Probabilities

Mathematics Department Efrei Paris

2024/2025 S3

P2 INT

Contents

1. Project 2024: Markov chains

1.1 Introduction

In ecology, an *ecosystem* is a community of living beings interacting with their environment. Probability theory is a tool that enables us to model the interactions of the components of an ecosystem in order to predict their evolution.

The aim of this project is to introduce a type of ecosystem modelling. In order to take several components into account, we propose that you work with graphs and discrete Markov chains.

1.2 A bit of theory

1.2.1 Definitions

Definition 1.1 (Oriented graph)
A finite oriented graph $G = (S, A)$ is defined by :

- a finite set of n vertices that we denote $S = \{s_1, s_2, \ldots, s_n\}$.
- a finite set $A = \{a_1, a_2, \ldots, a_m\}$ of ordered pairs of vertices, called *edges*: each a_i corresponds to an arrow when you draw the graph.

Definition 1.2 (Matrix associated to a weighted graph)
Let $G = (S, A)$ be an oriented graph. One can decide to associate a number to each edge (in which case one says that the graph is *weighted*). The set of all these values can be recorded in a matrix P such that the coefficient $p_{i,j}$ of the row i and column j represents the value associated to edge going from vertex j to vertex i.[1]

1.2.2 Example

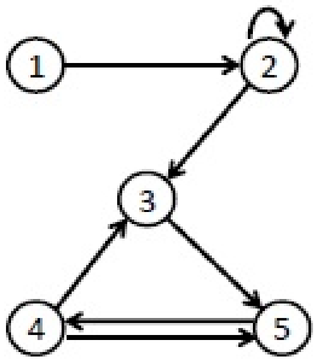

Figure 1.1 – $S = \{1,2,3,4,5\}$ is the set of vertices. There are 7 edges, such as the one going grom vertex 3 to vertex 5 for example.

For this graph, consider the following matrix:

$$P = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0,7 & 0 & 0 & 0 \\ 0 & 0,3 & 0 & 0,7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0,3 & 0 \end{pmatrix}$$

The second column, corresponding to vertex to 2, contains two non vanishing coefficients: $0,7$ and $0,3$ which can be interpreted as follows: the probability of going from vertex 2 (second column) to vertex 2 (second row) is 0,7. Similarly, the probability of going from vertex 2 (second column) to vertex 3 (third row) is 0,3.

Now let us return to the problem of an ecosystem. Suppose that, at the initial time, the ecosystem is composed of $0.3 = 30\%$ of species 2 and $0.7 = 70\%$ of species 5. We can then write this information

[1]Warning: in many references, the matrix is defined as the transposed of our definition.

in the following column vector:

$$X_0 = \begin{pmatrix} 0 \\ 0,3 \\ 0 \\ 0 \\ 0,7 \end{pmatrix}.$$

Suppose, moreover, that the matrix P mimics the time evolution of the system from one year to another. We can compute the distribution of the species in our ecosystem the following year, that is X_1.

$$X_1 = P.X_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 1 & 0,7 & 0 & 0 & 0 \\ 0 & 0,3 & 0 & 0,7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0,3 & 0 \end{pmatrix} \times \begin{pmatrix} 0 \\ 0,3 \\ 0 \\ 0 \\ 0,7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0,3 \times 0,7 \\ 0,3 \times 0,3 \\ 0,7 \times 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0,21 \\ 0,09 \\ 0,7 \\ 0 \end{pmatrix}$$

At the beginning of the year 1, the ecosystem will therefore be composed of 21% of species 2, 9% of species 3 and 70% of species 4.

Note that the evolution of our system is dictated only by the P matrix, which is constant. Thus, the prediction of the future distribution of a system will be entirely contained in the present distribution of that same system. We will then say that this process satisfies the *Markov property*.

1.3 Five states Mediterranean ecosystems

Example taken from the book Modélisation et simulation d'écosystèmes, P. Coquillard and D. Hill, Masson 1997

Originally, the Mediterranean forest, on limestone rock at low altitude, was almost certainly dominated by oaks (pubescent oaks). However, human activity eradicated these primitive forests and replaced them with pastureland, orchards, etc. The abandonment of all agricultural activity then encouraged the establishment of another species, the Aleppo pine, after passing through a state of garrigue. These replacement forests, which are highly flammable, are subject to recurrent fires (deliberate or accidental) and are therefore condemned to perpetual reconstitution.

To study the long-term evolution of this ecosystem, we will model this evolution by a five states Markov chain: oak grove (O), Vineyards and orchards (V), grass (G), garrigue (Ga) and pine forest (Pi). Time is discretised into time intervals of duration Δt (expressed in years). For any $n \in \mathbb{N}$, the vector

$$X_n = \begin{pmatrix} O \\ V \\ G \\ Ga \\ Pi \end{pmatrix}_n$$

will denote the distribution of this ecosystem at instant $t_n = n\Delta t$. We will suppose that X_0 is the uniform distribution.

The transition probability matrix of this Markov chain P is given by:

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

1. Draw the graph associated to this model.
2. Supposing that we are in state Pi, what is the probability of a trajectory of the type $Pi - G - Ga$? And of a trajectory of the type $Pi - O - V$?
3. Give an example of a trajectory with a vanishing probability.
4. Using a computer, compute the distribution of the ecosystem for $n = 20$.
   You can use Matlab (Help for matrices).
5. Plot the evolution of the ecosystem's distribution as a function of n.
   You can use Matlab (Help for plots and Help for 3D-plots).
   Looking at the numerical values obtained, explain the behaviour of the ecosystem at long times.
6. Using your Linear Algebra course (and Matlab for the computations), diagonalize P. In other words, find matrices A, A^{-1} and D (with D diagonal) such that $P = A.D.A^{-1}$. What do the matrices A, A^{-1} and D represent in terms of eigenvalues and eigenstates?
7. Using the previous question, find again the limiting distribution of the ecosystem when n tends to infinity.
8. The transition from Pi to G mimics the presence of fires. Study the long-term influence of fires, by varying its value in the matrix.

1.4 A more complex model for a more diverse ecosystem

Now imagine an ecosystem with 20 states. Let P be the transition probability matrix:

$$\begin{pmatrix}
0,1 & 0 & 0,8 & 0 & 0,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,2 & 0,4 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,1 & 0,2 & 0,05 & 0,9 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,5 & 0,2 & 0,05 & 0,05 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0,1 & 0,2 & 0,05 & 0,05 & 0,7 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,6 & 0,4 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,2 & 0,7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,2 & 0,1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0 & 0,3 & 0,3 & 0,3 & 0,4 & 0,1 & 0,3 & 0 & 0,15 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,3 & 0,05 & 0,3 & 0,1 & 0,1 & 0,1 & 0,2 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,15 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,1 & 0,05 & 0,05 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,1 & 0,1 & 0 & 0,1 & 0,2 & 0,1 & 0,2 & 0,3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0,1 & 0,05 & 0,2 & 0,1 & 0,1 & 0,2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0,2 & 0 & 0 & 0 & 0,05 & 0,1 & 0,05 & 0,1 & 0,1 & 0,2 & 0,1
\end{pmatrix}$$

9. Two states A and B are said to be *communicating states* if the graph contains a path from A to B and another path from B to A (careful: a path may contain more than one edge).
Give the groups of states that all communicate with each other.
Explain how this can be seen on the matrix.

10. Start from the distribution:

$$X_0 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

(where T denotes the transposed matrix).

a) Compute the distribution X_n of the system for $n = 2$, $n = 10$ and $n = 50$.

b) Plot X_n as a function of n.

c) Does the sequence $(X_n)_{n\in\mathbb{N}}$ admit a limiting distribution?

11. Now start from the distribution:

$$Y_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

a) Compute the distribution Y_n of the system for $n = 2$, $n = 10$ and $n = 50$.

b) Plot Y_n as a function of n.

c) Does the sequence $(Y_n)_{n\in\mathbb{N}}$ admit a limiting distribution?

12. Same questions starting from the initial distributions:

$$Z_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

and

$$W_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}^T$$

13. Finally, start from the distribution:

$$\alpha_0 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}^T$$

a) Plot α_n as a function of n.

b) Does the sequence $(\alpha_n)_{n\in\mathbb{N}}$ admit a limiting distribution?

c) Find a linear relation between the limiting distributions of $(\alpha_n)_{n\in\mathbb{N}}$, $(X_n)_{n\in\mathbb{N}}$ and $(W_n)_{n\in\mathbb{N}}$. Try to compute by hand an approximation of the relation found.

14. Bonus question: drawing on your previous answers, theorize about the existence of a limiting distribution for a Markov chain.

1.5 Back to coral reefs

Let us return to the framework of exercise 3.21: The warming of ocean waters leads to thermal stress, which significantly contributes to coral bleaching. Research indicates that a global temperature increase of 0.6°C could result in the extinction of all coral reefs. These ecosystems are vital to marine biodiversity, and their survival hinges on maintaining stable ocean temperatures.

Let us introduce a more optimistic scenario than the one initially envisaged: assume that there is a possibility of revolutionary scientific breakthroughs leading to a decrease of 0.2°C in a given decade. Thus, in each decade, the global average temperature could:

- ⋆ increase by 0.2°C with a probability of $p = 0.5$,
- ⋆ decrease by 0.2°C with a probability of $q = 0.2$,
- ⋆ remain constant.

The aim is to find the average lifespan of coral reefs.

14. Model this problem using Markov chains.

15. Use numerical computations to estimate the average lifespan of coral reefs.

16. Find again this value using theoretical computations.

1.6 To go further

Consider a Markov chain with $k \in \mathbb{N}^*$ states such that:

$$P = \frac{1}{20} \times \begin{pmatrix} 0 & 11 & 0 & \dots & \dots & 0 & 9 \\ 9 & 0 & 11 & 0 & & & 0 \\ 0 & 9 & 0 & 11 & 0 & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & & & \ddots & \ddots & 0 & 11 \\ 11 & 0 & \dots & \dots & 0 & 9 & 0 \end{pmatrix}_k .$$

17. Draw the graph of this chain.

Fix $k = 999$ and consider the initial distribution $X_0 = (x_i)_{1 \le i \le 999}$ such that $x_{500} = 1$ and $x_j = 0$ for $j \neq 500$.

18. Plot the distribution X_{200}. Find a good approximation of this distribution by a familiar probability distribution (give the values of the parameters that best approximate the distribution).

19. Now plot the distribution X_{201}, compare it with X_{200} and explain what is going on.

20. On the same plot, show the distributions $X_{200}, X_{250}, X_{300}, X_{350}$ and X_{400}.
Plot $E(X_n)$ and $\sigma(X_n)$ as a function of n for $1 \le n \le 500$.
Explain what will be the problem if we extend the plots beyond $n \ge 500$.

21. Plot the distributions X_{200000} and X_{200001} and explain what is happening.

1.6.1 Details on rendering

1.6.2 The rendering

The project will be carried out in teams of 5. You have until Sunday November 10 to register your teams in the dedicated section on Moodle. Students not assigned to a team after this date will be randomly assigned to groups by the teachers.

Each team is required to submit a report of no more than 20 pages. Grading will take into account presentation (use of LaTeX is encouraged), the quality of the curves presented, the conciseness of answers and interpretations. It is better to answer just one difficult question with a detailed, in-depth analysis, rather than giving superficial answers to all the questions.

Your report must be submitted via Moodle by 23h59 on Saturday, November 30. Any delay will be heavily sanctioned.

1.6.3 The oral presentation

For the oral presentation, a presentation with slides is expected. Pedagogy and clarity are the aim of this 15-minute oral. Please note that if you exceed the 15-minute presentation time, your teacher will stop you. Of course, slides must not include any handwritten or photographed text. All formulae and matrices must be typed. At the oral, you will be asked to give a precise account of the most important points in each section.

At the end of the presentation, there will be 15 minutes for questions: each student will be questioned individually on the project or on a point from the course.

Good luck with your project.