CHAPTER 13

FINITE MARKOV CHAINS

The theory of Markov chains is used in

- modelling and simulation,
- queueing theory: for instance, the study of the average waiting time for accessing the nerve-centres of a telematic network,
- robotics, to model the moves of the robot depending on its environment,
- signal theory.

The theory of Markov chains is based on conditional probabilities. We mainly study finite Markov chains whose underlying sample probability space is finite.

In this chapter, we define finite Markov chains, their transition matrices and their graphs. We then show how the graph of a finite Markov chain can be used to study its properties.

We recommend in the strongest possible terms the following handbook:

William Feller, *Probability Theory*, Vol. 1, John Wiley, New York (1968).

We also recommend:

Dean Isaacson, Richard Madsen, *Markov Chains Theory and Applications*, John Wiley, New York (1976).

13.1 Introduction

Up to now we have mainly studied independent experiments. For instance, a sequence of trials e_n that independently produce as results elements of the same trial space Ω yields a sequence of independent random variables X_n; we then have, by the independence, that

$$P\big((X_0,\ldots,X_n)=(\omega_0,\ldots,\omega_n)\big)=P(X_0=\omega_0)\cdots P(X_n=\omega_n)=p_0\cdots p_n$$

(see the binomial distribution corresponding to a coin-tossing game).

In the present chapter we will study the simplest possible generalization of this notion in which the result of the nth trial, $X_n = \omega$, no longer has a fixed probability $p(\omega)$, independent of the trials $e_0, \dots, e_{n-1}$, but, rather, a conditional probability, completely determined by the result of the $(n-1)$th trial e_{n-1}. We associate with each pair ω_i, ω_j the probability p_{ij}, which represents the probability that $X_n = \omega_j$ given that $X_{n-1} = \omega_i$, and we assume that this probability is the same for all possible ns. We will thus have the following, assuming for instance that at the initial moment $P(X_0 = \omega_j) = q_j$:

$$P\big((X_0, X_1) = (\omega_i, \omega_j)\big) = q_i p_{ij} \ ,$$
$$P\big((X_0, X_1, X_2) = (\omega_i, \omega_j, \omega_k)\big) = q_i p_{ij} p_{jk} \ .$$

The situation is often summed up by saying, inaccurately, that: 'the result of the nth trial depends only on the result of the $(n-1)$th trial'. Actually, the nth trial depends (via the $(n-1)$th trial) on the $(n-2)$th trial, which in turn depends on the $(n-3)$th one, etc.

Indeed, it is more accurate, but also more cumbersome, to say that the whole past of the sequence of trials can be coded, for conditioning its future evolution, in the knowledge of its present state. Formally, we have

$$P(X_{n+1} = \omega_{n+1} \,/\, X_n = \omega_n, \dots, X_0 = \omega_0) = P(X_{n+1} = \omega_{n+1} \,/\, X_n = \omega_n) \ ,$$

with, moreover, for all $k > 0$,

$$P(X_{n+k+1} = \omega \,/\, X_{n+k} = \omega') = P(X_{n+1} = \omega \,/\, X_n = \omega').$$

13.2 Generalities

13.2.1 Definitions

In this chapter we will adopt the usual notations for Markov chains, even though they are slightly different from the notations used in Chapter 12.

The trial space $\Omega = \{E_1, \dots, E_r\}$ is finite, and its elements $E_1, \dots, E_r$ are called the *states* of the system; we will abbreviate $E_1, \dots, E_r$ to $1, \dots, r$ when no ambiguity can occur. A sequence $(X_n)_{n \in \mathbb{N}}$ of random variables will be represented by a sequence of states of Ω, the nth state of the sequence representing the value of the random variable (X_n). The result x_n of the nth trial (or the value x_n of the nth random variable) will be called the state of the system at time n. $(X_n)_{n \in \mathbb{N}}$ is a Markov chain if the conditional probabilities at time n,

$$p_{ij} = P(X_{n+1} = E_j \,/\, X_n = E_i) \ ,$$

do not depend on n. The formal definition follows.

Definition 13.1 *A sequence of random variables (abbreviated r.v.'s)* $(X_n)_{n\in\mathbb{N}}$ *with values ranging over a finite set* $\Omega = \{E_1, \dots, E_r\}$ *is a Markov chain** *if and only if it satisfies the following equivalent conditions:*

(i) $P(X_{n+1} = E_{i_{n+1}} / X_0 = E_{i_0}, ..., X_n = E_{i_n}) = P(X_{n+1} = E_{i_{n+1}} / X_n = E_{i_n})$ *and*

$$\forall k > 0 \quad \forall n > 0 \quad P(X_{n+k+1} = E / X_{n+k} = E') = P(X_{n+1} = E / X_n = E').$$

(ii) $P(X_0 = E_{i_0}, X_1 = E_{i_1}, \dots, X_n = E_{i_n}) = q_{i_0} p_{i_0 i_1} p_{i_1 i_2} p_{i_{n-1} i_n}$, *where* $\forall i$, $q_i = P(X_0 = E_i)$ *is the initial probability distribution, and* $\forall i, j, n$, $p_{ij} = P(X_{n+1} = E_j / X_n = E_i)$ *is the conditional probability of obtaining the result* E_j *given that the preceding result is* E_i.

By Definition 13.1 we will thus have

1. $\forall i, \quad q_i \geq 0$ and $\sum_{i=1}^{r} q_i = 1$,
2. $\forall i \forall j, \quad p_{ij} \geq 0$ and $\sum_{k=1}^{r} p_{ik} = 1$.

Exercise 13.1 Show that, conversely, the q_is and the p_{ij}s satisfying 1 and 2 above, with $1 \leq i, j \leq r$, each define a Markov chain. ♢

13.2.2 Examples

We will see an example, a counterexample, and a model equivalent to the notion of Markov chain.

Example 13.2 Message passing. Consider passing a message 'yes' or 'no' in a population. Each individual forwards the message received with probability p and the opposite message with probability $1 - p$. Let X_n be the message forwarded by the nth individual. We have:

$$P(X_{n+1} = \text{'yes'} / X_n = \text{'yes'}) = P(X_{n+1} = \text{'no'} / X_n = \text{'no'}) = p,$$
$$P(X_{n+1} = \text{'yes'} / X_n = \text{'no'}) = P(X_{n+1} = \text{'no'} / X_n = \text{'yes'}) = 1 - p.$$

---

* Here we do *not* use the usual terminology, where a Markov chain is defined by the single property:

$$P(X_{n+1} = E_{i_{n+1}} / X_0 = E_{i_0}, \dots, X_n = E_{i_n}) = P(X_{n+1} = E_{i_{n+1}} / X_n = E_{i_n}) = p_{i_n i_{n+1}}(n)\ ,$$

i.e. $P(X_{n+1} = E_{i_{n+1}} / X_n = E_{i_n})$ can depend on n. The chains that we will consider are in fact the special case of the general Markov chains, when $P(X_{n+1} = E_{i_{n+1}} / X_n = E_{i_n}) = p_{ij}$ *does not depend* on n. Such chains are usually called *homogeneous Markov chains.*

These four equalities allow us, if we are given the probability distribution of the initial message, to fully describe the message passing. We see that only the message X_n forwarded by the nth individual affects X_{n+1}, and that it is not useful to memorize the whole preceding history of the message. The $(X_n)_{n\in\mathbb{N}}$s form a Markov chain, with $\Omega = \{\text{'yes'},\text{'no'}\}$. Let $E_1 = \{\text{'yes'}\}$ and $E_2 = \{\text{'no'}\}$; then $p_{11} = p$, $p_{12} = 1-p$, $p_{21} = 1-p$, $p_{22} = p$.

EXERCISE 13.2 The notations are as in Example 13.2. Assume, moreover, that $q_1 = P(X_0 = \text{'yes'}) = q_2 = P(X_0 = \text{'no'}) = 1/2$.

1. Show by induction on n that all the r.v.'s X_n have the same probability distribution as X_0.
2. Are the r.v.'s X_n independent? ♢

EXAMPLE 13.3 We now exhibit an example of a non-Markovian chain: the Polya urn model. An urn contains b black balls and r red balls. A ball is drawn at random. It is replaced and, moreover, c balls of the colour drawn are added. A new random drawing is performed and the whole process is iterated. Let $X^n \in \{B, R\} = \Omega$ be the colour of the ball drawn at the nth drawing. The sequence $(X^n)_{n\geq 1}$ is not a Markov chain. We thus have:

$$P(X^3 = B \,/\, X^2 = B) = \frac{b+c}{b+r+c}\,,$$

$$P(X^3 = B \,/\, X^2 = B, X^1 = B) = \frac{b+2c}{b+r+2c}\,.$$

EXERCISE 13.3 Return to the Polya urn model of Example 13.3 and recall that $P(X^n = B) = \big(b/(b+r)\big)$ for all n (see Exercise 12.5).

1. For what values of c is the sequence $(X^n)_{n\in\mathbb{N}}$ defined in Example 13.3 a Markov chain?
2. Let Y_n be the r.v. giving the number of black balls in the urn at time n. Is the sequence $(Y_n)_{n\in\mathbb{N}}$ a Markov chain? ♢

EXAMPLE 13.4 Markov chains as urn models. Any Markov chain can be represented as an urn model as follows: if $\Omega = \{E_1, \ldots, E_r\}$, r urns are available; each urn contains a fixed number of balls marked $E_1, \ldots, E_r$. In the jth urn, the probability of drawing a ball marked E_k is p_{jk}. At the initial trial, an urn is chosen according to the probability distribution q_j. From that chosen urn, a ball is drawn at random, and if it is marked E_j, the next drawing is made from the jth urn, and so on. The sequence X_n of the drawn markings is a Markov chain. We thus see that Markov chains can be modelled by urns.

13.2.3 Transition matrices

A Markov chain is thus characterized by

- on the one hand, the conditional probabilities p_{ij}, $1 \leq i, j \leq r$, where p_{ij} is the probability of a state i given that the state j occurred at the preceding trial, and is called the *transition probability* from i to j and
- on the other hand, the initial probability distributions q_i, $1 \leq i \leq r$.

The p_{ij}s form the *transition matrix* P

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1r} \\ p_{21} & p_{22} & \cdots & p_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ p_{r1} & p_{r2} & \cdots & p_{rr} \end{pmatrix}$$

and verify

$$\forall i, j, \quad p_{ij} \geq 0 \text{ and } \forall i, \quad \sum_{k=1}^{r} p_{ik} = 1. \tag{13.1}$$

A matrix verifying (13.1) is called a *stochastic matrix*.

Exercise 13.4 Sentry.
Assume that a sentry watches over a square stronghold having four turrets in the following way: he starts at random by one of the turrets and after each five-minute interval tosses a coin and goes to the first turret on his left (if TAILS turns up) or to the first turret on his right (if HEADS turns up).

1. Formalize the problem.
2. Let X_n be the number of the turret chosen as the nth watchtower, $n = 0, 1, \ldots$. Show that X_n is a Markov chain. What is the transition matrix? ♢

Exercise 13.5 Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain with transition matrix P. Let

$$L_n = \begin{pmatrix} P(X_n = 1) \\ \vdots \\ P(X_n = i) \\ \vdots \\ P(X_n = r) \end{pmatrix}$$

be the column describing the probability distribution of X_n.

1. Express L_{n+1} in terms of P and L_n.
2. Deduce L_n in terms of P and of the probability distribution L_0 of X_0.
3. Give a necessary and sufficient condition ensuring that all the X_ns have the same probability distribution. ♢

EXERCISE 13.6 Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain.

1. Show that the following properties are equivalent:
   - (i) There exists a k such that X_k and X_{k+1} are independent.
   - (ii) The columns of P are constant.
   - (iii) For each k, X_k and X_{k+1} are independent.
   - (iv) $\forall n$, $(X_0, \dots, X_n)$ are independent.
2. What is the distribution of X_n when the conditions of 1 are satisfied? ♢

13.2.4 Properties

In the sequel, X_n will denote a Markov chain with values in $\Omega = \{E_1, \dots, E_r\}$. Let $q_i^{(n)} = P(X_n = E_i)$ be the (unconditional) probability that the chain is in state i at time n, and let

$$p_{ij}^{(n)} = P(X_n = E_j \,/\, X_0 = E_i)$$

be the conditional probability of a transition from E_i to E_j in exactly n steps.

Let us first state five simple lemmata about conditional probabilities that will be quite useful when computing with Markov chains.

Lemma 13.5 $P(A \cap B \,/\, C) = P(A \,/\, B \cap C) \times P(B \,/\, C)$.

Lemma 13.6 $P(A \cap B \,/\, A) = P(B \,/\, A)$.

Lemma 13.7 $q_i^{(n)} = \sum_{j=1}^{r} q_j p_{ji}^{(n)}\ , \quad \forall i = 1, \dots, r.$

Lemma 13.8
(i) $\forall k \geq 0, \quad \forall i, j = 1, \dots, r\,, \quad p_{ij}^{(n)} = P(X_{n+k} = E_j \,/\, X_k = E_i)\ .$
(ii) $\forall n \geq 0, \quad \forall i, j = 1, \dots, r$, *the* $p_{ij}^{(n)}$ *can be computed recursively by*

$$p_{ij}^{(1)} = p_{ij}$$
$$p_{ij}^{(n+1)} = \sum_{k=1}^{r} p_{ik}\, p_{kj}^{(n)}\ .$$

(iii) *The* $p_{ij}^{(n)}$ *are the coefficients of the matrix* P^n *(the transition matrix* P *multiplied by itself* n *times).*

REMARK 13.9 All the matrices P^n are stochastic matrices.

Lemma 13.10 (Markov property)

$$\begin{aligned}P(X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k} = E_{i_{n+k}} \,/\, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}) \\ = p_{i_n i_{n+1}} \cdots p_{i_{n+k-1} i_{n+k}} \\ = P(X_1 = E_{i_{n+1}}, \ldots, X_k = E_{i_{n+k}} \,/\, X_0 = E_{i_n}).\end{aligned}$$

Proof. By induction on k.

1. If $k = 1$, it is simply the definition of Markov chains.
2. Assume that the result holds for $k \leq k_0$. Let

$$B = (X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k_0} = E_{i_{n+k_0}})$$

and $A = X_{n+k_0+1} = E_{i_{n+k_0+1}}$. We then have, by Lemma 13.5,

$$\begin{aligned}P(X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k_0} = E_{i_{n+k_0}}, X_{n+k_0+1} = E_{i_{n+k_0+1}} \\ / \, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}) \\ = P(X_{n+k_0+1} = E_{i_{n+k_0+1}} \\ / \, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}, X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k_0} = E_{i_{n+k_0}}) \\ \times P(X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k_0} = E_{i_{n+k_0}} \,/\, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}).\end{aligned}$$

We thus obtain, applying the induction hypothesis once with $k = 1$ and once with $k = k_0$,

$$\begin{aligned}P(X_{n+1} = E_{i_{n+1}}, \ldots, X_{n+k_0} = E_{i_{n+k_0}}, X_{n+k_0+1} = E_{i_{n+k_0+1}} \\ / \, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}) \\ = p_{i_{n+k_0} i_{n+k_0+1}} \times p_{i_n i_{n+1}} \cdots p_{i_{n+k_0-1} i_{n+k_0}} \\ = P(X_1 = E_{i_{n+k_0+1}} \,/\, X_0 = E_{i_{n+k_0}}) \\ \times P(X_1 = E_{i_{n+1}}, \ldots, X_k = E_{i_{n+k}} \,/\, X_0 = E_{i_n}).\end{aligned}$$

The same computation with $n = 0$ shows that

$$\begin{aligned}P(X_1 = E_{i_{n+1}}, \ldots, X_{k_0} = E_{i_{n+k_0}}, X_{k_0+1} = E_{i_{n+k_0+1}} \,/\, X_0 = E_{i_0}) \\ = p_{i_{n+k_0} i_{n+k_0+1}} \times p_{i_n i_{n+1}} \cdots p_{i_{n+k_0-1} i_{n+k_0}} \\ = P(X_1 = E_{i_{n+k_0+1}} \,/\, X_0 = E_{i_{n+k_0}}) \\ \times P(X_1 = E_{i_{n+1}}, \ldots, X_k = E_{i_{n+k_0}} \,/\, X_0 = E_{i_n}),\end{aligned}$$

and hence the inductive step and the result. □

Lemma 13.10 states that the probability that a Markov chain which went through the states $E_{i_0}, \ldots, E_{i_n}$ goes on through the states $E_{i_{n+1}}, \ldots, E_{i_{n+k}}$ is the same as the probability that a chain starting at time 0 from state E_{i_n} then goes through the states $E_{i_{n+1}}, \ldots, E_{i_{n+k}}$.

Lemma 13.10 has several immediate consequences for expressing properties of Markov chains.

Corollary 13.11 *Let E be a state of a Markov chain. Then*

$$P(X_{n+1} \neq E, \ldots, X_{n+k} \neq E, X_{n+k+1} = E \,/\, X_0 = E_{i_0}, \ldots, X_n = E_{i_n}) \\ = P(X_1 \neq E, \ldots, X_k \neq E, X_{k+1} = E \,/\, X_0 = E_{i_n}).$$

Exercise 13.7
1. Prove Corollary 13.11.
2. Show that, similarly: $P(X_{n+1} \in A_1, \ldots, X_{n+k} \in A_k \,/\, X_0 = E_{i_0}, \ldots, X_n = E_{i_k}) = P(X_1 \in A_1, \ldots, X_k \in A_k \,/\, X_0 = E_{i_k})$.
3. Show, finally: $P(X_{n+1} \in A_1, \ldots, X_{n+k} \in A_k \,/\, X_0 \in A'_0, \ldots, X_{n-1} \in A'_{n-1}, X_n = E_{i_k}) = P(X_1 \in A_1, \ldots, X_k \in A_k \,/\, X_0 = E_{i_k})$. This last equality can be called the generalized Markov property. ♢

13.3 Classification of states

Many properties of Markov chains are intrinsic: this means that they depend only on the transition probabilities (i.e. on the transition matrix), and they do not depend on the starting point of the chain (i.e. on the initial probability distribution). This is true for the properties studied in the present section.

13.3.1 Irreducible chains

We shall say that state E_j can be reached from state E_i if and only if there exists an n such that $p_{ij}^{(n)} > 0$. In other words, if there is a strictly positive probability of reaching E_j from E_i.

Definition 13.12 *A non-empty set of states C is said to be closed if no state outside C can be reached from any state E_i in C, i.e. C is closed if and only if $\forall E_i \in C$ and $\forall E_j \notin C$, $p_{ij} = 0$.*

If the singleton $C = \{E_i\}$ is closed, the state E_i is said to be absorbing. A Markov chain is irreducible if there exists no closed set other than the set of all states.

Lemma 13.13 *If C is closed, then $\forall n$, $p_{ij}^{(n)} = 0$.*

Proof. Straightforward by induction on n. □

Lemma 13.14 *Let C be a closed set; let P_C (resp. P_C^n) be the matrix deduced from P (resp. P^n) by deleting from P (resp. P^n) all rows and columns corresponding to the states $E_j \notin C$. Then*
1. *for all n, $P_C^n = (P_C)^n$ and*
2. *the sequence P_C^n is a sequence of stochastic matrices.*

This lemma intuitively means that we have a Markov chain on C which can be studied *per se* via the matrix P_C^n, i.e. independently of the states outside C.

Proposition 13.15 *A chain is irreducible if and only if any state can be reached from any other state.*

Proof. The 'if' part is straightforward because if there were a closed set $C \subsetneq \Omega$, then the states of the complement $\overline{C} = \Omega - C$ of C would not be reachable from the states of C.

As for the 'only if' part, we reason by contradiction and assume that there exist states E_i and E_j such that E_j is not reachable from E_i. Then, the least closed set C containing E_i cannot contain E_j, and thus C is a closed set strictly contained in the set of all states, $C \subsetneq \Omega$, a contradiction. □

13.3.2 Classification of states

A state E_i is said to be *persistent* (or *recurrent*) if the probability that the system starting from E_i eventually returns to E_i is equal to 1. Otherwise, it is said to be *transient*.

For all i, j in $\{1, \ldots, r\}$, let $f_{ij}^{(n)} = P(A_{ij}^{(n)})$, where $A_{ij}^{(n)}$ = 'the system starting from E_i reaches E_j for the first time after exactly n steps'. Let $f_{ij} = \sum_{n=1}^{\infty} f_{ij}^{(n)}$ (assuming $f_{ij}^{(0)} = 0$). Then f_{ij} is the probability that the system starting from E_i eventually reaches E_j. We thus have

$$E_i \text{ persistent} \iff f_{ii} = 1$$

and

$$E_i \text{ transient} \iff f_{ii} < 1\,.$$

If $f_{ii} = 1$, then the $f_{ii}^{(n)}$s for $n \in \mathbb{N}$ form a probability distribution, and we can define

$$\mu_i = \sum_{n=1}^{\infty} n f_{ii}^{(n)}\,,$$

which represents the average waiting time in order for a system starting from E_i to come back to E_i (i.e. the mean of the r.v. T = number of the first return of the chain to state E_i).

Exercise 13.8 We can compute the $f_{ij}^{(n)}$s recursively by the recurrence relation

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)}\,. \qquad \diamondsuit$$

The definition and computation suggested in the above exercise are not very simple. Fortunately, we have, for the case of finite Markov chains that are of interest to us, a simple characterization of transient states, from which we will also deduce a characterization of persistent states by noticing that any state is either transient or persistent.

Proposition 13.16 *A state E_i of a Markov chain is absorbing if and only if it satisfies either one the following equivalent conditions:*

(i) $p_{ij} = \begin{cases} 1 & \text{if } j = i, \\ 0 & \text{if } j \neq i. \end{cases}$

(ii) $P(X_n = E_i\ ,\ \forall n\ /\ X_0 = E_i) = 1$, *where* $(X_n = E_i\ ,\ \forall n)$ *denotes the event* $X_0 = X_1 = \cdots = X_n = \cdots = E_i$.

EXERCISE 13.9 Prove this result. ♢

EXERCISE 13.10 Prove that any absorbing state is persistent. ♢

The following results are immediate consequences of Proposition 13.29; we will not prove them here.

Proposition 13.17 *E_i is transient if and only if there exists an E_j such that E_j can be reached from E_i and E_i cannot be reached from E_j.*

Two states of a chain are *of the same type* if and only if either they are both transient or they are both persistent. We have:

Theorem 13.18 *All states of an irreducible chain are of the same type.*

Theorem 13.19 *If E_i is persistent, there exists a unique closed and irreducible set $C(E_i)$ containing E_i such that for any E_j, E_k in $C(E_i)$, $f_{jk} = f_{kj} = 1$. $C(E_i)$ is called the class of the persistent state E_i.*

Thus, the system, starting from any state of $C(E_i)$, will eventually reach another state of $C(E_i)$ and will never get out of $C(E_i)$.

Corollary 13.20 *A transient state cannot be reached from a persistent state.*

From now on, we will assume a technical restriction enabling us to give a simple classification of the states of a Markov chain: we will assume that all states are aperiodic, i.e. that there exists no integer $k > 1$ such that $p_{ii}^{(n)} \neq 0$ if and only if n is a multiple of k. We can then characterize transient (resp. persistent) states by the following so-called ergodicity conditions. (This terminology stems from the fact that an aperiodic persistent state of a finite Markov chain is said to be ergodic.)

Theorem 13.21 *Let X_n be a Markov chain all of whose states are aperiodic:*

(i) *E_i is transient if and only if*

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$$

and in that case, $\forall j\,,\ \sum_{n=1}^{\infty} p_{ji}^{(n)} < \infty$.

(ii) *E_i is persistent (or ergodic), if and only if*

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$$

and in that case $\mu_i < \infty$*; moreover, for all* j ,

$$\sum_{n=0}^{\infty} p_{ij}^{(n)} = \begin{cases} \infty & \text{if } j \in C(E_i), \\ 0 & \text{if } j \notin C(E_i), \end{cases}$$

$$\text{and} \qquad \lim_{n\to\infty} p_{ji}^{(n)} = f_{ji}\mu_i^{-1}\,.$$

The proof of this theorem is too complex to be given here, but it has many useful consequences.

Definition 13.22 *The potential matrix of a Markov chain is the matrix* $U = (u_{ij})$ *defined by* $u_{ij} = \sum_{n=0}^{\infty} p_{ij}^{(n)}$, *where* $U = (1-P)^{-1}$ *is the inverse of the matrix* $1 - P$.

Corollary 13.23 *A state* i *of a Markov chain is persistent if and only if* $u_{ii} = \infty$.

Proposition 13.24 *Given a Markov chain* X_n, *we define the random variable* $N_i = \sum_{n=0}^{\infty} \delta_{ni}$, *where*

$$\delta_{ni} = \begin{cases} 1 & \text{if } X_n = i, \\ 0 & \text{if } X_n \neq i. \end{cases}$$

N_i represents the number of occurrences of state i from time 0 (inclusive) on:

(i) *If i is persistent,* $P(N_i = \infty \ / \ X_0 = i) = 1$.

(ii) *If i is transient,* $P(N_i < \infty \,/\, X_0 = i) = 1$*; N_i then has a geometric distribution and*

$$\forall k, \quad P(N_i = k \,/\, X_0 = i) = (1 - f_{ii}) f_{ii}^{k-1}\,.$$

(iii) *If i and j are transient, and $i \neq j$,* $P(N_j < \infty \ / \ X_0 = i) = 1$*; N_j then has the distribution*

$$\forall k, \quad P(N_j = k \,/\, X_0 = i) = f_{ij}(1 - f_{jj}) f_{jj}^{k-1}\,.$$

Proof. We first show (iii).

$$P(N_j = k \,/\, X_0 = i) = P(N_j = k, \sum_m A_m \,/\, X_0 = i)\ ,$$

where $A_m = (X_m = j, X_{m-1} \neq j, \ldots, X_1 \neq j, X_0 = i)$ is the event 'the system, starting from the initial state i, visits state j for the first time at the mth step'. We have the following, noting that the A_ms form a partition for $m \in \mathbb{N}$:

$$\begin{aligned}
P(N_j = k \,/\, X_0 = i) &= P(N_j = k, \textstyle\sum_m A_m \,/\, X_0 = i) \\
&= \sum_m P(N_j = k, A_m \,/\, X_0 = i) \\
&= \sum_m P(N_j = k, A_m \,/\, A_m) \times P(A_m \,/\, X_0 = i) \qquad (13.2) \\
&= \sum_m P(A_m \,/\, X_0 = i) \times P(N_j = k \,/\, A_m) \qquad (13.3) \\
&= \sum_m f_{ij}^{(m)} \times P(N_j = k \qquad (13.4) \\
&\qquad / \, X_m = j, X_{m-1} \neq j, \ldots, X_1 \neq j, X_0 = i) \\
&= \sum_m f_{ij}^{(m)} \times P(N_j = k \,/\, X_0 = j) \qquad (13.5) \\
&= P(N_j = k \,/\, X_0 = j) \times \sum_m f_{ij}^{(m)} \\
&= P(N_j = k \,/\, X_0 = j) \times f_{ij}\ , \qquad (13.6)
\end{aligned}$$

where
(13.2) follows from Lemma 13.5 applied with $A = A_m$, $B = (N_j = k)$ and $C = (X_0 = i)$,
(13.3) follows from Lemma 13.6 applied with $A = A_m$ and $B = (N_j = k)$,
(13.4) follows from the definition of the $f_{ij}^{(m)}$s and
(13.5) follows from the generalized Markov property proved in Exercise 13.7.

A similar argument shows that

$$\begin{aligned}
P(N_j = k \,/\, X_0 = j) &= f_{jj} \times P(N_j = k-1, X_0 = j \,/\, X_0 = j) \\
&= f_{jj} \times P(N_j = k-1 \,/\, X_0 = j) \\
&= f_{jj}^{k-1} \times P(N_j = 1 \,/\, X_0 = j) \\
&= f_{jj}^{k-1} \times (1 - f_{jj})
\end{aligned}$$

by noting that $P(N_j = 1, X_0 = j \,/\, X_0 = j) = 1 - f_{jj}$ is the probability that the chain starting from the initial state i never comes back to that state. The probability distributions given in (ii) and (iii) follow immediately.

We then check that

$$P(N_i < \infty \,/\, X_0 = i) = \sum_{k \geq 1} P(N_i = k \,/\, X_0 = i) = \sum_{k \geq 1} f_{ii}^{k-1} \times (1 - f_{ii}) = 1\,. \quad \square$$

EXERCISE 13.11 Show the following properties:
1. $u_{ij} = E(N_j \,/\, X_0 = i)$. For $i \neq j$, and i and j transient, u_{ij} thus represents the average number of occurrences of j for a chain starting from the initial state i.
2. $u_{ij} = f_{ij}u_{jj}$.
3. If j is transient, then $\forall i$, $E(N_j \,/\, X_0 = i) < \infty$. ♢

Corollary 13.25 *The states of a Markov chain can be uniquely partitioned into*

$$\Omega = T \cup C_1 \cup \cdots \cup C_k \ ,$$

such that T consists of all the transient states, and each C_i is an irreducible closed set of persistent states. If E_j is in C_i, then $\forall E_k \in C_i$, $f_{jk} = 1$ and $\forall E_k \notin C_i$, $f_{jk} = 0$. Moreover, any finite Markov chain has at least one persistent state, and it is impossible that $\forall E_j \in C_i$, $\mu_j < \infty$.

Proposition 13.26 *Let E_i be a state of a Markov chain. The least irreducible closed set $C(E_i)$ containing E_i is*

$$C(E_i) = \{E_j \ / \ u_{ij} > 0\}.$$

Consequently, a chain is irreducible if and only if $\forall E_i, E_j$, $u_{ij} > 0$; moreover, for finite chains, we will have $\forall E_i, E_j$, $u_{ij} = \infty$.

The asset of the preceding characterizations and classifications is that they are easily generalized to infinite Markov chains. Their liability is that they are difficult to use. For finite Markov chains, there is a much simpler characterization of transient and persistent states, via a graph associated with the Markov chain. We must first recall some basic notions about graphs (see Chapter 10 for more details).

13.3.3 Graph of a finite Markov chain

Strongly connected graphs were defined in Chapter 10, Section 10.1.7. It is easy to see that a graph is *strongly connected* if it satisfies the following equivalent conditions:

(i) Any two vertices x and y are on some circuit.
(ii) For any two vertices x and y, there exist both a path with origin x and target y and a path with origin y and target x.

On the vertices of a graph, we can define an equivalence relation $x \equiv y$ if and only if

- either $x = y$,
- or x and y are on the same circuit.

We check that $\equiv$ is indeed an equivalence relation, where the equivalence classes modulo $\equiv$ form a partition of the set of vertices of the graph and are called the strongly connected components of the graph. Note that the equivalence classes modulo $\equiv$ are the maximal strongly connected subgraphs of the graph, and hence are called the strongly connected components. With each graph G we can associate a *reduced graph*, whose vertices are the strongly connected components of G and whose edges are defined as follows: there is an edge whose origin is the strongly connected component C and whose target is the strongly connected component C' if and only if there exists a vertex $x \in C$ and a vertex $x' \in C'$ such that (x, x') is an edge of G (i.e. there is at least one edge going from a vertex of C to a vertex of C'). The reduced graph is, by construction, a graph without circuits.

Proposition 13.27 *Define on a graph without circuits the following relation: $x < y$ if and only if there exists a path going from x to y. The relation $<$ is an ordering.*

We now have all the tools needed for studying the classification of Markov chains by means of graphs.

Definition 13.28 (Graph of a Markov chain) *The transition matrix of a Markov chain is represented by a directed graph $G = (S, A)$, where S is a finite set of vertices and A is the 'edge' relation on S. G is equipped with a labelling $l: A \longrightarrow]0, 1]$, and is defined as follows:*

- $S = \{E_1, \ldots, E_r\} = \{1, \ldots, r\} = \Omega$.
- $(i, j) \in A$ *if and only if* $p_{ij} > 0$, *and in this case* $l((i, j)) = p_{ij}$.

EXERCISE 13.12 We consider two urns U_1 and U_2. Initially, each urn contains five balls. There are altogether (in urns U_1 and U_2) two black balls, four red balls, and four green balls. At each step, a ball is drawn from each urn and replaced in the other urn. Let X_n be the number of black balls in U_1 after n swaps.

1. Check that the X_ns form a Markov chain.
2. Draw the graph of the chain and write its transition matrix. Is the chain irreducible?
3. Find the distribution of X_0 in order for all the X_ns to have the same distribution. (Hint: refer to Exercise 13.5.) ♢

EXERCISE 13.13 Let U_1 and U_2 be two urns. Initially, each urn contains two black balls and seven red balls. At each step, we draw a ball from each urn and replace it in the other urn. Let X_n be the number of black balls in U_1 at step n, i.e. after n swaps; the initial state corresponds to $n = 0$.

1. Show that X_n is a Markov chain. What are its states? Compute the initial probabilities $q_i = P(X_0 = i)$.
2. Compute the transition probabilities of X_n. Deduce the transition matrix of the chain.

3. Draw the graph of the chain. Is the chain irreducible?

4. Let $L(n)$ be the vector of the probabilities of the states of the chain at step n. Compute $L(0), L(1)$, and $L(2)$. ♢

We have now a simple characterization of persistent and transient states.

Proposition 13.29

1. *A Markov chain is irreducible if and only if its graph is strongly connected.*

2. *A state of a Markov chain is persistent if and only if its strongly connected component is a maximal element for the order $<$ on the reduced graph associated with the graph of the Markov chain and defined in Proposition* 13.27.

3. *A state of a Markov chain is absorbing if and only if it is persistent and its strongly connected component is reduced to that single state.*

Example 13.30 Consider the Markov chain with transition matrix

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

The associated graph is described by Figure 13.1, where the strongly connected components are circled with hyphenated lines: they are thus $C_1 = \{1\}$, $C_2 = \{2\}$, $C_3 = \{3, 4, 5\}$, and $C_4 = \{6, 7\}$

The associated reduced graph is described in Figure 13.2.

The persistent states are thus the states of C_1 and C_4; we can decompose the set of states of the chain into $S = T \cup C_1 \cup C_4$, where C_1 and C_4 are the closed irreducible sets of persistent states and $T = C_2 \cup C_3$ is the set of transient states. State 1 is absorbing.

Exercise 13.14 Draw the graph of the Markov chain of Exercise 13.4. Which are the absorbing, persistent, transient states? ♢

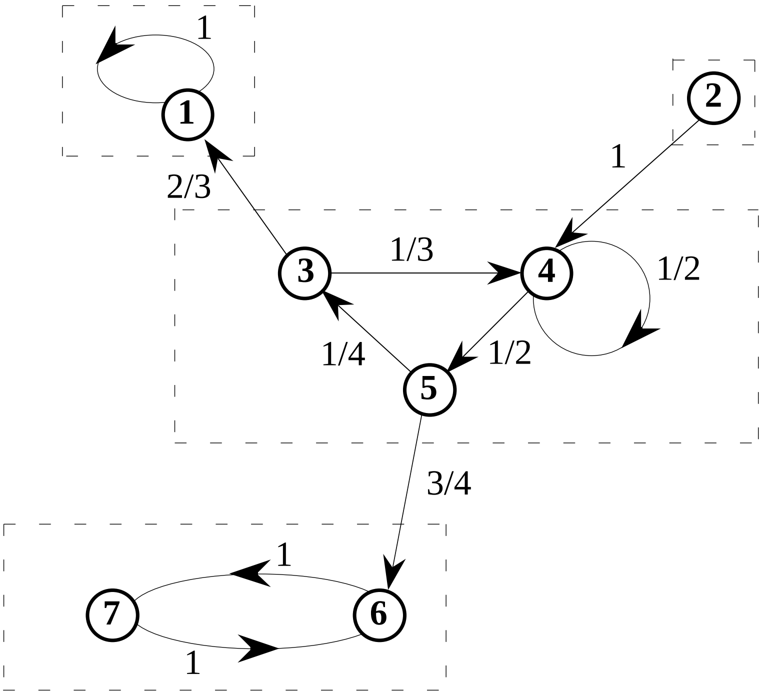

Figure 13.1

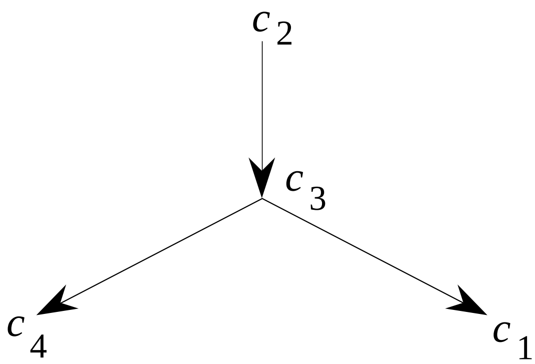

Figure 13.2

13.3.4 Probability and average waiting time for absorption

Consider a Markov chain decomposed as $S = T \cup C_1 \cup \cdots \cup C_k$. Assume that the chain starts in state E_i. If E_i is a persistent state in C_j, the chain will remain in C_j forever. If E_i is a transient state in T, then since the chain is finite, it will (almost surely) after some finite time go in one of the irreducible closed sets of persistent states C_j and will remain there forever. Then, for an arbitrary state E_i in the set $T = \{E_m \,/\, m = 1, \ldots, p\}$ of transient states, there arise the problems of determining

1. the probability that, starting from E_i, the chain ends up in C_j (called the

probability of ultimate absorption of E_i in C_j) and
2. the average waiting time after which this absorption occurs (called the *average waiting time before absorption* and corresponding to the average number of steps before absorption).

In each sequence, let C_j be a closed irreducible set of persistent states. Let λ_{in} be the probability that the chain, starting from E_i, is absorbed in C_j at the nth step exactly, i.e.

$$\lambda_{in} = P(X_n \in C_j \ , \ X_k \notin C_j \ \forall k < n \ / \ X_0 = E_i) \, .$$

Then:

1. The probability λ_i of ultimate absorption of E_i in C_j is $\lambda_i = \sum_{n=1}^{\infty} \lambda_{in}$.
2. The average waiting time before absorption is given by $t = \sum_{n=1}^{\infty} n\lambda_{in}$.

Several methods are available for computing λ_{in} and λ_i , by which we can deduce t.

1. We note that

$$\sum_{k \in C_j} p_{ik}^{(n)} = \lambda_{i1} + \lambda_{i2} + \cdots + \lambda_{in} \, .$$

Hence, when n goes to infinity,

$$\lambda_i = \lim_{n \to \infty} \sum_{k \in C_j} p_{ik}^{(n)} \, ,$$

which we can obtain by studying the matrix

$$P' = \lim_{n \to \infty} P^n .$$

2. We can also write that the first passage in C_j occurs at step n if at step $n-1$ the system was still in T and if it goes from T to C_j at step n. Thus

$$\lambda_{in} = \sum_{m \in T} p_{im}^{(n-1)} \Big(\sum_{k \in C_j} p_{mk} \Big) \, ,$$

which defines a direct computation of λ_{in} in terms of the elements of the submatrix of transient elements and of its powers; the latter computation can be simpler than the former one.

3. Finally, we can write that the first passage in C_j occurs at step n if the first step took us to a transient state m and if, starting from m, we have reached C_j after $n-1$ steps exactly. This gives

$$\lambda_{in} = \sum_{m \in T} p_{im} \lambda_{m(n-1)} , \quad \forall n > 1,$$

and summing on n,

$$\lambda_i - \lambda_{i1} = \sum_{m \in T} p_{im} \lambda_m, \text{ with } \lambda_{i1} = \sum_{k \in C_j} p_{ik} . \tag{13.7}$$

We thus obtain a system of $|T| = card(T)$ linear equations in the unknowns $\lambda_1, \ldots, \lambda_{|T|}$, and we can show that this system has a unique solution.

EXERCISE 13.15 Show that when $k = 1$, i.e. when there is a single closed irreducible set C_k, it is the case that $\forall i\,,\ \lambda_i = 1$. ♢

4. Note, finally, that we can use the generating functions and their partial fraction expansions in order to compute the transition probabilities in n steps $p_{ij}^{(n)}$ and deduce λ_{in} and λ_i.

We introduce the generating functions $Z_i = P_{ij}(z) = \sum_{n=0}^{\infty} p_{ij}^{(n)} z^n$ and show that they are solutions of a linear equation system of the form $Z_i - z \sum_{k=1}^{r} p_{ij} Z_j = b_i$. The solutions Z_i of such a system are rational functions that are decomposed into partial fractions and then expanded into geometric power series in order to obtain the coefficients $p_{ij}^{(n)}$ of Z_i. (See Feller.)

EXERCISE 13.16 We represent the curriculum for a student on a two-year course by a Markov chain defined as follows: the success probabilities at both the first-year and second-year final exams are p; the failure probabilities at both the first-year and second-year final exams are q; the probability of dropping out of the course at the end of each year is r; we have $p + q + r = 1$. The states of the Markov chain are the two years of study denoted by 1 and 2, together with a 'drop-out' state denoted by a and a 'success' state denoted by s. The average student will doubtless be interested in the probability of success and the average waiting time for reaching success. These are computed in 4 and 7 below.

1. Draw the graph of the chain and give the transition matrix. (States may be represented by $\{1, 2, a, s\}$, where 1 and 2 represent the first and second years, and a and s represent 'drop-out' and 'success'.)
2. Is the chain irreducible? What are the transient (resp. absorbing, persistent) states?
3. Let $\lambda_{i,t}^n$ for $i = 1, 2$ and $t = a, s$ be the probabilities that the chain, given that it started from initial state i, reaches state t (with $t \in \{a, s\}$) for the first time at step n, i.e. $\lambda_{i,t}^n = P(X_0 = i, X_1 \neq t, \ldots, X_{n-1} \neq t, X_n = t \,/\, X_0 = i)$. Compute the probability that the chain, given that it started from the initial state i, reaches state a (resp. s) for the first time at time $n+1$ in terms of the $\lambda_{i,t}^n$s.

4. Let $\lambda_{i,t}$ for $i = 1, 2$ and $t = a, s$ be the probabilities that the chain, given that it started from initial state i, ends up in t (with $t \in \{a, s\}$). Compute the ultimate absorption probabilities $\lambda_{i,t}$ in terms of the data p, q, r. Compute these ultimate absorption probabilities assuming $p = 0.6$, $q = 0.3$, and $r = 0.1$.

5. Let M be the matrix

$$M = \begin{pmatrix} q & p \\ 0 & q \end{pmatrix} .$$

Show by induction on n that, $\forall n \geq 0$,

$$M^n = \begin{pmatrix} q^n & nq^{n-1}p \\ 0 & q^n \end{pmatrix}$$

(with the conventions that $0^0 = 1$ and that $M^0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$). Deduce the values of the $\lambda_{i,t}^n$s for $i = 1, 2$, $t = a, s$ and $n \in \mathbb{N}$.

6. Compute the generating function of the random variable N_i giving the number of years of study before leaving the course, assuming that the student started at year i, with $i = 1, 2$. N_i , equivalently, represents the time needed in order for the chain, given that it started from initial state i, to reach either one of the states in the set $\{s, a\}$. Deduce the average number m_i of years of study in order for a student who started at year i, with $i = 1, 2$, to leave the course.

7. Compute the average number m_i^s (resp. m_i^a) of years of study ending with final success (resp. final failure) for a student who started at year i, with $i = 1, 2$. Compute these average waiting times assuming $p = 0.6$, $q = 0.3$ and $r = 0.1$.

8. In this example, what is the intuitive meaning of the Markov chain hypothesis? $\diamondsuit$