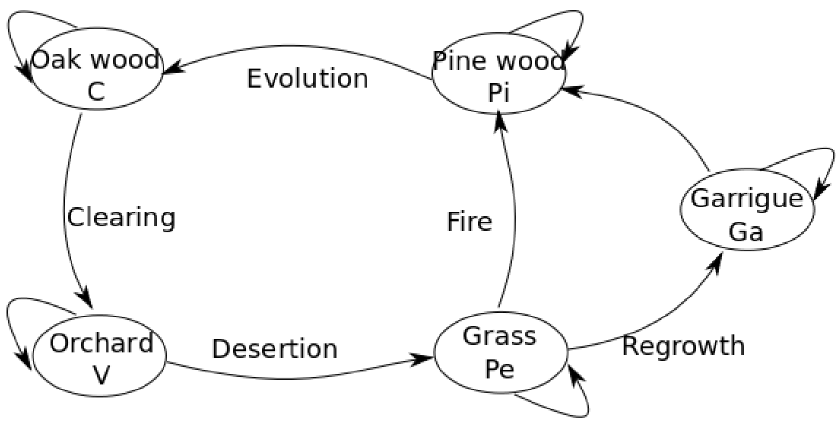
Exercise 4 – TP

Let us consider the evolution of Mediterranean ecosystems[1]. Originally, the Mediterranean forest, on limestone at low altitude, was dominated by oak trees (Quercus pubescens). But human activity decimated those original

[1]Example from the book *Modélisation et simulation d'écosystèmes*, P. Coquillard et D. Hill, Masson 1997

forests to substitute them for grazing, farmlands, orchards, etc. Then, farming was progressively stopped; but instead of yielding natural regrowth of oak trees, after a garrigue state, the implantation of pine trees (Pinus halepensis) was favored. Yet, these substituted highly inflammable forest, are often victims of fires (arson or natural). Therefore, they are permanently partly destroyed and recreated.

Here are the diagram and transition matrix of the Markov chain modelling the evolution of this ecosystem. This system has five states denoted as $X = C, V, Pe, Ga, Pi$.



$$P = \begin{pmatrix} 0,8 & 0,2 & 0 & 0 & 0 \\ 0 & 0,7 & 0,3 & 0 & 0 \\ 0 & 0 & 0,4 & 0,6 & 0 \\ 0 & 0 & 0 & 0,2 & 0,8 \\ 0,1 & 0 & 0,25 & 0 & 0,65 \end{pmatrix}$$

a) Is there any absorbing state?

b) Let $x^{(n)}$ be the state of the system at time n. One initialize it at $x^{(0)} = [1, 0, 0, 0, 0]$ because there were originally only oak trees.
Using numerical simulations of the Markov chain, study the evolution of the system; in other words, plot the evolution of $x^{(n)}$ when n varies and plot the histogram of the number of passages by the different states after $n = 1000$ steps. Use Python for the simulations.
What is the state of the system at $n = 2$, $n = 10$, $n = 50$, $n = 1000$? (in other words, give the values of $x^{(2)}$, $x^{(10)}$, $x^{(50)}$, $x^{(1000)}$). What is the probability of being in each state at $n = 2$, $n = 10$, $n = 50$, $n = 1000$?

c) Are there any recurrent states? Which ones?

d) What do you conclude? What teminology do we use for this kind of system?