

Course1 : Introduction to Markov Chains

To model the evolution over time (the dynamic) of biological systems, such as that of an organism, substance, an ecosystem, often chosen for *random models*. The simplest of these models are random Markov chains, which in the particular case studied here are easy to use because there is virtually no mathematical prerequisites. They will give us the opportunity of a first familiarization with the matrix algebra that we will deepen in the following lessons.

1 An example in ecology

We are interested in developing of a natural forest in the temperate region of a fallow plot (for example, abandonment of cultivated area or following a fire). Our simplified model has 3 states. State 1 is that of a vegetation consisting of herbs or other pioneer species, the state 2 corresponds to the presence of shrubs whose rapid development requires maximum sunshine and the state 3 the largest trees that may develop in a semi sunny environment. If we denote by h, a, f these three states (for *grass, shrubs, forest*), all possible states for a given point of this plot is the set $S = \{h, a, f\}$. On the plot is ground mark a large number of points (one thousand) distributed on a regular grid and recorded at fixed time interval (every 3 years) the condition of vegetation in each of these points (by choosing one the three states which is the closest).

By observing the evolution during a time interval can be determined for each state $i \in \{h, a, f\}$ the proportion of points that are passed to the state $j \in \{h, a, f\}$, and note p_{ij} this proportion. If the different proportions and identified (there are 9) change a little the next time interval, it can be assumed unchanged over time and may be regarded as the probability for any point of moving from state i to state j for a period of time. For example, suppose in this plot, these probabilities are as follows:

$$\mathbb{P} = \begin{array}{ccc} & \begin{matrix} h & a & f \end{matrix} & \\ \begin{matrix} h \\ a \\ f \end{matrix} & \begin{pmatrix} 0,5 & 0,45 & 0,05 \\ 0,1 & 0,5 & 0,4 \\ 0 & 0,1 & 0,9 \end{pmatrix} & \begin{matrix} h \\ a \\ f \end{matrix} \end{array}$$

If X_0 is the condition of a point at time $t = 0$ and X_1 state the same point at $t = 1$, for example, the probability of transition from the state shrub $t = 0$ to the state forest at $t = 1$ is written $P(X_1 = f/X_0 = a)$ is worth $P(X_1 = f/X_0 = a) = 0,4$, Similarly $P(X_1 = h/X_0 = h) = 0,5$.

The set of states $S = \{h, a, f\}$ and transition matrix \mathbb{P} is an example of a *Markov chain*. One can also represent this Markov chain by the diagram in the following *points and arrows* (1).

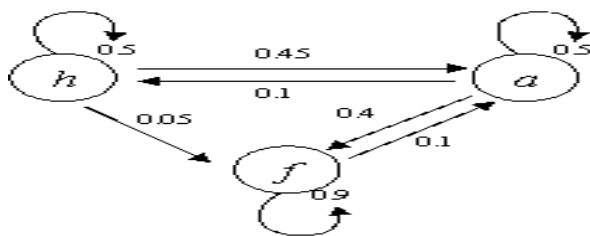


Figure 1: Diagram points and arrows corresponding to the example of the dynamics of natural forest with three states, *grass, shrubs* and *forest*, studied below. Note that a burning forest does not become grass but is replaced by shrubs.

In this diagram, each state is represented by a circled letter and each nonzero coefficient of the transition matrix \mathbb{P} by an arrow from one state to another state.

In this model, we can calculate the probability of any succession of states, called the *trajectory* of the Markov chain. For example the probability that a point on the plot, there is the succession of states (h, h, a, f, f) is calculated as follows:

$$\begin{aligned} P(X_0 = h, X_1 = h, X_2 = a, X_3 = f, X_4 = f) \\ &= \pi_0(h)P(X_1 = h/X_0 = h)P(X_2 = a/X_1 = h)P(X_3 = f/X_2 = a)P(X_4 = f/X_3 = f) \\ &= \pi_0(h)p_{hh}p_{ha}p_{af}p_{ff} = \pi_0(h)(0,5)(0,45)(0,4)(0,9) = 0,081\pi_0(h). \end{aligned}$$

where $\pi_0(h)$ is the probability of being in state h at the initial time $t = 0$.

2 Evolution of the distribution of states over time

The observation of the state is the different points of the plot at the initial time t_0 determines the initial proportions of each of the 3 states $\pi_0 = (\pi_0(h), \pi_0(a), \pi_0(f))$. This is done for each point is the state where it is and we calculate the proportion of points of each possible state. You can see each proportion as the probability for a point in the plot to be one of the states at the initial time. Thus, if one has for example $\pi_0 = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$, this means that half of the plot points are initially in the state h , fourth in the state a and the fourth in state f . But we can also interpret this considering that any state has 50% chance of being in state h , 25% of being in state a and 25% in the state f . This is why the proportion of individuals in the study population located in each state,

$$\begin{array}{c|c|c|c} S & h & a & f \\ \hline \pi_0 & \pi_0(h) & \pi_0(a) & \pi_0(f) \end{array}$$

is called the *law of initial probability* π_0 or the *initial distribution*. When choosing a mod'elisation by a Markov chain, the goal is often to determine the evolution of the distribution of states over time. For example, if the parcel considered above is covered for one third of forest at the initial time, this proportion will she grow up, move towards 100%, on the contrary tend to zero or approach a limit value, a kind of ecological balance?

We will see that if we know the initial distribution we can calculate the distribution at instant $t = 1$, then at time $t = 2$ and so on. In other words, we compute the law π_1

$$\begin{array}{c|c|c|c} S & h & a & f \\ \hline \pi_1 & \pi_1(h) & \pi_1(a) & \pi_1(f) \end{array}$$

and more generally the law π_t for all $t > 0$ and thus model the dynamics of this population. Here's how it is done. To calculate the probability π_1 of the three states at time $t = 1$

$$\pi_1 := (P(X_1 = h), P(X_1 = a), P(X_1 = f)) = (\pi_1(h), \pi_1(a), \pi_1(f)),$$

we note that, due to *the formula of total probability*¹, the probability is equal to $\pi_1(h)$

$$P(X_1 = h/X_0 = h)P(X_0 = h) + P(X_1 = h/X_0 = a)P(X_0 = a) + P(X_1 = h/X_0 = f)P(X_0 = f)$$

which can be written here $\pi_1(h) = 0,5 \cdot \pi_0(h) + 0,1 \cdot \pi_0(a) + 0 \cdot \pi_0(f)$ given the values of transition probabilities given by the first column of the matrix \mathbb{P} .

We deduce that $\pi_1(h)$ is the *dot product* of the vector π_0 with the first column of the matrix \mathbb{P} . Similarly, we verify that $\pi_1(a)$ is the scalar product of the vector π_0 with the second column of the matrix \mathbb{P} and $\pi_1(f)$ is the scalar product of the vector π_0 with troixime column of the matrix \mathbb{P} . We summarize this by saying that the vector π_1 is the product of the vector π_0 by the matrix \mathbb{P} , which can be written simply

$$\pi_1 = \pi_0 \cdot \mathbb{P}$$

or, more simply $\pi_1 = \pi_0 \mathbb{P}$, as if it were the product of two numbers (but here it is the product of a vector and a matrix). This very short means that the vector $\pi_1 = (\pi_1(h), \pi_1(a), \pi_1(f))$ is the product of the vector $\pi_0 = (\pi_0(h), \pi_0(a), \pi_0(f))$ by the matrix \mathbb{P} , which can also be written in matrix form:

$$(\pi_1(h), \pi_1(a), \pi_1(f)) = (\pi_0(h), \pi_0(a), \pi_0(f)) \begin{pmatrix} 0,5 & 0,45 & 0,05 \\ 0,1 & 0,5 & 0,4 \\ 0 & 0,1 & 0,9 \end{pmatrix}.$$

¹If $\Omega = B_1 \cup B_2$ et $B_1 \cap B_2 = \emptyset$, on a $\mathbb{P}(A) = \mathbb{P}(A/B_1)\mathbb{P}(B_1) + \mathbb{P}(A/B_2)\mathbb{P}(B_2)$. And more generally, if $\Omega = B_1 \cup \dots \cup B_n$ and $B_i \cap B_j = \emptyset$ if $i \neq j$, on a $\mathbb{P}(A) = \mathbb{P}(A/B_1)\mathbb{P}(B_1) + \dots + \mathbb{P}(A/B_n)\mathbb{P}(B_n)$.

3 General Overview

The example of the dynamics of a natural forest in temperate areas that we studied in the previous two paragraphs is a special case of *Markov chain*. More generally, is modeled by a Markov chain evolution over time of quantities X that can take a finite number of states $X = x_1, X = x_2, \dots, X = x_n$ and passing the state x_i at time t to state x_j at the next time $t + 1$ with probability p_{ij} given. The numbers $p_{ij} = P(X_{t+1} = x_j / X_t = x_i)$ then check all the inequalities $0 \leq p_{ij} \leq 1$ and their sum is 1, $\sum_{j=0}^n p_{ij} = 1$ (as if the chain is in state x_i at a time, it will necessarily be in one of the possible states x_1, \dots, x_n the next moment and then $p_{i1} + p_{i2} + \dots + p_{in} = 1$). The term $P(X_{t+1} = x_j / X_t = x_i)$ is called a *conditional probability* and is the “probability that the quantity x_j worth X at time $t + 1$ knowing that it was worth x_i at time t ”.

To define a Markov chain, so you have two basic ingredients:

1. The state space $S := \{x_1, \dots, x_n\}$ is assumed that the finished
2. The *transition matrix*² (or transition)

$$\mathbb{P} = (p_{ij})_{1 \leq i \leq n, 1 \leq j \leq n} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ p_{11} & p_{12} & \dots & p_{1n} \\ \vdots & & & \vdots \\ p_{n1} & p_{n2} & \dots & p_{nn} \end{pmatrix} \begin{matrix} x_1 \\ \vdots \\ x_n \end{matrix}$$

Note that this matrix is called *stochastic matrix* because its coefficients are all between 0 and 1 and the sum of the coefficients of each line is 1 (which is not generally true for columns).

Called law of *initial distribution* or *initial probability* vector π_0

$$\begin{array}{c|ccc|c} S & x_1 & x_2 & \dots & x_n \\ \hline \pi_0 & \pi_0(x_1) = P(X_0 = x_1) & \pi_0(x_2) = P(X_0 = x_2) & \dots & \pi_0(x_n) = P(X_0 = x_n) \end{array}$$

which each component represents the probability of being in one of the states at the initial time. The sum of the components of this vector of probabilities is equal to 1 because for sure be in one of the 'possible states. Knowledge of the transition matrix \mathbb{P} can calculate the evolution over time of the initial distribution. For this, we simply the product of the vector by this matrix π_0 , once for distribution π_1 at time $t = 1$, then again to the know at time $t = 2 \dots$ etc.

4 Relevance of Markov models

Finally add some comments regarding the relevance of modeling by Markov chains. Markov chains models indeed provide very useful, especially in biology but they also, of course, defects. Among them, the simplifying assumption of a **finite number of possible states** is relatively easy to circumvent (where necessary) as there are Markov chains with an infinite number of states. Similarly one can overcome the assumption of **invariance in time of transition probabilities** (in fact it was assumed that the probability of moving from one state to a state x_i x_j did not change over time) Whereas in Markov chains which are not homogeneous, that is to say having transition matrices modified over time. Of course the study of these models gnralis'es (Markov chains or infinite non-homogeneous) require the use of more sophisticated mathematical tools.

By cons, there is another defect, it is a real limitation of the model: it is the **spatial invariance**. For the calculation of transition probabilities, it was indeed an implicit assumption of spatial homogeneity is rarely satisfied in practice. For example a plant site was certainly not the same probability of being employed next period by a species using the adjacent sites already are or they are not. And unfortunately this can not be taken into account by the Markov models we have presented here. We will therefore should be used with caution when a Markov chain character of isotropic environment is not at all satisfied.

²It is customary to note the coefficients of a matrix (p_{ij}) , where i is the row index and j the column index, so the number noted in Figure p_{ij} **i**-th **Line** and **j**-th **Column** ; hold “**LiCo**”.

5 Lectures

For more information, it is strongly recommended to read the Wikipedia article on Markov chains at:
http://en.wikipedia.org/wiki/Markov_chain
(including the delightful history of Doudou hamsters!).