Chapter 1

Discrete-Time Markov Chains

We consider in this chapter a collection of random variables $X = \{X_n, \ n \in \mathbb{N}\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with values in a countable set $S$ and satisfying the Markov property, that is the past and the future of $X$ are independent when its present state is known. Time is represented here by the subscript $n$, which is the reason we refer to discrete time. The set $S$ is often called the state space.

1.1. Definitions and properties

**Definition 1.1.** A stochastic process $X = \{X_n, \ n \in \mathbb{N}\}$ on a state space $S$ is a discrete-time Markov chain if:

- for all $n \geq 0$, $X_n \in S$,
- for all $n \geq 1$ and for all $i_0, \ldots, i_{n-1}, i_n \in S$, we have:

$$\mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \mathbb{P}\{X_n = i_n \mid X_{n-1} = i_{n-1}\}.$$  

**Definition 1.2.** A discrete-time Markov chain $X = \{X_n, \ n \in \mathbb{N}\}$ on a state space $S$ is said to be homogeneous if, for all $n, k \in \mathbb{N}$ and, for all $i, j \in S$, we have:

$$\mathbb{P}\{X_{n+k} = j \mid X_k = i\} = \mathbb{P}\{X_n = j \mid X_0 = i\}\text{.}$$

All the following Markov chains are considered homogeneous. The term Markov chain in this chapter will thus designate a homogeneous discrete-time Markov chain.
We consider, for all \( i, j \in S \), \( P_{i,j} = \Pr\{X_n = j \mid X_{n-1} = i\} \) and we define the transition probability matrix \( P \) of the Markov chain \( X \) as:

\[
P = (P_{i,j})_{i,j \in S}.
\]

We then have, by definition, for all \( i, j \in S \),

\[
P_{i,j} \geq 0 \quad \text{and} \quad \sum_{j \in S} P_{i,j} = 1.
\]

A matrix for which these two properties hold is called a stochastic matrix. For all \( n \in \mathbb{N} \), we write \((P^n)_{i,j}\) the coefficient \((i, j)\) of the matrix \(P^n\), where we define \(P^0 = I\), with \(I\) the identity matrix whose dimension will be contextually given further on – here it is equal to the number of states \(|S|\) of \(S\). We write \(\alpha = (\alpha_i, i \in S)\) the row vector containing the initial distribution of the Markov chain \(X\), defined by:

\[
\alpha_i = \Pr\{X_0 = i\}.
\]

For all \(i \in S\), we thus have:

\[
\alpha_i \geq 0 \quad \text{and} \quad \sum_{i \in S} \alpha_i = 1.
\]

**Theorem 1.1.–** The process \(X = \{X_n, n \in \mathbb{N}\}\) on the state space \(S\) is a Markov chain with initial distribution \(\alpha\) and transition probability matrix \(P\) if and only if for all \(n \geq 1\) and for all \(i_0, \ldots, i_n \in S\), we have:

\[
\Pr\{X_n = i_n, \ldots, X_0 = i_0\} = \alpha_{i_0} P_{i_0,i_1} \cdots P_{i_{n-1},i_n}.
\]  

**Proof.–** If \(X = \{X_n, n \in \mathbb{N}\}\) is a Markov chain then:

\[
\Pr\{X_n = i_n, \ldots, X_0 = i_0\} = \Pr\{X_n = i_n \mid X_{n-1} = i_{n-1}\} \Pr\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}
\]

\[
= \Pr\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} P_{i_{n-1},i_n}.
\]

Iterating this calculation \(n - 1\) times over \(\Pr\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}\), we obtain:

\[
\Pr\{X_n = i_n, \ldots, X_0 = i_0\} = \Pr\{X_0 = i_0\} P_{i_0,i_1} \cdots P_{i_{n-1},i_n}
\]

\[
= \alpha_{i_0} P_{i_0,i_1} \cdots P_{i_{n-1},i_n}.
\]
Conversely, if relation [1.1] is satisfied then we have \( P\{X_0 = i_0\} = \alpha_{i_0} \) and, for \( n \geq 1 \),

\[
P\{X_n = i_n \mid X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\} = \frac{P\{X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}}{P\{X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}}
= \frac{\alpha_{i_0} P_{i_0, i_1} \cdots P_{i_{n-1}, i_n}}{\alpha_{i_0} P_{i_0, i_1} \cdots P_{i_{n-2}, i_{n-1}}}
= P_{i_{n-1}, i_n}
= P\{X_n = i_n \mid X_{n-1} = i_{n-1}\}.
\]

\( X \) is, therefore, a Markov chain with initial distribution \( \alpha \) and transition probability matrix \( P \).

This result shows that a discrete-time Markov chain is completely determined by its initial distribution \( \alpha \) and transition probability matrix \( P \).

In the following sections, we will often use products of infinite matrices or vector-matrix or matrix-vector products of infinite dimension. Remember that, in general, these products are not associative except when the affected matrices or vectors have non-negative coefficients, which is the case in this chapter. More details on this subject are given in the first chapter of [KEM 66] and in section 4.3.

**THEOREM 1.2.**— If \( X \) is a Markov chain on the state space \( S \), with initial distribution \( \alpha \) and transition probability matrix \( P \) then, for all \( i, j \in S \) and, for all \( n \geq 0 \), we have:

1) \( P\{X_n = j \mid X_0 = i\} = (P^n)_{i,j} \);

2) \( P\{X_n = j\} = (\alpha P^n)_j \).

**PROOF.**—

1) For all \( m, n \geq 0 \), we have:

\[
P\{X_{n+m} = j \mid X_0 = i\}
= \sum_{k \in S} P\{X_{n+m} = j, X_n = k \mid X_0 = i\}
= \sum_{k \in S} P\{X_{n+m} = j \mid X_n = k, X_0 = i\} P\{X_n = k \mid X_0 = i\}
= \sum_{k \in S} P\{X_{n+m} = j \mid X_n = k\} P\{X_n = k \mid X_0 = i\}
= \sum_{k \in S} P\{X_m = j \mid X_0 = k\} P\{X_n = k \mid X_0 = i\},
\]
where the third equality uses the Markov property and the fourth uses the homogeneity of $X$.

Defining $P_{i,j}(n) = \mathbb{P}\{X_n = j \mid X_0 = i\}$, this last relation becomes:

$$P_{i,j}(n + m) = \sum_{k \in S} P_{i,k}(n)P_{k,j}(m),$$

that is if $P(n)$ denotes the matrix with coefficients $P_{i,j}(n)$,

$$P(n + m) = P(n)P(m).$$

These equations are called the Chapman–Kolmogorov equations. In particular, as $P(1) = P$, we have:

$$P(n) = P(n - 1)P = P(n - 2)P^2 = \cdots = P^n.$$

2) We obtain, using point 1,

$$\mathbb{P}\{X_n = j\} = \sum_{i \in S} \mathbb{P}\{X_n = j \mid X_0 = i\}\mathbb{P}\{X_0 = i\}$$

$$= \sum_{i \in S} \alpha_i (P^n)_{i,j}$$

$$= (\alpha P^n)_j,$$

which completes the proof.

In particular, this result shows that if $P$ is stochastic then $P^n$ is also stochastic, for all $n \geq 2$.

**Theorem 1.3.** If $X$ is a Markov chain then, for all $n \geq 0$, $0 \leq k \leq n$, $m \geq 1$, for all $i_k, \ldots, i_n \in S$ and $j_1, \ldots, j_m \in S$, we have:

$$\mathbb{P}\{X_{n+m} = j_m, \ldots, X_{n+1} = j_1 \mid X_n = i_n, \ldots, X_k = i_k\}$$

$$= \mathbb{P}\{X_m = j_m, \ldots, X_1 = j_1 \mid X_0 = i_n\}.$$
PROOF.– Using theorem 1.1, we have:

\[
P\{X_{n+m} = j_m, \ldots, X_{n+1} = j_1 \mid X_n = i_n, \ldots, X_k = i_k\} = \frac{P\{X_{n+m} = j_m, \ldots, X_{n+1} = j_1, X_n = i_n, \ldots, X_0 = i_0\}}{P\{X_n = i_n, \ldots, X_k = i_k\}}
\]

\[
= \sum_{i_0, \ldots, i_{k-1} \in S} \frac{P\{X_n = i_n, \ldots, X_0 = i_0\}}{P\{X_n = i_n, \ldots, X_k = i_k\}}
\]

\[
= \sum_{i_0, \ldots, i_{k-1} \in S} \alpha_{i_0} P_{i_0,i_1} \cdots P_{i_{k-1},i_k} P_{i_k,i_{k+1}} \cdots P_{i_{n-1},i_n} P_{i_n,j_1} \cdots P_{j_{m-1},j_m}
\]

\[
= P_{i_n,j_1} \cdots P_{j_{m-1},j_m}
\]

\[
= P\{X_m = j_m, \ldots, X_1 = j_1 \mid X_0 = i_n\},
\]

which completes the proof. \]

The Markov property seen so far stated that the past and the future are independent when the present is known at a given deterministic time \(n\). The strong Markov property allows us to extend this independence when the present is known at a particular random time which is called a stopping time.

1.2. Strong Markov property

Let \(X = \{X_n, \ n \in \mathbb{N}\}\) be a Markov chain on the state space \(S\), defined on the probability space \((\Omega, \mathcal{F}, P)\). For all \(n \geq 0\), we denote by \(\mathcal{F}_n\) the \(\sigma\)-algebra of events expressed as a function of \(X_0, \ldots, X_n\), that is:

\[
\mathcal{F}_n = \\{\omega \in \Omega \mid (X_0(\omega), \ldots, X_n(\omega)) \in B_n\}, B_n \in \mathcal{P}(S^{n+1})\},
\]

where, for a set \(E\), \(\mathcal{P}(E)\) denotes the set of all subsets of \(E\) and \(S^{n+1}\) is the set of all \((n+1)\)-dimensional vectors, whose entries are states of \(S\). For all \(i \in S\), we write \(\delta^i = (\delta^i_j, \ j \in S)\) the probability distribution concentrated on the state \(i\), defined by:

\[
\delta^i_j = 1_{\{i=j\}}.
\]

THEOREM 1.4.– If \(X = \{X_n, \ n \in \mathbb{N}\}\) is a Markov chain on the state space \(S\) then, for all \(n \geq 0\) and for all \(i \in S\), conditional on \(\{X_n = i\}\), the process \(\{X_{n+p}, \ p \in \mathbb{N}\}\)
is a Markov chain with initial distribution $\delta^i$ and transition probability matrix $P$, independent of $(X_0, \ldots, X_n)$. This means that for all $A \in \mathcal{F}_n$, for all $m \geq 1$ and for all $j_1, \ldots, j_m \in S$, we have:

$$\mathbb{P}\{ X_{n+m} = j_m, \ldots, X_{n+1} = j_1, A \mid X_n = i \} = \mathbb{P}\{ X_m = j_m, \ldots, X_1 = j_1 \mid X_0 = i \} \mathbb{P}\{ A \mid X_n = i \}.$$ 

PROOF. – It is sufficient to prove the result when:

$$A = \{X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}.$$ 

Indeed, $A$ is a countable union of disjoint events of this form, therefore, the general case can be deduced using the $\sigma$-additivity property. It is also sufficient to consider the case where $i_n = i$ as in the contrary case the two sides are null.

Let $A = \{X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\}$. We have, by the Markov property and applying theorem 1.3,

$$\mathbb{P}\{ X_{n+m} = j_m, \ldots, X_{n+1} = j_1, A \mid X_n = i \} = \mathbb{P}\{ X_m = j_m, \ldots, X_1 = j_1 \mid X_0 = i, A \} \mathbb{P}\{ A \mid X_n = i \} = \mathbb{P}\{ X_m = j_m, \ldots, X_1 = j_1 \mid X_0 = i \} \mathbb{P}\{ A \mid X_n = i \},$$

which completes the proof. 

DEFINITION 1.3.– A random variable $T$ with values in $\mathbb{N} \cup \{\infty\}$ is called a stopping time for the process $X$ if for all $n \geq 0$, $\{T = n\} \in \mathcal{F}_n$.

In the following section, we often use the variable $\tau(j)$ that counts the number of transitions necessary to reach state $j$, defined by:

$$\tau(j) = \inf\{n \geq 1 \mid X_n = j\},$$

where $\tau(j) = \infty$ if this set is empty. For all $j \in S$, $\tau(j)$ is a stopping time since $\{\tau(j) = 0\} = \emptyset \in \mathcal{F}_0$, $\{\tau(j) = 1\} = \{X_1 = j\} \in \mathcal{F}_1$ and, for $n \geq 2$,

$$\{\tau(j) = n\} = \{X_n = j, X_k \neq j, 1 \leq k \leq n - 1\} \in \mathcal{F}_n.$$

Let $T$ be a stopping time, and $\mathcal{F}_T$ the $\sigma$-algebra of events expressed as a function of $X_0, \ldots, X_T$, that is:

$$\mathcal{F}_T = \{ B \in \mathcal{F} \mid \forall n \in \mathbb{N}, B \cap \{T = n\} \in \mathcal{F}_n \}.$$
**Theorem 1.5:** Strong Markov Property. If $X = \{X_n, \ n \in \mathbb{N}\}$ is a Markov chain and $T$ a stopping time for $X$ then, for all $i \in S$, conditional on $\{T < \infty\} \cap \{X_T = i\}$, the process $\{X_{T+n}, \ n \in \mathbb{N}\}$ is a Markov chain with initial distribution $\delta_i$ and transition probability matrix $P$, independent of $(X_0, \ldots, X_T)$. This means that for all $A \in \mathcal{F}_T$, for all $m \geq 1$ and for all $j_1, \ldots, j_m \in S$, we have:

$$
P \{X_{T+m} = j_m, \ldots, X_{T+1} = j_1, A \mid T < \infty, X_T = i\} = \sum_{n=0}^{\infty} \frac{P \{X_{T+n} = j_m, \ldots, X_n = j_1, A \mid T = n, X_T = i\} P \{X_n = i\}}{P \{T < \infty, X_T = i\}} \frac{P \{X_{T+1} = j_1, A \mid T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}}.
$$

where the fifth equality is obtained using Theorem 1.4 because $A \cap \{T = n\} \in \mathcal{F}_n$. ■

**Proof.** We have:

$$
P \{X_{T+m} = j_m, \ldots, X_{T+1} = j_1, A \mid T < \infty, X_T = i\} = \sum_{n=0}^{\infty} \frac{P \{X_{T+n} = j_m, \ldots, X_T = j_1, A, T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}} P \{X_T = i\} \frac{P \{X_{T+1} = j_1, A \mid T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}}.
$$

$$
= \sum_{n=0}^{\infty} \frac{P \{X_{T+n} = j_m, \ldots, X_T = j_1, A \mid T = n, X_T = i\} P \{X_n = i\}}{P \{T < \infty, X_T = i\}} \frac{P \{X_{T+1} = j_1, A \mid T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}}.
$$

$$
= \sum_{n=0}^{\infty} \frac{P \{X_{T+n} = j_m, \ldots, X_T = j_1, A \mid T = n, X_T = i\} P \{X_n = i\}}{P \{T < \infty, X_T = i\}} \frac{P \{X_{T+1} = j_1, A \mid T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}}.
$$

$$
= \sum_{n=0}^{\infty} \frac{P \{X_{T+n} = j_m, \ldots, X_T = j_1, A \mid T = n, X_T = i\} P \{X_n = i\}}{P \{T < \infty, X_T = i\}} \frac{P \{X_{T+1} = j_1, A \mid T < \infty, X_T = i\}}{P \{T < \infty, X_T = i\}}.
$$

where the fifth equality is obtained using Theorem 1.4 because $A \cap \{T = n\} \in \mathcal{F}_n$. ■
1.3. Recurrent and transient states

Let us recall that the random variable $\tau(j)$ that counts the number of transitions necessary to reach state $j$ is defined by:

$$\tau(j) = \inf \{n \geq 1 \mid X_n = j\},$$

where $\tau(j) = \infty$ if this set is empty.

For all $i, j \in S$ and, for all $n \geq 1$, we define:

$$f^{(n)}_{i,j} = \mathbb{P}\{\tau(j) = n \mid X_0 = i\} = \mathbb{P}\{X_n = j, X_{k} \neq j, 1 \leq k \leq n - 1 \mid X_0 = i\}.$$

For $n = 1$, we have, of course, $f^{(1)}_{i,j} = \mathbb{P}\{\tau(j) = 1 \mid X_0 = i\} = \mathbb{P}\{X_1 = j \mid X_0 = i\} = P_{i,j}$. Hence $f^{(n)}_{i,i}$ is the probability, starting from $i$, that the first return to state $i$ occurs at time $n$ and, for $i \neq j$, $f^{(n)}_{i,j}$ is the probability, starting from $i$, that the first visit to state $j$ occurs at time $n$.

**Theorem 1.6.** For all $i, j \in S$ and, for all $n \geq 1$, we have:

$$(P^n)_{i,j} = \sum_{k=1}^{n} f^{(k)}_{i,j} (P^{n-k})_{j,j}, \quad \text{[1.2]}$$

recalling that $(P^0)_{i,j} = 1_{\{i = j\}}$.

**Proof.** For $i, j \in S$ and $n \geq 1$, we have $X_n = j \implies \tau(j) \leq n$, by definition of $\tau(j)$. From this we obtain:

$$(P^n)_{i,j} = \mathbb{P}\{X_n = j \mid X_0 = i\}$$

$$= \mathbb{P}\{X_n = j, \tau(j) \leq n \mid X_0 = i\}$$

$$= \sum_{k=1}^{n} \mathbb{P}\{X_n = j, \tau(j) = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{n} \mathbb{P}\{X_n = j \mid \tau(j) = k, X_0 = i\} \mathbb{P}\{\tau(j) = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{n} f^{(k)}_{i,j} \mathbb{P}\{X_n = j \mid X_k = j, \tau(j) = k, X_0 = i\}$$

$$= \sum_{k=1}^{n} f^{(k)}_{i,j} \mathbb{P}\{X_n = j \mid X_k = j\}$$

$$= \sum_{k=1}^{n} f^{(k)}_{i,j} (P^{n-k})_{j,j},$$
where the fifth equality comes from the fact that \( \{ \tau(j) = k \} = \{ X_k = j, \tau(j) = k \} \) and the penultimate equality uses the Markov property since \( \tau(j) \) is a stopping time.

For all \( i, j \in S \), we define \( f_{i,j} \) as:

\[
 f_{i,j} = P\{ \tau(j) < \infty \mid X_0 = i \} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.
\]

The quantity \( f_{i,i} \) is the probability, starting from \( i \), that the first return to state \( i \) occurs in a finite time and, for \( i \neq j \), \( f_{i,j} \) is the probability, starting from \( i \), that the first visit to state \( j \) occurs in a finite time.

The calculation of \( f_{i,j}^{(n)} \) and \( f_{i,j} \) can be carried out using the following result.

**THEOREM 1.7.** For all \( i, j \in S \) and, for all \( n \geq 1 \), we have:

\[
 f_{i,j}^{(n)} = \begin{cases} 
 P_{i,j} & \text{if } n = 1 \\
 \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} f_{\ell,j}^{(n-1)} & \text{if } n \geq 2 
\end{cases}
\]

and

\[
 f_{i,j} = P_{i,j} + \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} f_{\ell,j}.
\]

**PROOF.** From the definition of \( f_{i,j}^{(n)} \), we have, for \( n = 1 \), \( f_{i,j}^{(1)} = P_{i,j} \). For \( n \geq 2 \), we have:

\[
 f_{i,j}^{(n)} = P\{ X_n = j, X_k \neq j, 1 \leq k \leq n - 1 \mid X_0 = i \}
\]

\[
 = \sum_{\ell \in S \setminus \{j\}} P\{ X_n = j, X_k \neq j, 2 \leq k \leq n - 1, X_1 = \ell \mid X_0 = i \}
\]

\[
 = \sum_{\ell \in S \setminus \{j\}} P\{ X_1 = \ell \mid X_0 = i \}
\]

\[
 \times P\{ X_n = j, X_k \neq j, 2 \leq k \leq n - 1 \mid X_1 = \ell, X_0 = i \}
\]

\[
 = \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} P\{ X_n = j, X_k \neq j, 2 \leq k \leq n - 1 \mid X_1 = \ell, X_0 = i \}.
\]
Successively using the Markov property and the homogeneity of the Markov chain, we obtain:

\[ f^{(n)}_{i,j} = \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} \mathbb{P}\{X_n = j, X_k \neq j, 2 \leq k \leq n - 1 \mid X_1 = \ell\} \]

\[ = \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} \mathbb{P}\{X_{n-1} = j, X_k \neq j, 1 \leq k \leq n - 2 \mid X_0 = \ell\} \]

\[ = \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} f^{(n-1)}_{\ell,j}. \]

Summing over \( n \), we obtain, using Fubini’s theorem, the second relation.

**Definition 1.4.** A state \( i \in S \) is called recurrent if \( f_{i,i} = 1 \) and transient if \( f_{i,i} < 1 \). A Markov chain is called recurrent (respectively transient) if all its states are recurrent (respectively transient).

**Definition 1.5.** A state \( i \in S \) is called absorbing if \( P_{i,i} = 1 \). All absorbing states are recurrent. Indeed, if \( i \) is an absorbing state then, by definition, we have \( f^{(n)}_{i,i} = 1_{\{n=1\}} \) and so \( f_{i,i} = 1 \), which means that the state \( i \) is recurrent.

**Theorem 1.8.** The state \( j \) is recurrent if and only if:

\[ \sum_{n=1}^{\infty} (P^n)_{j,j} = \infty. \]

**Proof.** Let us resume equation [1.2] for \( i = j \), that is:

\[ (P^n)_{j,j} = \sum_{k=1}^{n} f_{j,j}^{(k)} (P^{n-k})_{j,j}. \]

Summing over \( n \), using Fubini’s theorem and since \( (P^0)_{j,j} = 1 \), we obtain:

\[ \sum_{n=1}^{\infty} (P^n)_{j,j} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{j,j}^{(k)} (P^{n-k})_{j,j} \]

\[ = \sum_{k=1}^{\infty} f_{j,j}^{(k)} \sum_{n=k}^{\infty} (P^{n-k})_{j,j} \]

\[ = f_{j,j} \sum_{n=0}^{\infty} (P^n)_{j,j} = f_{j,j} \left(1 + \sum_{n=1}^{\infty} (P^n)_{j,j}\right). \]
It follows that if \( u_j = \sum_{n=1}^{\infty} (P^n)_{j,j} < \infty \) then \( f_{j,j} = u_j/(1 + u_j) < 1 \), which means that state \( j \) is transient.

Conversely, let \( u_j(N) = \sum_{n=1}^{N} (P^n)_{j,j} \) and assume that \( \lim_{N \to \infty} u_j(N) = \infty \). We then have, again using equation [1.2] taken for \( i = j \), for all \( N \geq 1 \),

\[
 u_j(N) = \sum_{n=1}^{N} \sum_{k=1}^{n} f_{j,j}^{(k)} (P^{n-k})_{j,j} \\
 = \sum_{k=1}^{N} f_{j,j}^{(k)} \sum_{n=k}^{N} (P^{n-k})_{j,j} \\
 \leq \sum_{k=1}^{N} f_{j,j}^{(k)} \sum_{n=0}^{N} (P^n)_{j,j} \\
 \leq f_{j,j} \left( 1 + u_j(N) \right),
\]

and, therefore, we obtain:

\[
 f_{j,j} \geq \frac{u_j(N)}{1 + u_j(N)} = \frac{1}{1 + \frac{1}{u_j(N)}} \to 1 \text{ when } N \to \infty,
\]

which shows that \( f_{j,j} = 1 \) or, in other words, that state \( j \) is recurrent. \( \blacksquare \)

**Corollary 1.1.**– If state \( j \) is transient then, for all \( i \in S \),

\[
 \sum_{n=1}^{\infty} (P^n)_{i,j} < \infty,
\]

and, therefore,

\[
 \lim_{n \to \infty} (P^n)_{i,j} = 0 \text{ and } \lim_{n \to \infty} \mathbb{P}\{X_n = j\} = 0.
\]

**Proof.**– Let us resume equation [1.2], that is:

\[
 (P^n)_{i,j} = \sum_{k=1}^{n} f_{i,j}^{(k)} (P^{n-k})_{j,j}.
\]
Here again, summing over $n$, using Fubini’s theorem and because $(P^0)_{j,j} = 1$, we obtain:

$$\sum_{n=1}^{\infty} (P^n)_{i,j} = \sum_{n=1}^{\infty} \sum_{k=1}^{n} f_{i,j}^{(k)} (P^{n-k})_{j,j}$$

$$= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \sum_{n=k}^{\infty} (P^{n-k})_{j,j}$$

$$= f_{i,j} \sum_{n=0}^{\infty} (P^n)_{j,j}$$

$$= f_{i,j} \left( 1 + \sum_{n=1}^{\infty} (P^n)_{j,j} \right).$$

If the state $j$ is transient then, from theorem 1.8, we have $\sum_{n=1}^{\infty} (P^n)_{j,j} < \infty$ and since $f_{i,j} \leq 1$, we obtain $\sum_{n=1}^{\infty} (P^n)_{i,j} < \infty$ and so:

$$\lim_{n \to \infty} (P^n)_{i,j} = 0.$$

If $\alpha$ is the initial distribution of $X$, we have:

$$\mathbb{P}\{X_n = j\} = \sum_{i \in S} \alpha_i (P^n)_{i,j}$$

and by the dominated convergence theorem, we obtain:

$$\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = 0,$$

which completes the proof.

### 1.4. State classification

**Definition 1.6.** A state $j \in S$ is said to be accessible from a state $i \in S$ if there exists an integer $n \geq 0$, such that $(P^n)_{i,j} > 0$. We then write $i \rightarrow j$.

**Definition 1.7.** We say that two states $i$ and $j$ communicate if they are accessible from one another. We then write $i \leftrightarrow j$.

**Theorem 1.9.** This communication relation is an equivalence relation, which means that for all $i, j, k \in S$, we have:
1) \( i \leftrightarrow i \) (reflexivity).

2) \( i \leftrightarrow j \quad \iff \quad j \leftrightarrow i \) (symmetry).

3) \( i \leftrightarrow j, j \leftrightarrow k \implies i \leftrightarrow k \) (transitivity).

**Proof.**— Every state \( i \) is accessible from itself because \((P^0)_{i,i} = 1 > 0\). The relation is, therefore, reflexive. It is also symmetric, by definition. As for transitivity, if \( i \leftrightarrow j \) and \( j \leftrightarrow k \) then there exist integers \( n \) and \( m \) such that \((P^n)_{i,j} > 0\) and \((P^m)_{j,k} > 0\). We then obtain:

\[
(P^{n+m})_{i,k} = \sum_{\ell \in S} (P^n)_{i,\ell} (P^m)_{\ell,k} \geq (P^n)_{i,j} (P^m)_{j,k} > 0.
\]

Thus we have proved that \( i \rightarrow k \). In the same way, we prove that \( k \rightarrow i \). \( \blacksquare \)

As with any equivalence relation, the equivalence classes form a partition of the state space \( S \), that is their classes are not empty and disjoint and their union is equal to \( S \). An equivalence class groups all communicating states. For all \( i \in S \), the equivalence class \( C(i) \) of the state \( i \) is defined by:

\[
C(i) = \{ j \in S \mid i \leftrightarrow j \}.
\]

These equivalence classes are the strongly connected components of the graph related to the Markov chain \( X \).

**Definition 1.8.**— A Markov chain \( X \) is said to be irreducible if all its states communicate, namely if it possesses a single equivalence class or if its graph admits a single strongly connected component.

**Theorem 1.10.**— For all \( i, j \in S \), we have:

1) \( i \leftrightarrow j \) and \( j \) recurrent \( \implies \) \( i \) recurrent.

2) \( i \leftrightarrow j \) and \( j \) transient \( \implies \) \( i \) transient.

**Proof.**— If \( i \leftrightarrow j \) then there exist integers \( \ell \geq 0 \) and \( m \geq 0 \) such that \((P^\ell)_{i,j} > 0\) and \((P^m)_{j,i} > 0\). For all \( n \geq 0 \), we have:

\[
(P^{\ell+n+m})_{i,i} = \sum_{k \in S} (P^\ell)_{i,k} (P^{n+m})_{k,i} \geq (P^\ell)_{i,j} (P^{n+m})_{j,i}.
\]
and

$$(P^{n+m})_{j,i} = \sum_{k \in S} (P^n)_{j,k}(P^m)_{k,i} \geq (P^n)_{j,j}(P^m)_{j,i},$$

and thus:

$$(P^{\ell+m+n})_{i,i} \geq (P^{\ell})_{i,j}(P^n)_{j,j}(P^m)_{j,i}.$$  

Summing over $n$, we obtain:

$$\sum_{n=1}^{\infty} (P^n)_{i,i} \geq \sum_{n=1}^{\infty} (P^{\ell+n+m})_{i,i} \geq (P^{\ell})_{i,j}(P^m)_{j,i} \sum_{n=1}^{\infty} (P^n)_{j,j}.$$  

If $j$ is recurrent then, from theorem 1.8, the sum, above, is equal to $\infty$ that proves, again using theorem 1.8, that $i$ is also recurrent. Part (2) is proved by contradiction. Indeed, if $i \leftrightarrow j$ and if $j$ is transient then the state $i$ cannot be recurrent, since following (1), reversing the roles of $i$ and $j$ would imply that $j$ is recurrent.

This theorem shows that recurrence and transience are class properties, which means that if a state of a given equivalence class is recurrent (respectively transient) then all the states of the same class are also recurrent (respectively transient).

### 1.5. Visits to a state

For every state $j \in S$, we denote by $N_j$ the total number of visits to state $j$, except the initial state, that is:

$$N_j = \sum_{n=1}^{\infty} 1_{\{X_n = j\}}.$$  

It is easy to see that, for every state $j \in S$, we have $\{\tau(j) < \infty\} = \{N_j > 0\}$.

**Theorem 1.11.** For all $i, j \in S$ and, for all $\ell \geq 0$, we have:

$$P\{N_j \geq \ell \mid X_0 = i\} = f_{i,j}(f_{j,j})^\ell.$$
PROOF.– Let us consider the random variable $N_{j,m}$ that counts the number of visits to state $j$ from time $m$, which is:

$$N_{j,m} = \sum_{n=m}^{\infty} 1\{X_n=j\}.$$ 

Note that $N_j = N_{j,1}$. For all $\ell \geq 1$, we have, by definition of $\tau(j)$,

$$\{N_j > \ell\} \subseteq \{\tau(j) < \infty\} = \{N_j > 0\}.$$

We, therefore, obtain:

$$\mathbb{P}\{N_j > \ell \mid X_0 = i\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{N_j > \ell, \tau(j) = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{N_{j,k+1} > \ell - 1, \tau(j) = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{N_{j,k+1} > \ell - 1 \mid \tau(j) = k, X_0 = i\} \mathbb{P}\{\tau(j) = k \mid X_0 = i\}$$

$$= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \mathbb{P}\{N_{j,k+1} > \ell - 1 \mid X_{\tau(j)} = j, \tau(j) = k, X_0 = i\}$$

$$= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \mathbb{P}\{N_{j,k+1} > \ell - 1 \mid X_k = j, \tau(j) = k, X_0 = i\}$$

$$= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \mathbb{P}\{N_{j,k+1} > \ell - 1 \mid X_k = j\}$$

$$= \sum_{k=1}^{\infty} f_{i,j}^{(k)} \mathbb{P}\{N_j > \ell - 1 \mid X_0 = j\}$$

$$= f_{i,j} \mathbb{P}\{N_j > \ell - 1 \mid X_0 = j\},$$

where we use the definition of $\tau(j)$ for the second equality and the fact that $X_{\tau(j)} = j$ when $\tau(j) < \infty$ for the fourth equality. The sixth equality uses the Markov property since $\tau(j)$ is a stopping time and the seventh equality uses the homogeneity of the Markov chain $X$. Taking $i = j$, we obtain, for $\ell \geq 1$,

$$\mathbb{P}\{N_j > \ell \mid X_0 = j\} = f_{j,j} \mathbb{P}\{N_j > \ell - 1 \mid X_0 = j\},$$
therefore, for all $\ell \geq 0$,
\[
\mathbb{P}\{N_j > \ell \mid X_0 = j\} = (f_{j,j})^\ell \mathbb{P}\{N_j > 0 \mid X_0 = j\} = (f_{j,j})^\ell \mathbb{P}\{\tau(j) < \infty \mid X_0 = j\} = (f_{j,j})^{\ell+1},
\]
which gives, for $i, j \in S$ and, for all $\ell \geq 0$,
\[
\mathbb{P}\{N_j > \ell \mid X_0 = i\} = f_{i,j}(f_{j,j})^\ell,
\]
which completes the proof.

**Corollary 1.2.** For all $i, j \in S$, we have:

\[
\mathbb{E}\{N_j \mid X_0 = i\} = \begin{cases} 
0 & \text{if } f_{i,j} = 0, \\
\frac{f_{i,j}}{1 - f_{j,j}} & \text{if } f_{i,j} \neq 0 \text{ and } f_{j,j} < 1, \\
\infty & \text{if } f_{i,j} \neq 0 \text{ and } f_{j,j} = 1.
\end{cases}
\]

**Proof.** If $f_{i,j} = 0$ then, using theorem 1.11, we have $\mathbb{P}\{N_j > \ell \mid X_0 = i\} = 0$ for all $\ell \geq 0$, which means that $\mathbb{P}\{N_j = 0 \mid X_0 = i\} = 1$ or $\mathbb{E}\{N_j \mid X_0 = i\} = 0$. Assume that we have $f_{i,j} \neq 0$. If $f_{j,j} = 1$ then we have, using theorem 1.11, $\mathbb{P}\{N_j > \ell \mid X_0 = i\} = f_{i,j} > 0$ for all $\ell \geq 0$, which means that $\mathbb{E}\{N_j \mid X_0 = i\} = \infty$. If $f_{j,j} < 1$ then, again using theorem 1.11, we have $\mathbb{P}\{N_j < \infty \mid X_0 = i\} = 1$ and
\[
\mathbb{E}\{N_j \mid X_0 = i\} = \sum_{\ell=0}^{\infty} \mathbb{P}\{N_j > \ell \mid X_0 = i\} = f_{i,j} \sum_{\ell=0}^{\infty} (f_{j,j})^\ell = \frac{f_{i,j}}{1 - f_{j,j}},
\]
which completes the proof.

**Corollary 1.3.** Let $i$ and $j$ be two states of $S$.

- The state $j$ is recurrent if and only if $\mathbb{P}\{N_j = \infty \mid X_0 = j\} = 1$.
  In this case, $\mathbb{P}\{N_j = \infty \mid X_0 = i\} = f_{i,j}$.
- The state $j$ is transient if and only if $\mathbb{P}\{N_j < \infty \mid X_0 = j\} = 1$.
  In this case, $\mathbb{P}\{N_j < \infty \mid X_0 = i\} = 1$.

**Proof.** From theorem 1.11, we have, for all $\ell \geq 0$,
\[
\mathbb{P}\{N_j > \ell \mid X_0 = i\} = f_{i,j}(f_{j,j})^\ell.
\]
If $j$ is recurrent then $f_{j,j} = 1$ and thus we have $\mathbb{P}\{N_j > \ell \mid X_0 = i\} = f_{i,j}$ for all $\ell \geq 0$, which means that $\mathbb{P}\{N_j = \infty \mid X_0 = i\} = f_{i,j}$ which is equal to 1 if $i = j$. Conversely, if $\mathbb{P}\{N_j = \infty \mid X_0 = j\} = 1$ then we have $\mathbb{P}\{N_j > \ell \mid X_0 = j\} = 1$ for all $\ell \geq 0$ and, therefore, $f_{j,j} = 1$.

If $j$ is transient then $f_{j,j} < 1$ and thus $\mathbb{P}\{N_j > \ell \mid X_0 = i\}$ tends to 0 when $\ell$ tends to infinity, which means that $\mathbb{P}\{N_j < \infty \mid X_0 = i\} = 1$. Conversely, if $\mathbb{P}\{N_j < \infty \mid X_0 = j\} = 1$ then necessarily $f_{j,j} < 1$.

**Theorem 1.12.** For all $i, j \in S$, we have:

$$\mathbb{E}\{N_j \mid X_0 = i\} = \sum_{n=1}^{\infty} (P^n)_{i,j}.$$ 

**Proof.** Using the monotone convergence theorem, we have:

$$\mathbb{E}\{N_j \mid X_0 = i\} = \mathbb{E}\left\{ \sum_{n=1}^{\infty} 1\{X_n = j\} \mid X_0 = i\right\}$$

$$= \sum_{n=1}^{\infty} \mathbb{E}\{1\{X_n = j\} \mid X_0 = i\}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\{X_n = j \mid X_0 = i\}$$

$$= \sum_{n=1}^{\infty} (P^n)_{i,j},$$

which completes the proof.

Thus we have shown that a state $j$ is recurrent if and only if, starting from $j$, the Markov chain $X$ visits state $j$ an infinite number of times. If state $j$ is recurrent and if the chain starts in a state $i \neq j$ then, if it reaches state $j$, with probability $f_{i,j}$, it will return to $j$ infinitely many times, otherwise, with probability $1 - f_{i,j}$, it will never reach state $j$. If state $j$ is transient then, whichever state $i$ the chain starts in, it will only visit state $j$ a finite number of times and the average number of visits to $j$ will also be finite.

In the following, if $B$ is an event, we will say, if there is need to simplify the notation, that we have $B$, $\mathbb{P}$-almost surely, or $B$, $\mathbb{P}$-a.s. if $\mathbb{P}\{B\} = 1$. Similarly, for a state $i \in S$, we will say that we have $B$, $\mathbb{P}_i$-almost surely or $B$, $\mathbb{P}_i$-a.s. if $\mathbb{P}\{B \mid X_0 = i\} = 1$. For example corollary 1.3 can be also written: a state $j \in S$ is recurrent if and only if $N_j = \infty$, $\mathbb{P}_j$-a.s. A state $j \in S$ is transient if and only
if $N_j < \infty$, $\mathbb{P}_j$-a.s. Likewise, in the statement of theorem 1.16, we could write $\tau(j) < \infty$, $\mathbb{P}$-a.s., in place of $\mathbb{P}\{\tau(j) < \infty\} = 1$.

1.6. State space decomposition

**Lemma 1.1.** For all $i, j \in S$ such that $i \neq j$, we have:

$$i \rightarrow j \iff f_{i,j} > 0.$$  

**Proof.** Let $i$ and $j$ be two states of $S$ such that $i \neq j$.

If $i \rightarrow j$ then, by definition and since $(P^0)_{i,j} = 0$, there exists an integer $n \geq 1$ such that $(P^n)_{i,j} > 0$. We then obtain, from relation [1.2], that for this integer $n$, we have:

$$\sum_{k=1}^{n} f_{i,j}^{(k)} (P^{n-k})_{j,j} > 0.$$  

This implies that there exists an integer $k \in \{1, \ldots, n\}$ such that $f_{i,j}^{(k)} > 0$. Hence,

$$f_{i,j} = \sum_{\ell=1}^{\infty} f_{i,j}^{(\ell)} \geq f_{i,j}^{(k)} > 0.$$  

Conversely, if $f_{i,j} > 0$ then there exists an integer $n \geq 1$ such that $f_{i,j}^{(n)} > 0$. For this integer $n$ relation [1.2] gives, since $(P^0)_{j,j} = 1$,

$$(P^n)_{i,j} = \sum_{k=1}^{n} f_{i,j}^{(k)} (P^{n-k})_{j,j} \geq f_{i,j}^{(n)} > 0,$$

which means that $i \rightarrow j$.

The following theorem generalizes point 1 of theorem 1.10.

**Theorem 1.13.** For all $i, j \in S$, we have:

$$i \rightarrow j \text{ and } i \text{ recurrent } \implies j \text{ recurrent and } f_{i,j} = f_{j,i} = 1.$$

**Proof.** If $i = j$ then the result is trivial. Let $i$ and $j$ be two states such that $i \neq j$ and assume that $i \rightarrow j$ and that $i$ is recurrent. From lemma 1.1, we have $f_{i,j} > 0$ and since $i$ is recurrent, we have $f_{i,i} = 1$. From theorem 1.11, we obtain, for all $\ell \geq 0$,
By taking the limit when \( \ell \) tends to infinity, we have:

\[
P\{ N_i = \infty \mid X_0 = j \} = f_{j,i}.
\]

Since the state \( i \) is recurrent, corollary 1.3 allows us to assert that \( N_i = \infty \), \( P_i \)-a.s., thus, for all \( m \geq 1 \), we also have \( N_{i,m} = \infty \), \( P_i \)-a.s., where \( N_{i,m} \) is defined in the proof of theorem 1.11. Using the same reasoning as the one employed in the proof of theorem 1.11, we have:

\[
1 = P\{ \tau(j) = \infty \mid X_0 = i \} + \sum_{k=1}^{\infty} P\{ \tau(j) = k \mid X_0 = i \}
\]

\[
= 1 - f_{i,j} + \sum_{k=1}^{\infty} P\{ \tau(j) = k, N_{i,k+1} = \infty \mid X_0 = i \}
\]

\[
= 1 - f_{i,j} + \sum_{k=1}^{\infty} f_{i,j}^{(k)} P\{ N_{i,k+1} = \infty \mid \tau(j) = k, X_0 = i \}
\]

\[
= 1 - f_{i,j} + \sum_{k=1}^{\infty} f_{i,j}^{(k)} P\{ N_i = \infty \mid X_0 = j \}
\]

\[
= 1 - f_{i,j} + f_{i,j} P\{ N_i = \infty \mid X_0 = j \}
\]

\[
= 1 - f_{i,j} + f_{i,j} f_{j,i}.
\]

We then obtain:

\[
1 = 1 - f_{i,j} + f_{i,j} f_{j,i},
\]

that is:

\[
f_{i,j} = f_{i,j} f_{j,i}.
\]

Since \( f_{i,j} > 0 \), it follows that \( f_{j,i} = 1 \). Using lemma 1.1, we also obtain that \( j \rightarrow i \). Therefore, we have \( i \leftrightarrow j \) and \( i \) recurrent. Theorem 1.10 then states that \( j \) is recurrent. We finally obtain \( j \neq i \), \( j \rightarrow i \) and \( j \) recurrent. The same approach by interchanging the roles of \( i \) and \( j \), gives us \( f_{i,j} = 1 \).
**Definition 1.9.**—A non-empty subset $C$ of states of $S$ is said to be closed if for all $i \in C$ and, for all $j \notin C$, we have $P_{i,j} = 0$.

Recall that every subset $C$ of states of $S$, we have:

$$\sum_{j \in C} (P^n)_{i,j} = P\{X_n \in C \mid X_0 = i\}.$$  

**Lemma 1.2.**—Let $C$ be a non-empty subset of states of $S$. If $C$ is closed then, for all $i \in C$ and $n \geq 0$, we have:

$$\sum_{j \in C} (P^n)_{i,j} = 1.$$  

**Proof.**—The property always holds for $n = 0$ since $P^0 = I$, the identity matrix. For $n = 1$, since $C$ is closed, we have, for all $i \in C$ and $j \notin C$, $P_{i,j} = 0$, and thus:

$$\sum_{j \in C} P_{i,j} = \sum_{j \in S} P_{i,j} = 1.$$  

Let us assume that the property holds for the integer $n - 1$, with $n \geq 1$. We then have, using Fubini’s theorem, for all $i \in C$, since $P_{i,m} = 0$ if $m \notin C$,

$$\sum_{j \in C} (P^n)_{i,j} = \sum_{j \in C} \sum_{m \in S} P_{i,m} (P^{n-1})_{m,j}$$

$$= \sum_{j \in C} \sum_{m \in C} P_{i,m} (P^{n-1})_{m,j}$$

$$= \sum_{m \in C} P_{i,m} \sum_{j \in C} (P^{n-1})_{m,j}$$

$$= 1,$$  

which completes the proof.  

Therefore, a subset of states is closed if the chain cannot get out of it. In an equivalent manner, a non-empty subset $C$ of states of $S$ is closed if the submatrix of $P$ containing the transition probabilities between states of $C$ is stochastic. By extension, we will say that the state space $S$ is itself closed.

**Definition 1.10.**—A non-empty subset $C$ of states of $S$ is said to be irreducible if, for all $i, j \in C$, we have $i \leftrightarrow j$. This subset is called recurrent (respectively transient) if all its states are recurrent (respectively transient).

By definition, the equivalence classes of $X$ are irreducible sets.
THEOREM 1.14.– If $i$ is a recurrent state then its equivalence class $C(i)$ is closed.

PROOF.– Let $i$ be a recurrent state. Its equivalence class is defined by:
$$C(i) = \{ j \in S \mid i \leftrightarrow j \}.$$ 

From theorem 1.10 we see that the class $C(i)$ is recurrent. Let us assume that $C(i)$ is not closed. Consequently, there exists a state $k \in C(i)$ and a state $j \notin C(i)$ such that $P_{k,j} > 0$, which means that $k \rightarrow j$. The state $k$, being recurrent, we obtain, using theorem 1.13, that $f_{j,k} = 1$. Lemma 1.1 then states that $j \rightarrow k$, which means that we have $k \leftrightarrow j$, and thus $j \in C(i)$, which contradicts the hypothesis. The class $C(i)$ is therefore closed.

THEOREM 1.15.– Let $i$ be a transient state. We have:
$$C(i) \text{ finite } \implies C(i) \text{ not closed}.$$

PROOF.– Let $i$ be a transient state. From theorem 1.10, we see that its equivalence class $C(i)$ is transient. Let us assume that $C(i)$ is closed. We then have, for all $k \in C(i)$, by lemma 1.2,
$$\sum_{\ell \in C(i)} (P^n)_{k,\ell} = 1.$$

The class $C(i)$ being transient, we have, from corollary 1.1, for all $k, \ell \in C(i)$,
$$\sum_{n=1}^{\infty} (P^n)_{k,\ell} < \infty$$
and, since $C(i)$ is finite, we have:
$$\sum_{\ell \in C(i)} \sum_{n=1}^{\infty} (P^n)_{k,\ell} < \infty.$$

However, using Fubini’s theorem, we obtain:
$$\sum_{\ell \in C(i)} \sum_{n=1}^{\infty} (P^n)_{k,\ell} = \sum_{n=1}^{\infty} \sum_{\ell \in C(i)} (P^n)_{k,\ell} = \sum_{n=1}^{\infty} 1 = \infty,$$
which leads to a contradiction. Therefore, if $C(i)$ is finite, $C(i)$ cannot be closed.

From this we can deduce that the state space $S$ is composed of equivalence classes, by definition irreducible, which can be transient or recurrent. Recurrent classes are necessarily closed whereas transient classes can be closed or non-closed. Only infinite transient classes can be closed, while finite transient classes are necessarily non-closed.
The general structure of a transition probability matrix $P$ can thus be represented by regrouping the set of transient states $B$, which can contain many classes, and the set of recurrent states $C$. We decompose the set of recurrent states $C$ into recurrent classes $C_1, C_2, \ldots, C_j, \ldots$ which are necessarily closed, as shown in theorem 1.14. The form of $P$ is thus:

$$P = \begin{pmatrix}
  P_B & P_{B,C_1} & P_{B,C_2} & \cdots & P_{B,C_j} & \cdots \\
  0 & P_{C_1} & 0 & \cdots & 0 & \cdots \\
  0 & 0 & P_{C_2} & \cdots & 0 & \cdots \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\
  0 & 0 & 0 & \cdots & P_{C_j} & 0 \\
  \vdots & \vdots & \vdots & \ddots & 0 & \ddots 
\end{pmatrix},$$

where $P_{C_j}$ (respectively $P_B$) is the transition probability matrix between states of $C_j$ (respectively $B$) and $P_{B,C_j}$ is the transition probability submatrix from the states of $B$ to the states of $C_j$. If matrices $P_{B,C_j}$ are all null then the matrix $P$ is a block diagonal matrix and each matrix $P_B, P_{C_1}, \ldots, P_{C_j}, \ldots$ is the transition probability matrix of a Markov chain that is transient for the states of $B$ and recurrent for the states of $C_j$, $j \geq 1$.

1.7. Irreducible and recurrent Markov chains

**Theorem 1.16.–** If $X$ is an irreducible and recurrent Markov chain then for all $i, j \in S$, we have $f_{i,j} = 1$ and $\Pr\{\tau(j) < \infty \} = 1$.

**Proof.–** If chain $X$ is irreducible and recurrent, we have, from theorem 1.13, $f_{i,j} = \Pr\{\tau(j) < \infty \mid X_0 = i\} = 1$ for all $i, j \in S$. We then deduce that, for all $j \in S$,

$$\Pr\{\tau(j) < \infty \} = \sum_{i \in S} \Pr\{X_0 = i\} \Pr\{\tau(j) < \infty \mid X_0 = i\} = \sum_{i \in S} \Pr\{X_0 = i\} = 1,$$

which completes the proof. 

**Definition 1.11.–** We call measure on $S$ every row vector $v = (v_j, j \in S)$ such that $0 \leq v_j < \infty$. We say that the Markov chain $X$ has an invariant measure $v$ if $v$ is a measure on $S$ and if $v = vP$. The measure $v$ is said to be positive if $v_j > 0$, for all $j \in S$.

**Theorem 1.17.–** If the Markov chain $X$ is irreducible and recurrent then it has, up to a multiplicative constant, a unique positive invariant measure.
PROOF.—Existence. We denote by $\gamma^i_j$ the average number of visits to state $j$, starting from state $i$, until the first return to state $i$, that is:

$$\gamma^i_j = \mathbb{E}\left\{ \sum_{n=1}^{\tau(i)} 1\{X_n = j\} \mid X_0 = i \right\}$$

and we define the row vector $\gamma_i = (\gamma^i_j, j \in S)$. By definition of $\tau(i)$, we clearly have $\gamma^i_i = 1$. From theorem 1.16, we have $\tau(i) < \infty$, $\mathbb{P}$-a.s. We then obtain, using Fubini’s theorem,

$$\gamma^i_j = \sum_{k=1}^{\infty} \mathbb{E}\left\{ \sum_{n=1}^{\tau(i)} 1\{X_n = j\} \mid X_0 = i \right\}$$

$$= \sum_{k=1}^{\infty} \mathbb{P}\{X_n = j, \tau(i) = k \mid X_0 = i\}$$

$$= \sum_{n=1}^{\infty} \mathbb{P}\{X_n = j, \tau(i) \geq n \mid X_0 = i\} \quad [1.5]$$

$$= \sum_{\ell \in S} \sum_{n=1}^{\infty} \mathbb{P}\{X_{n-1} = \ell, \tau(i) \geq n, X_n = j \mid X_0 = i\}$$

$$= \sum_{\ell \in S} \sum_{n=1}^{\infty} \mathbb{P}\{X_{n-1} = \ell, \tau(i) > n - 1, X_0 = i\} \times \mathbb{P}\{X_{n-1} = \ell, \tau(i) \geq n \mid X_0 = i\}$$

$$= \sum_{\ell \in S} \sum_{n=1}^{\infty} \mathbb{P}\{X_{n-1} = \ell, \tau(i) \geq n \mid X_0 = i\} P_{\ell,j}, \quad [1.6]$$

where the last equality uses the homogeneity of $X$ and the Markov property, since $\tau(i)$ is a stopping time. Since $X_{\tau(i)} = i$, we have, when $X_0 = i$,

$$\sum_{n=1}^{\tau(i)} 1\{X_n = j\} = \sum_{n=0}^{\tau(i)-1} 1\{X_n = j\}.$$
We then get, again using Fubini’s theorem,
\[
\gamma^i_j = \mathbb{E}\left\{ \sum_{n=0}^{\tau(i)-1} 1_{\{X_n=j\}} \mid X_0 = i \right\}
\]
\[
= \sum_{k=1}^{\infty} \mathbb{E}\left\{ \left( \sum_{n=0}^{\tau(i)-1} 1_{\{X_n=j\}} \right) \mid X_0 = i \right\}
\]
\[
= \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} \mathbb{P}\{X_n = j, \tau(i) = k \mid X_0 = i\}
\]
\[
= \sum_{k=1}^{\infty} \sum_{n=1}^{k} \mathbb{P}\{X_{n-1} = j, \tau(i) = k \mid X_0 = i\}
\]
\[
= \sum_{n=1}^{\infty} \mathbb{P}\{X_{n-1} = j, \tau(i) \geq n \mid X_0 = i\}.
\]

By combining relations [1.6] and [1.7], we obtain:
\[
\gamma^i_j = \sum_{\ell \in S} \gamma^i_{\ell} P_{\ell,j},
\]
that is:
\[
\gamma^i = \gamma^i P.
\]

Moreover, by the irreducibility of \(X\), there exist two integers \(n\) and \(m\) such that \((P^n)_{i,j} > 0\) and \((P^m)_{j,i} > 0\) and since, by definition, \(\gamma^i_i = 1\), we have:
\[
0 < (P^n)_{i,j} = \gamma^i_j (P^n)_{i,j} \leq \sum_{\ell \in S} \gamma^i_{\ell} (P^n)_{\ell,j} = (\gamma^i P^n)_{j} = \gamma^i_j
\]
and
\[
\gamma^i_j (P^m)_{j,i} \leq \sum_{\ell \in S} \gamma^i_{\ell} (P^m)_{\ell,i} = (\gamma^i P^m)_{i} = \gamma^i_i = 1,
\]
hence \(\gamma^i_j \leq 1/(P^m)_{j,i} < \infty\). Therefore, \(\gamma^i\) is a positive invariant measure.
– Uniqueness. Let \( i \in S \) and let \( \lambda \) be an invariant measure such that \( \lambda_i = 1 \). For all \( N \geq 1 \), we have, since \( \lambda = \lambda P \), by iterating the induction below, for all \( j \in S \),

\[
\lambda_j = P_{i,j} + \sum_{i_1 \neq i} \lambda_{i_1} P_{i_1,j}
\]

\[
= P_{i,j} + \sum_{i_1 \neq i} P_{i,i_1} P_{i_1,j} + \sum_{i_1 \neq i_2 \neq i} \lambda_{i_2} P_{i_2,i_1} P_{i_1,j}
\]

\[= \cdots
\]

\[
= P_{i,j} + \sum_{n=1}^{N} \sum_{i_1 \neq i_1 \ldots, i_n \neq i} P_{i,i_{n-1}} P_{i_{n-1},i_{n-2}} \cdots P_{i_1,j}
\]

\[
\geq P_{i,j} + \sum_{n=1}^{N} \sum_{i_1 \neq i_1 \ldots, i_n \neq i} P_{i,i_{n-1}} P_{i_{n-1},i_{n-2}} \cdots P_{i_1,j}
\]

\[
= P_{i,j} + \sum_{n=1}^{N} P \{ X_{n+1} = j, X_{\ell} \neq i, 1 \leq \ell \leq n \mid X_0 = i \}
\]

\[
= P_{i,j} + \sum_{n=1}^{N} P \{ X_{n+1} = j, \tau(i) \geq n + 1 \mid X_0 = i \}
\]

\[
= \sum_{n=0}^{N} P \{ X_{n+1} = j, \tau(i) \geq n + 1 \mid X_0 = i \}
\]

\[
= \sum_{n=1}^{N+1} P \{ X_n = j, \tau(i) \geq n \mid X_0 = i \}.
\]

When \( N \) tends to infinity and using relation [1.5], we obtain:

\[
\lambda_j \geq \sum_{n=1}^{\infty} P \{ X_n = j, \tau(i) \geq n \mid X_0 = i \} = \gamma_j^i.
\]
Thus we have shown that the measure $\mu = \lambda - \gamma^i$ is an invariant measure that satisfies $\mu_i = 0$. Since the Markov chain is irreducible, for all $j \in S$, there exists an integer $m$ such that $(P^m)_{j,i} > 0$. We then have, for all $j \in S$,

$$0 \leq \mu_j(P^m)_{j,i} \leq \sum_{\ell \in S} \mu_{\ell}(P^m)_{\ell,i} = (\mu P^m)_i = \mu_i = 0.$$ 

This yields $\mu_j = 0$, for all $j \in S$, that is $\lambda = \gamma^i$.

Let us recall that a state $i$ is recurrent if $f_{i,i} = \mathbb{P}\{\tau(i) < \infty \mid X_0 = i\} = 1$. For every state $i$, we denote by $m_i$ the expected return time to state $i$, that is:

$$m_i = \mathbb{E}\{\tau(i) \mid X_0 = i\}.$$ 

By definition of $\tau(i)$, we have:

$$m_i = \begin{cases} \sum_{n=1}^{\infty} n f_{i,i}^{(n)} & \text{if } i \text{ is recurrent,} \\ \infty & \text{if } i \text{ is transient.} \end{cases}$$

From relation [1.4], we also have, using Fubini’s theorem and the monotone convergence theorem,

$$m_i = \sum_{j \in S} \gamma_j^i. \quad [1.8]$$

**Definition 1.12.**— A recurrent state $i$ is said to be positive recurrent if $m_i < \infty$ and null recurrent if $m_i = \infty$. If all the states of a Markov chain are positive recurrent (respectively null recurrent), then the Markov chain is said to be positive recurrent (respectively null recurrent).

Note that if a state $i$ is such that $m_i < \infty$ then we have $f_{i,i} = 1$, that is state $i$ is recurrent.

We denote by $\mathbb{I}$ the column vector whose components are all equal to 1. Its dimension is determined by the context in which it is used.

**Definition 1.13.**— An invariant probability on $S$ is an invariant measure $\nu = (v_j, \ j \in S)$ such that $v \mathbb{I} = 1$. The invariant probability $\nu$ is said to be positive if $v_j > 0$, for all $j \in S$. 
Corollary 1.4.– Let X be an irreducible Markov chain. X is positive recurrent if and only if it has an invariant probability. In this case, the invariant probability is unique and positive, we denote it by \( \pi = (\pi_j, \ j \in S) \) and it is given by:

\[
\pi_j = 1/m_j.
\]

Proof.– Let X be an irreducible Markov chain. If X is positive recurrent then, by definition, we have \( m_j < \infty \), for all \( j \in S \), and from theorem 1.17, it has, up to a multiplicative constant, a unique positive invariant measure. Defining:

\[
\pi_j = \frac{\gamma_j^i}{m_i},
\]

the measure \( \pi = (\pi_j, \ j \in S) \) is, from relation [1.8], the unique invariant probability of X and it is, moreover, positive. Finally, \( \pi_j \) is independent of \( i \) because all measures \( \gamma^i \) are proportional. Taking \( i = j \), we have:

\[
\pi_j = \frac{\gamma_j^j}{m_j} = \frac{1}{m_j}.
\]

Conversely, let \( \pi \) be an invariant probability. We then have, for all \( j \in S \) and for all \( n \geq 0 \),

\[
\pi_j = \sum_{i \in S} \pi_i (P^n)_{i,j}.
\]

Let us assume that X is transient. Then, letting \( n \) tend to infinity, using corollary 1.1 and the dominated convergence theorem, we obtain \( \pi_j = 0 \), for all \( j \in S \), which contradicts the fact that \( \pi \) is a probability. This proves that X is recurrent. From theorem 1.17, X has, up to a multiplicative constant, a unique positive invariant measure. This invariant measure is, therefore, proportional to \( \pi \), which shows that \( \pi \) is unique and positive.

For all \( i \in S \), the measure \( \lambda^i = (\lambda_j^i, \ j \in S) \), given by:

\[
\lambda_j^i = \frac{\pi_j}{\pi_i},
\]
is a positive invariant measure such that \( \lambda_i^i = 1 \). Hence, using relation [1.8] and the first step of the proof of uniqueness in theorem 1.17 we obtain that:

\[
m_i = \sum_{j \in S} \gamma_j^i \leq \sum_{j \in S} \lambda_j^i = \sum_{j \in S} \frac{\pi_j}{\pi_i} = \frac{1}{\pi_i} < \infty.
\]

Thus every state \( i \in S \) is positive recurrent, which means that the Markov chain is positive recurrent.

For all \( i, j \in S \), we define \( m_{i,j} = \mathbb{E}\{\tau(j) \mid X_0 = i\} \), which is the expected hitting time of state \( j \), starting from state \( i \). We then have \( m_{i,i} = m_i \). These expected hitting times are given by the following theorem.

**Theorem 1.18.** For all \( i, j \in S \), we have:

\[
m_{i,j} = 1 + \sum_{k \in S, k \neq j} P_{i,k} m_{k,j}. \tag{1.9}
\]

**Proof.** For all \( i, j \in S \), we have, by conditioning with respect to \( X_1 \) and using the Markov property,

\[
m_{i,j} = \mathbb{E}\{\tau(j) \mid X_0 = i\}
= \sum_{k \in S} P_{i,k} \mathbb{E}\{\tau(j) \mid X_1 = k, X_0 = i\}
= P_{i,j} \mathbb{E}\{\tau(j) \mid X_1 = j, X_0 = i\} + \sum_{k \in S, k \neq j} P_{i,k} \mathbb{E}\{\tau(j) \mid X_1 = k, X_0 = i\}
= P_{i,j} + \sum_{k \in S, k \neq j} P_{i,k} [1 + \mathbb{E}\{\tau(j) \mid X_0 = k\}]
= 1 + \sum_{k \in S, k \neq j} P_{i,k} m_{k,j},
\]

which completes the proof.

**Theorem 1.19.** Let \( X \) be a Markov chain on the state space \( S \) with transition probability matrix \( P \). Let \( \pi \) be a measure on \( S \). The two following statements are equivalent:

1) \( \pi = \pi P \).

2) For every partition \( A, B \) of \( S \), we have \( \sum_{i \in A} \pi_i \sum_{j \in B} P_{i,j} = \sum_{i \in B} \pi_i \sum_{j \in A} P_{i,j} \).
PROOF.– Let \( \pi \) be a measure on \( S \). Let us assume that \( \pi = \pi P \). Thus we have, for all \( j \in S \),
\[
\pi_j = \sum_{i \in S} \pi_i P_{i,j}.
\]

Let \( A, B \) be a partition of \( S \), that is two non-empty sets such that \( A \cap B = \emptyset \) and \( A \cup B = S \). Summing over \( j \in A \), we obtain, using Fubini’s theorem,
\[
\sum_{j \in A} \pi_j = \sum_{i \in S} \pi_i \sum_{j \in A} P_{i,j},
\]
or also, since \( A \) and \( B \) form a partition of \( S \),
\[
\sum_{j \in A} \pi_j = \sum_{i \in A} \pi_i \sum_{j \in A} P_{i,j} + \sum_{i \in B} \pi_i \sum_{j \in A} P_{i,j}.
\]

Moreover, since \( P \) is a stochastic matrix, we have:
\[
\sum_{j \in A} P_{i,j} = 1 - \sum_{j \in B} P_{i,j}.
\]

Hence:
\[
\sum_{j \in A} \pi_j = \sum_{i \in A} \pi_i \left( 1 - \sum_{j \in B} P_{i,j} \right) + \sum_{i \in B} \pi_i \sum_{j \in A} P_{i,j},
\]
that is:
\[
\sum_{i \in A} \pi_i \sum_{j \in B} P_{i,j} = \sum_{i \in B} \pi_i \sum_{j \in A} P_{i,j}.
\]

Conversely, by taking successively, for all \( \ell \in S \), \( A = \{ \ell \} \) and \( B = S \setminus \{ \ell \} \), we obtain:
\[
\pi_\ell \sum_{j \in S \setminus \{ \ell \}} P_{\ell,j} = \sum_{i \in S \setminus \{ \ell \}} \pi_i P_{i,\ell},
\]
that is:

$$\pi_\ell (1 - P_{\ell, \ell}) = \sum_{i \in S \setminus \{\ell\}} \pi_i P_{i, \ell},$$

thus, for all $\ell \in S$,

$$\pi_\ell = \sum_{i \in S} \pi_i P_{i, \ell},$$

hence $\pi = \pi P$. 

1.8. Aperiodic Markov chains

Let $E$ be a set of positive integers. We denote by $\gcd(E)$ the greatest common divisor of $E$, that is the largest integer that divides all integers of $E$. Let $\{u_n, \ n \geq 1\}$ be a sequence of positive integers. For $k \geq 1$, the sequence $d_k = \gcd\{u_1, \ldots, u_k\}$ is decreasing and is lower bounded by 1, therefore, it converges to a limit $d \geq 1$ called the gcd of the sequence $\{u_n, \ n \geq 1\}$. Since the numbers $d_k$ are integers, the limit $d$ is reached in a finite number of steps, thus there exists a positive integer $k_0$ such that $d = \gcd\{u_1, \ldots, u_k\}$, for all $k \geq k_0$.

**Definition 1.14.** The period $d(i)$ of a state $i \in S$ is defined by:

$$d(i) = \gcd\{n \geq 1 \mid (P^n)_{i,i} > 0\},$$

using the convention $d(i) = 0$ if $(P^n)_{i,i} = 0$, for all $n \geq 1$. If $d(i) = 1$ then the state $i$ is said to be aperiodic.

**Theorem 1.20.** If $i \leftrightarrow j$ then $d(i) = d(j)$.

**Proof.** If $i = j$ then the result is trivial, therefore, we assume that $i \neq j$. If $i \leftrightarrow j$ then, since $i \neq j$, there exist integers $\ell \geq 1$ and $m \geq 1$ such that $(P^\ell)_{i,j} > 0$ and $(P^m)_{j,i} > 0$. Thus we have:

$$(P^{\ell+m})_{i,i} = \sum_{k \in S} (P^\ell)_{i,k} (P^m)_{k,i} \geq (P^\ell)_{i,j} (P^m)_{j,i} > 0,$$
which shows that \( \ell + m \) is a multiple of \( d(i) \). Moreover, we also have, in the same way:

\[
(P^{\ell+m})_{j,j} = \sum_{k \in S} (P^m)_{j,k} (P^\ell)_{k,j} \geq (P^m)_{j,i} (P^\ell)_{i,j} > 0,
\]

which shows that \( \ell + m \) is a multiple of \( d(j) \).

Let \( \mathbb{N}^* \) be the set of positive integers. The set \( d(i)\mathbb{N}^* \) is then the set of all the multiples of \( d(i) \). Let \( E_i = \{ n \geq 1 \mid (P^n)_{i,i} > 0 \} \). By definition, we have \( d(i) = \gcd(E_i) \) and if \( n \in E_i \) then \( n \in d(i)\mathbb{N}^* \). We then have \( E_i \subseteq d(i)\mathbb{N}^* \) and \( d(i) = \gcd(d(i)\mathbb{N}^*) \).

We have just seen that \( E_i \neq \emptyset \), since \( \ell + m \in E_i \). Let \( n \in E_i \). We have, since \( (P^n)_{i,i} > 0 \),

\[
(P^{n+\ell+m})_{j,j} \geq (P^m)_{j,i} (P^n)_{i,i} (P^\ell)_{i,j} > 0,
\]

therefore, \( n + \ell + m \in E_j \) and also \( n + \ell + m \in d(j)\mathbb{N}^* \). However, we have seen that \( \ell + m \in d(j)\mathbb{N}^* \) so we have \( n \in d(j)\mathbb{N}^* \). Thus we have shown that \( E_i \subseteq d(j)\mathbb{N}^* \), therefore, \( d(j) \leq d(i) \).

The roles played by states \( i \) and \( j \) being symmetric, we also have \( d(i) \leq d(j) \), that is \( d(i) = d(j) \).

It follows that periodicity is a class property, that is all the states of the same class have the same period.

**Definition 1.15.**— A Markov chain is said to be aperiodic if all its states have the same period equal to 1.

We note, in particular, that if a state \( i \) is such that \( P_{i,i} > 0 \) then \( i \) is aperiodic.

Let us now recall some results on the \( \gcd \) of positive integers. We denote by \( \mathbb{Z} \) the set of integers.

**Definition 1.16.**— We say that a non-empty subset \( I \) of \( \mathbb{Z} \) is an ideal of \( \mathbb{Z} \) if the following two properties are satisfied:

- If \( x \in I \) and \( y \in I \) then \( x + y \in I \).
- If \( x \in I \) and \( \lambda \in \mathbb{Z} \) then \( \lambda x \in I \).
Let us observe that if $I$ is an ideal of $\mathbb{Z}$, then $0 \in I$ and that if $x \in I$ then $-x \in I$. For all $a \in \mathbb{Z}$, we denote by $a\mathbb{Z}$ the set of all multiples of $a$, that is $a\mathbb{Z} = \{\lambda a, \ \lambda \in \mathbb{Z}\}$. It is easy to check whether $a\mathbb{Z}$ is an ideal of $\mathbb{Z}$.

**Definition 1.17.** An ideal $I$ of $\mathbb{Z}$ is said to be principal if there exists a unique integer $a \geq 0$ such that $I = a\mathbb{Z}$.

**Lemma 1.3.** Every ideal of $\mathbb{Z}$ is principal.

**Proof.** Let $I$ be an ideal of $\mathbb{Z}$. We have to show that there exists a unique integer $a \geq 0$ such that $I = a\mathbb{Z}$.

If $I = \{0\}$ then it is clear that $I = 0\mathbb{Z}$, therefore, $I$ is principal. It is now assumed that $I \neq \{0\}$.

- **Existence:** since $I \neq \{0\}$, $I$ contains positive elements since if $x \in I$ then $-x \in I$. We denote by $a$ the smallest positive element of $I$. By definition of an ideal, the multiples of one of its elements are also its elements. Therefore, we have $a\mathbb{Z} \subseteq I$. Conversely, if $x \geq 0$ and $x \in I$ then the Euclidean division of $x$ by $a$ leads to:

\[ x = aq + r, \text{ where } 0 \leq r \leq a - 1. \]

Since $x \in I$ and $-aq \in I$ we have $r = x - aq \in I$ which means that $r = 0$ since $a$ is the smallest positive element of $I$. Thus, we have $x = aq$, that is $x \in a\mathbb{Z}$. Finally, if $x \leq 0$ and $x \in I$ then $-x \in I$. It follows from the above that we have $-x \in a\mathbb{Z}$ and, therefore, $x \in a\mathbb{Z}$, hence $I \subseteq a\mathbb{Z}$. Thus we have shown that $I = a\mathbb{Z}$.

- **Uniqueness:** if $a$ and $b$ are two positive integers such that $a\mathbb{Z} = b\mathbb{Z}$ then $a$ and $b$ are multiples of each other, and thus $a = b$. \[\blacksquare\]

**Lemma 1.4.** Let $n_1, \ldots, n_k$ be positive integers. There is a common divisor $d$ to these $k$ integers, of the form $d = \lambda_1 n_1 + \cdots + \lambda_k n_k$, where $\lambda_1, \ldots, \lambda_k$ are integers. Such a common divisor is a multiple of any other common divisor to these $k$ integers. It is the gcd of the integers $n_1, \ldots, n_k$.

**Proof.** Let $I = \{\lambda_1 n_1 + \cdots + \lambda_k n_k, \ \lambda_1, \ldots, \lambda_k \in \mathbb{Z}\}$ be the set of all linear combinations, with coefficients in $\mathbb{Z}$, of the integers $n_1, \ldots, n_k$. It is easy to check that $I$ is an ideal of $\mathbb{Z}$. From lemma 1.3, it is a principal ideal, therefore, there exists a unique integer $d \geq 0$ such that $I = d\mathbb{Z}$. In particular, $d \in I$, thus $d$ is of the form $d = \lambda_1 n_1 + \cdots + \lambda_k n_k$. \[\blacksquare\]
However, for all \( j = 1, \ldots, k \), taking \( \lambda_j = 1 \) and the other \( \lambda_\ell \) equal to 0, we obtain \( n_j \in I \), which shows that, for all \( j = 1, \ldots, k \), \( n_j \) is a multiple of \( d \). It follows that \( d \) is a common divisor of all integers \( n_1, \ldots, n_k \).

Let \( x \) be another common divisor of the integers \( n_1, \ldots, n_k \). The integers \( n_1, \ldots, n_k \) and, therefore, also \( \lambda_1 n_1 + \cdots + \lambda_k n_k \) are multiples of \( x \), for all \( \lambda_1, \ldots, \lambda_k \). Thus \( d \) is also a multiple of \( x \), which leads to \( d = \gcd\{n_1, \ldots, n_k\} \).

**LEMMA 1.5.**— Let \( n_1, \ldots, n_k \) be positive integers and \( d \) be their gcd. There exists an integer \( N > 0 \) such that \( n \geq N \) implies that there exist non-negative integers \( c_1, \ldots, c_k \) such that \( nd = c_1 n_1 + \cdots + c_k n_k \).

**PROOF.**— From lemma 1.4, there exist integers \( \lambda_1, \ldots, \lambda_k \) such that \( d = \lambda_1 n_1 + \cdots + \lambda_k n_k \). Defining:

\[
N_1 = \sum_{\{j|\lambda_j > 0\}} \lambda_j n_j \quad \text{and} \quad N_2 = \sum_{\{j|\lambda_j < 0\}} (-\lambda_j) n_j,
\]

we have \( d = N_1 - N_2 \), with \( N_1 > 0 \) and \( N_2 \geq 0 \) since \( d \geq 1 \). In the proof of lemma 1.4, we have also shown that the integers \( n_1, \ldots, n_k \) are multiples of \( d \). Therefore, \( N_1 \) and \( N_2 \) are also multiples of \( d \). Let \( N \) be the positive integer defined by \( N = N_2^2/d \).

The integer \( n \geq N \) can be written as \( n = N + \ell \) and the Euclidean division of \( \ell \) by \( N_2/d \) allows us to write \( \ell = \delta N_2/d + b \) with \( \delta \geq 0 \) and \( 0 \leq b < N_2/d \). We then obtain:

\[
nd = Nd + \ell d = N_2^2 + \delta N_2 + bd
\]

and by replacing \( d \) by \( N_1 - N_2 \) on the right hand side, we can write:

\[
nd = (N_2 - b + \delta)N_2 + bN_1.
\]

Since \( d \geq 1 \) and \( b < N_2/d \), we have \( N_2 - b > 0 \), which shows, using the expressions of \( N_1 \) and \( N_2 \) given in [1.10], that \( nd \) is written as a linear combination with positive coefficients of the integers \( n_j \), for all \( n \geq N \).

**THEOREM 1.21.**— If \( d(i) \) is the period of the state \( i \) then there exists a positive integer \( N \) such that for all \( n \geq N \), we have:

\[
(P^{nd(i)})_{i,i} > 0.
\]
If \( d(i) = 0 \) then the result is true for any value of \( N \). We assume, therefore, that \( d(i) \geq 1 \). By definition of the period of a state, there exists a finite number of positive integers \( n_\ell, \ell = 1, \ldots, k \) such that \((P^{n_\ell})_{i,i} > 0 \) and \( d(i) = \gcd\{n_1, \ldots, n_k\} \). From lemma 1.5, there exists an integer \( N > 0 \) such that \( n \geq N \) implies that there exist non-negative integers \( c_1, \ldots, c_k \) such that \( nd(i) = c_1 n_1 + \cdots + c_k n_k \). Therefore, we have, for all \( n \geq N \),

\[
(P^{nd(i)})_{i,i} = (P^{c_1 n_1 + \cdots + c_k n_k})_{i,i} \geq \prod_{\ell=1}^{k} (P^{c_1 n_1})_{i,i} \geq \prod_{\ell=1}^{k} ((P^{n_\ell})_{i,i})^{c_\ell} > 0,
\]

which completes the proof.

### 1.9. Convergence to equilibrium

**Theorem 1.22.** Let \( X \) be an irreducible and aperiodic Markov chain with an invariant probability denoted by \( \pi \). For all \( j \in S \), we have:

\[
\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = \pi_j,
\]

for every initial distribution. In particular, for all \( i, j \in S \), we have:

\[
\lim_{n \to \infty} (P^n)_{i,j} = \pi_j.
\]

**Proof.** Let us point out that, from corollary 1.4, the invariant probability \( \pi \) is unique and positive. We denote by \( \alpha \) the initial distribution of \( X \). The proof is based on a coupling argument. Let \( Y = \{Y_n, \ n \in \mathbb{N}\} \) be a Markov chain on the same state space \( S \) as \( X \), with initial distribution \( \pi \), with the same transition probability matrix \( P \) as \( X \) and independent of \( X \). Let \( 0 \) be an arbitrary state of \( S \). We define:

\[
T = \inf\{n \geq 1 \mid X_n = Y_n = \ell\}.
\]

- Step 1. Let us show that \( \mathbb{P}\{T < \infty\} = 1 \).

The Markov chains \( X \) and \( Y \) being independent, the process \( W = \{W_n, \ n \in \mathbb{N}\} \), defined by \( W_n = (X_n, Y_n) \), is a Markov chain on \( S \times S \) with initial distribution \( \beta \) given by:

\[
\beta_{i,k} = \alpha_i \pi_k,
\]
and with transition probability matrix $\tilde{P}$ given by:

$$\tilde{P}_{(i,k),(j,l)} = P_{i,j}P_{k,l}.$$  

We can easily show that the Markov chain $W$ has a positive invariant probability $\tilde{\pi}$ given by:

$$\tilde{\pi}_{(i,k)} = \pi_i \pi_k.$$  

It is also simple to see, by induction, that for all $n \geq 0$, we have:

$$(\tilde{P}^n)_{(i,k),(j,l)} = (P^n)_{i,j}(P^n)_{k,l}.$$  

Let us also point out that we have $T = \inf\{n \geq 1 \mid W_n = (\ell, \ell)\}$. Let $i, j, k, l$ be four states of $S$. $X$ being irreducible, there exist integers $m$ and $h$ such that $(P^m)_{i,j} > 0$ and $(P^h)_{k,l} > 0$. The Markov chain $X$ being aperiodic, there exist, from theorem 1.21, integers $N_j$ and $N_l$ such that for all $n \geq N = \max\{N_j, N_l\}$, $(P^n)_{i,j} > 0$ and $(P^n)_{l,l} > 0$. Therefore, we have, for all $n \geq N$,

$$(\tilde{P}^{n+m+h})_{(i,k),(j,l)} = (P^{n+m+h})_{i,j}(P^{n+m+h})_{k,l} \\
\geq (P^n)_{i,j}(P^{n+h})_{j,j}(P^h)_{k,l}(P^{n+m})_{l,l} \\
> 0,$$

which shows that the Markov chain $W$ is irreducible. Since, in addition, it has an invariant probability $\tilde{\pi}$, corollary 1.4 allows us to conclude that $W$ is positive recurrent. Since $W$ is irreducible and recurrent, we have, from theorem 1.16, $\mathbb{P}\{T < \infty\} = 1$.

---

Step 2. Concatenation of chains $X$ and $Y$.

Since the time $T$ is finite $\mathbb{P}$-a.s. we define the process $Z = \{Z_n, n \in \mathbb{N}\}$, for all $n \in \mathbb{N}$, by:

$$Z_n = \begin{cases} X_n & \text{if } n < T, \\
Y_n & \text{if } n \geq T. \end{cases}$$

We show in this step that $Z$ is a Markov chain.
We clearly have, by definition of $T$, $Z_0 = X_0$. Let $n \geq 1$, $1 \leq k \leq n$ and $i_k, \ldots, i_n \in S$. We have:

$$
P\{Y_n = i_n, \ldots, Y_k = i_k \mid W_k = (\ell, \ell)\} = \sum_{h_n, \ldots, h_{k+1} \in S} P\{W_n = (h_n, i_n), \ldots, W_k = (\ell, i_k) \mid W_k = (\ell, \ell)\} = 1\{i_k = \ell\} \sum_{h_n, \ldots, h_{k+1} \in S} \tilde{P}_{(\ell, \ell), (h_{k+1}, i_{k+1})} \cdots \tilde{P}_{(h_{n-1}, i_{n-1}), (h_n, i_n)} = 1\{i_k = \ell\} P_{\ell, h_{k+1}} P_{\ell, i_{k+1}} \cdots P_{i_{n-1}, i_n}.
$$

Proceeding in the same way for the Markov chain $X$, we obtain:

$$
P\{Y_n = i_n, \ldots, Y_k = i_k \mid W_k = (\ell, \ell)\} = P\{X_n = i_n, \ldots, X_k = i_k \mid W_k = (\ell, \ell)\}. \quad [1.11]
$$

Using the fact that:

$$\{T = k\} = \{W_k = (\ell, \ell), W_{k-1} \neq (\ell, \ell), \ldots, W_1 \neq (\ell, \ell)\}$$

and that $\{T = k\} \subseteq \{W_k = (\ell, \ell)\}$, we have:

$$
P\{Z_n = i_n, \ldots, Z_0 = i_0, T = k\} = P\{Y_n = i_n, \ldots, Y_k = i_k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0, T = k\} = P\{Y_n = i_n, \ldots, Y_k = i_k, W_k = (\ell, \ell), T = k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\} = P\{Y_n = i_n, \ldots, Y_k = i_k \mid W_k = (\ell, \ell)\} \times P\{T = k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\} = P\{X_n = i_n, \ldots, X_k = i_k \mid W_k = (\ell, \ell)\} \times P\{T = k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\} = P\{X_n = i_n, \ldots, X_k = i_k, W_k = (\ell, \ell), T = k, X_{k-1} = i_{k-1}, \ldots, X_0 = i_0\} = P\{X_n = i_n, \ldots, X_0 = i_0, T = k\},
$$

where the third and antepenultimate equalities use the Markov property, since $T$ is a stopping time, and where the fourth uses relation [1.11]. This relation being true for all $1 \leq k \leq n$, we have:

$$
P\{Z_n = i_n, \ldots, Z_0 = i_0, T \leq n\} = P\{X_n = i_n, \ldots, X_0 = i_0, T \leq n\}.$$
In addition, by definition of $Z$, we have:

$$\mathbb{P}\{Z_n = i_n, \ldots, Z_0 = i_0, T > n\} = \mathbb{P}\{X_n = i_n, \ldots, X_0 = i_0, T > n\},$$

and, therefore,

$$\mathbb{P}\{Z_n = i_n, \ldots, Z_0 = i_0\} = \mathbb{P}\{X_n = i_n, \ldots, X_0 = i_0\},$$

which proves, from theorem 1.1, that $Z$ is, as $X$, a Markov chain with initial distribution $\alpha$ and transition probability matrix $P$.

– Step 3. Passage to the limit.

Thanks to Step 2, we have, for all $n \geq 0$, $\mathbb{P}\{X_n = j\} = \mathbb{P}\{Z_n = j\}$ and by definition of $Z$,

$$\mathbb{P}\{Z_n = j\} = \mathbb{P}\{X_n = j, T > n\} + \mathbb{P}\{Y_n = j, T \leq n\}.$$

Thus we obtain, since $\mathbb{P}\{Y_n = j\} = (\pi P^n)_j = \pi_j$,

$$|\mathbb{P}\{X_n = j\} - \pi_j| = |\mathbb{P}\{Z_n = j\} - \mathbb{P}\{Y_n = j\}|$$

$$= |\mathbb{P}\{X_n = j, T > n\} + \mathbb{P}\{Y_n = j, T \leq n\} - \mathbb{P}\{Y_n = j\}|$$

$$= |\mathbb{P}\{X_n = j, T > n\} - \mathbb{P}\{Y_n = j, T > n\}|$$

$$\leq \max\{\mathbb{P}\{X_n = j, T > n\}, \mathbb{P}\{Y_n = j, T > n\}\}$$

$$\leq \mathbb{P}\{T > n\},$$

and $\mathbb{P}\{T > n\}$ approaches 0 when $n$ approaches infinity because $T$ is finite $\mathbb{P}$-a.s. Therefore, we have shown that for every initial distribution $\alpha$, we have, for all $j \in S$,

$$\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = \pi_j.$$

By choosing for $\alpha$ the probability distribution $\delta^i$ concentrated on state $i$, defined, for all $j \in S$, by $\delta^i_j = 1_{i=j}$, we obtain:

$$\mathbb{P}\{X_n = j\} = (\delta^i P^n)_j = (P^n)_{i,j}.$$
Therefore, we have, for all \( i, j \in S \),

\[
\lim_{n \to \infty} (P^n)_{i,j} = \pi_j,
\]

which completes the proof.

The limiting probability distribution \( \pi \) is often called the stationary distribution of \( X \). The study of process \( X \) or a function of it at finite instants with an initial distribution different from the stationary distribution is referred to as transient regime analysis, and when time approaches infinity or when the initial distribution is equal to the stationary distribution, it is referred to as stationary regime analysis.

**Theorem 1.23.** Let \( X \) be an irreducible, aperiodic and null recurrent Markov chain. Then, for all \( j \in S \), we have:

\[
\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = 0,
\]

for every initial distribution. In particular, for all \( i, j \in S \), we have:

\[
\lim_{n \to \infty} (P^n)_{i,j} = 0.
\]

**Proof.** We return to the coupling argument used in the proof of theorem 1.22. We denote by \( \alpha \) the initial distribution of \( X \). Let \( Y = \{Y_n, \ n \in \mathbb{N}\} \) be a Markov chain on the same state space \( S \) as \( X \), with initial distribution \( \mu \), with the same transition probability matrix \( P \) as \( X \) and independent of \( X \). We have shown that the process \( W = \{W_n, \ n \in \mathbb{N}\} \) defined by \( W_n = (X_n, Y_n) \) is a Markov chain on \( S \times S \) and that the aperiodicity of \( X \) results in the irreducibility of \( W \).

If \( W \) is transient then, from corollary 1.1, we have, for all \( j \in S \),

\[
\lim_{n \to \infty} \mathbb{P}\{W_n = (j, j)\} = 0.
\]

Moreover, taking \( \mu = \alpha \), we have, for all \( j \in S \),

\[
[\mathbb{P}\{X_n = j\}]^2 = \mathbb{P}\{X_n = j, Y_n = j\} = \mathbb{P}\{W_n = (j, j)\},
\]

hence:

\[
\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = 0,
\]
which completes the proof in this case.

If \( W \) is recurrent then, from Step 3 of the proof of theorem 1.22, we have, for every initial distribution \( \alpha \) of \( X \) and \( \mu \) of \( Y \),

\[
\lim_{n \to \infty} |\mathbb{P}\{X_n = j\} - \mathbb{P}\{Y_n = j\}| = 0,
\]

and, for all \( k \geq 0 \),

\[
\lim_{n \to \infty} |\mathbb{P}\{X_{n-k} = j\} - \mathbb{P}\{Y_{n-k} = j\}| = 0.
\]

By taking \( \mu = \alpha P^k \), we have, for \( n \geq k \),

\[
\mathbb{P}\{Y_{n-k} = j\} = (\mu P^{n-k})_j = (\alpha P^n)_j = \mathbb{P}\{X_n = j\},
\]

hence, for all \( k \geq 0 \),

\[
\lim_{n \to \infty} |\mathbb{P}\{X_{n-k} = j\} - \mathbb{P}\{X_n = j\}| = 0. \tag{1.12}
\]

Moreover, since \( X \) is null recurrent, we have, for all \( j \in S \),

\[
m_j = \mathbb{E}\{\tau(j) \mid X_0 = j\} = \sum_{k=0}^{\infty} \mathbb{P}\{\tau(j) > k \mid X_0 = j\} = \infty.
\]

Therefore, for all \( \varepsilon > 0 \), there exists an integer \( K \geq 0 \) such that:

\[
\sum_{k=0}^{K} \mathbb{P}\{\tau(j) > k \mid X_0 = j\} \geq \frac{2}{\varepsilon}. \tag{1.13}
\]

If \( A_1, \ldots, A_\ell \) are events, we have:

\[
\mathbb{P}\{A_1\} + \mathbb{P}\{\overline{A_1} \cap A_2\} + \ldots + \mathbb{P}\{\overline{A_1} \cap \ldots \cap \overline{A_{\ell-1}} \cap A_\ell\} = \mathbb{P}\{\bigcup_{i=1}^{\ell} A_i\} \leq 1.
\]
By taking, for all \( n \geq K \) and for \( k = 1, \ldots, K + 1 \), \( A_k = \{ X_{n-k+1} = j \} \), we obtain:

\[
1 \geq \sum_{k=n-K}^{n} \mathbb{P}\{ X_{n} \neq j, X_{n-1} \neq j, \cdots, X_{k+1} \neq j, X_{k} = j \}
\]

\[
= \sum_{k=n-K}^{n} \mathbb{P}\{ X_{n} \neq j, X_{n-1} \neq j, \cdots, X_{k+1} \neq j \mid X_{k} = j \} \mathbb{P}\{ X_{k} = j \}
\]

\[
= \sum_{k=n-K}^{n} \mathbb{P}\{ X_{n-k} \neq j, X_{n-k-1} \neq j, \cdots, X_{1} \neq j \mid X_{0} = j \} \mathbb{P}\{ X_{k} = j \}
\]

\[
= \sum_{k=n-K}^{n} \mathbb{P}\{ \tau(j) > n-k \mid X_{0} = j \} \mathbb{P}\{ X_{k} = j \}
\]

\[
= \sum_{k=0}^{K} \mathbb{P}\{ \tau(j) > k \mid X_{0} = j \} \mathbb{P}\{ X_{n-k} = j \}, \tag{1.14}
\]

where the third equality comes from the homogeneity of \( X \) and the fourth equality from the definition of the variable \( \tau(j) \). Combining [1.13] and [1.14], we conclude that for \( n \geq K \), there exists an integer \( k \in \{0, 1, \ldots, K\} \), depending on \( n \), such that:

\[
\mathbb{P}\{ X_{n-k} = j \} \leq \frac{\varepsilon}{2}.
\]

Relation [1.12] states that for all \( \varepsilon > 0 \), there exists an integer \( N \geq K \) such that for \( n \geq N \), we have:

\[
\max_{k=0,1,\ldots,K} |\mathbb{P}\{ X_{n-k} = j \} - \mathbb{P}\{ X_{n} = j \}| \leq \frac{\varepsilon}{2}.
\]

Combining these two inequalities, we obtain, for \( n \geq N \),

\[
\mathbb{P}\{ X_{n} = j \} \leq \mathbb{P}\{ X_{n-k} = j \} + |\mathbb{P}\{ X_{n-k} = j \} - \mathbb{P}\{ X_{n} = j \}| \leq \varepsilon,
\]

which completes the proof. The second result is obtained by choosing \( \alpha = \delta^i \), as we did in the proof of the second result of theorem 1.22. \( \blacksquare \)
1.10. Ergodic theorem

Let \( j \) be a fixed state of the state space \( S \) of a Markov chain \( X \). Let us recall that \( \tau(j) \) denotes the number of transitions necessary to reach state \( j \), that is:

\[
\tau(j) = \inf\{n \geq 1 \mid X_n = j\}.
\]

In the same way, we define the sequence \( \tau_\ell(j) \) of successive passage times to state \( j \), by:

\[
\tau_0(j) = 0 \quad \text{and} \quad \tau_\ell(j) = \inf\{n \geq \tau_{\ell-1}(j) + 1 \mid X_n = j\}, \quad \text{for } \ell \geq 1,
\]

with the convention \( \inf \emptyset = \infty \). We thus have \( \tau(j) = \tau_1(j) \). The length \( S_\ell(j) \) of the \( \ell \)th excursion to state \( j \) is defined, for \( \ell \geq 1 \), by:

\[
S_\ell(j) = \begin{cases} 
  \tau_\ell(j) - \tau_{\ell-1}(j) & \text{if } \tau_{\ell-1}(j) < \infty \\
  0 & \text{otherwise}.
\end{cases}
\]

**Lemma 1.6.** For \( \ell \geq 2 \), conditional on \( \tau_{\ell-1}(j) < \infty \), \( S_\ell(j) \) is independent of \( \{X_0, \ldots, X_{\tau_{\ell-1}(j)}\} \) and

\[
\mathbb{P}\{S_\ell(j) = n \mid \tau_{\ell-1}(j) < \infty\} = \mathbb{P}\{\tau(j) = n \mid X_0 = j\}.
\]

**Proof.** Let us apply the strong Markov property at stopping time \( T = \tau_{\ell-1}(j) \). We have \( X_T = j \) when \( T < \infty \). Therefore, from theorem 1.5, conditional on \( T < \infty \), the process \( \{X_{T+n}, \ n \in \mathbb{N}\} \) is a Markov chain, with initial distribution \( \delta^j \) and transition probability matrix \( P \), independent of \( \{X_0, \ldots, X_T\} \). However, we have:

\[
S_\ell(j) = \inf\{n \geq 1 \mid X_{T+n} = j\},
\]

therefore, \( S_\ell(j) \) is the first passage time to state \( j \) for the Markov chain \( \{X_{T+n}, \ n \in \mathbb{N}\} \), which means that:

\[
\mathbb{P}\{S_\ell(j) = n \mid \tau_{\ell-1}(j) < \infty\} = \mathbb{P}\{\tau(j) = n \mid X_0 = j\},
\]

which completes the proof. \( \blacksquare \)
THEOREM 1.24.— Let \( j \) be a recurrent state. Conditional on \( \tau(j) < \infty \), for all \( \ell \geq 2 \), the random variables \( S_1(j), \ldots, S_\ell(j) \) are independent and the variables \( S_2(j), \ldots, S_\ell(j) \) are identically distributed. If \( X_0 = j \) then \( S_1(j) \) has the same distribution and we have \( \mathbb{E}\{S_1(j) \mid X_0 = j\} = \mathbb{E}\{\tau(j) \mid X_0 = j\} = m_j \).

PROOF.— From lemma 1.6, we have, for all \( \ell \geq 2 \),

\[
\mathbb{P}\{S_\ell(j) = n \mid \tau_{\ell-1}(j) < \infty\} = \mathbb{P}\{\tau(j) = n \mid X_0 = j\} = f^{(n)}_{j,j}.
\]

State \( j \) being recurrent, we have \( \mathbb{P}\{\tau(j) < \infty \mid X_0 = j\} = f_{j,j} = 1 \). Since \( \tau(j) = \tau_1(j) \), it follows, by summing the previous relation over \( n \) and for \( \ell = 2 \), that:

\[
\mathbb{P}\{S_2(j) < \infty \mid \tau(j) < \infty\} = \mathbb{P}\{\tau(j) < \infty \mid X_0 = j\} = f_{j,j} = 1.
\]

Therefore, we have, conditional on \( \tau(j) < \infty \), \( S_2 < \infty \), \( \mathbb{P}\)-a.s. and since \( S_2 = \tau_2(j) - \tau_1(j) \), we obtain, conditional on \( \tau(j) < \infty \), \( \tau_2(j) < \infty \), \( \mathbb{P}\)-a.s. By induction, it follows that for all \( \ell \geq 1 \), conditional on \( \tau(j) < \infty \), we have \( \tau_\ell(j) < \infty \), \( \mathbb{P}\)-a.s., and, therefore, \( X_{\tau_\ell(j)} = j \).

For all \( \ell \geq 2 \) and for all \( k_1, \ldots, k_\ell \geq 1 \), we have, using lemma 1.6 and the strong Markov property:

\[
\mathbb{P}\{S_\ell(j) = k_\ell \mid S_{\ell-1}(j) = k_{\ell-1}, \ldots, S_1(j) = k_1\} = \mathbb{P}\{S_\ell(j) = k_\ell \mid X_{\tau_{\ell-1}(j)} = j, \tau_{\ell-1}(j) < \infty, S_{\ell-1}(j) = k_{\ell-1}, \ldots, S_1(j) = k_1\} = \mathbb{P}\{\tau(j) = k_\ell \mid X_0 = j\} = f^{(k_\ell)}_{j,j}.
\]

We, therefore, obtain by induction, since \( S_1(j) = \tau(j) \),

\[
\mathbb{P}\{S_\ell(j) = k_\ell, \ldots, S_1(j) = k_1 \mid \tau(j) < \infty\} = \mathbb{P}\{S_\ell(j) = k_\ell \mid \tau(j) < \infty\} \cdots \mathbb{P}\{S_1(j) = k_1 \mid \tau(j) < \infty\} = f^{(k_\ell)}_{j,j} f^{(k_{\ell-1})}_{j,j} \cdots f^{(k_2)}_{j,j} \mathbb{P}\{\tau(j) = k_1 \mid \tau(j) < \infty\}.
\]

Therefore, conditional on \( \tau(j) < \infty \), the variables \( S_1(j), S_2(j), \ldots, S_\ell(j) \) are independent and \( S_2(j), \ldots, S_\ell(j) \) are identically distributed. If \( X_0 = j \) then we have \( \tau(j) < \infty \) with probability 1, since \( j \) is recurrent and

\[
\mathbb{P}\{\tau(j) = k_1 \mid \tau(j) < \infty\} = \mathbb{P}\{\tau(j) = k_1\} = f^{(k_1)}_{j,j},
\]
so $S_1$ also has the same distribution and, by definition,

$$\mathbb{P}\{S_1(j) \mid X_0 = j\} = \mathbb{P}\{\tau(j) \mid X_0 = j\} = m_j,$$

which completes the proof.

**Theorem 1.25.** Strong Law of Large Numbers. Let $Y_1, Y_2, \ldots$ be a sequence of independent, non-negative random variables, identically distributed with mean $\mathbb{E}(Y_1) = \mu$. Then:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = \mu, \quad \text{P}-\text{a.s.}$$

**Proof.** When $\mu < \infty$, this result is classic and a proof can be found in [WIL 91], for instance. The case where $\mu = \infty$ can be obtained easily. Indeed, let us fix $0 < N < \infty$ and define $Y_k^{(N)} = \min\{Y_k, N\}$. The variables $Y_1^{(N)}, Y_2^{(N)}, \ldots$ are non-negative, independent and identically distributed, with $Y_k \geq Y_k^{(N)}$. Therefore, we have:

$$\frac{1}{n} \sum_{k=1}^{n} Y_k \geq \frac{1}{n} \sum_{k=1}^{n} Y_k^{(N)} \to \mathbb{E}(Y_1^{(N)}), \quad \text{P}-\text{a.s., when } n \to \infty.$$ 

Since the sequence $Y_1^{(N)}$ is increasing with $N$, we have, by the monotone convergence theorem,

$$\lim_{N \to \infty} \mathbb{E}(Y_1^{(N)}) = \mathbb{E}(Y_1) = \infty.$$

Thus we obtain:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y_k = \infty, \quad \text{P}-\text{a.s.},$$

which completes the proof.
For \( n \geq 1 \) and \( j \in S \), we denote by \( V_j(n) \) the number of visits of the Markov chain \( X \) to state \( j \) up to time \( n - 1 \) and by \( V_j \) the total number of visits of the Markov chain \( X \) to state \( j \), that is:

\[
V_j(n) = \sum_{k=0}^{n-1} 1_{\{X_k = j\}} \quad \text{and} \quad V_j = \sum_{k=0}^{\infty} 1_{\{X_k = j\}}.
\]

**THEOREM 1.26.** -- **ERGODIC THEOREM.** -- For all \( j \in S \), we have:

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k = j\}} = \frac{1_{\{\tau(j) < \infty\}}}{m_j}, \quad \text{P}-\text{a.s.},
\]

where \( m_j = \mathbb{E}\{\tau(j) \mid X_0 = j\} \) is the expected return time to state \( j \), with the convention \( 1/\infty = 0 \).

**PROOF.** -- Let \( j \in S \) and \( \alpha = (\alpha_i, \; i \in S) \) be the initial distribution of the Markov chain \( X \). We then have, by definition, \( V_j = N_j + 1_{\{X_0 = j\}} \).

If state \( j \) is transient then, from corollary 1.3, we have, for all \( i \in S \),

\[
P\{V_j < \infty \mid X_0 = i\} = P\{N_j < \infty \mid X_0 = i\} = 1,
\]

hence:

\[
P\{V_j < \infty\} = \sum_{i \in S} \alpha_i P\{V_j < \infty \mid X_0 = i\} = 1.
\]

Since \( V_j < \infty \), \( P\)-a.s., we have, when \( n \) tends to infinity,

\[
\frac{V_j(n)}{n} \leq \frac{V_j}{n} \to 0, \quad P\text{-a.s.}
\]

Moreover, if \( j \) is transient then we have \( m_j = \infty \), and \( 1_{\{\tau(j) < \infty\}}/m_j = 0 \), which completes the proof when \( j \) is transient.
If state $j$ is recurrent then two cases arise depending on whether $\tau(j) = \infty$ or $\tau(j) < \infty$.

If $\tau(j) = \infty$ then, by definition, we have $N_j = 0$, and $V_j(n) = V_j = 1\{X_0 = j\}$. It follows that, when $n$ tends to infinity,

$$\frac{V_j(n)}{n} = \frac{V_j}{n} \rightarrow 0, \quad P\text{-a.s.}$$

In fact, since $j$ is recurrent, we have $\tau(j) = \infty \Rightarrow X_0 = j$, and $V_j = 0$. On the other hand, since $\tau(j) = \infty$, we have $1\{\tau(j) < \infty\} = 0$, that is $1\{\tau(j) < \infty\}/m_j = 0$.

Let us now consider the case where $\tau(j) < \infty$.

If $X_0 = i \neq j$ then we have, for all $n \geq 1$, by definition of $S_{\ell}(j)$,

$$S_1(j) + \cdots + S_{V_j(n)}(j) \leq n - 1 < S_1(j) + \cdots + S_{V_j(n)+1}(j),$$

where the left hand side is equal to 0 when $V_j(n) = 0$. Indeed, the left hand side is the last passage time to state $j$ before time $n$ and the right hand side is the first passage time to state $j$ after time $n - 1$. We then obtain:

$$\frac{S_1(j)}{V_j(n)} + \frac{S_2(j) + \cdots + S_{V_j(n)}(j)}{V_j(n)} < \frac{n}{V_j(n)} \leq \frac{S_1(j)}{V_j(n)} + \frac{S_2(j) + \cdots + S_{V_j(n)+1}(j)}{V_j(n)}.$$  \[1.15\]

From theorem 1.24, the random variables $S_2(j), \ldots$ are independent and identically distributed, with mean $m_j$. Since $\tau(j) = \tau_1(j) < \infty$ and since state $j$ is recurrent, we have $\tau_{\ell}(j) < \infty$, for all $\ell \geq 1$, and thus $X_{\tau_1(j)} = j$ and

$$\lim_{n \rightarrow \infty} V_j(n) = V_j = \sum_{k=0}^{\infty} 1\{X_k = j\} = \sum_{\ell=0}^{\tau_1(j)-1} \sum_{k=\tau_1(j)}^{\tau_{\ell+1}(j)-1} 1\{X_k = j\} = \sum_{\ell=1}^{\infty} 1 = \infty.$$

It follows, since $S_1(j) = \tau_1(j) < \infty$, that $\lim_{n \rightarrow \infty} S_1(j)/V_j(n) = 0$. Moreover, from theorem 1.25 on the strong law of large numbers, we have:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=2}^{n} S_k(j) = m_j, \quad P\text{-a.s.}$$
By taking the limit in [1.15], we obtain:
\[
\lim_{n \to \infty} \frac{n}{V_j(n)} = m_j, \quad \mathbb{P}_j\text{-a.s.}
\]
that is:
\[
\lim_{n \to \infty} \frac{V_j(n)}{n} = \frac{1}{m_j}, \quad \mathbb{P}_j\text{-a.s.}
\]

If \( X_0 = j \), using theorem 1.24, the random variables \( S_1(j), S_2(j), \ldots \) are independent and identically distributed and we have \( \mathbb{P}\{S_1(j)\} = m_j \). As above, since \( \tau(j) = \tau_1(j) < \infty \) and since state \( j \) is recurrent, we have \( \tau_l(j) < \infty \), for all \( \ell \geq 1 \) and, therefore, \( X_{\tau_l(j)} = j \) and
\[
\lim_{n \to \infty} \frac{V_j(n)}{n} = V_j = \sum_{k=0}^{\infty} 1\{X_k=j\} = \sum_{\ell=0}^{\tau_{\ell+1}(j)-1} \sum_{k=\tau_\ell(j)}^{\tau_{\ell+1}(j)-1} 1\{X_k=j\} = \sum_{\ell=0}^{\infty} 1 = \infty.
\]

Finally, since \( X_0 = j \), we have, for \( n \geq 1 \), by definition of the variables \( S_\ell(j) \),
\[
S_1(j) + \cdots + S_{V_j(n)-1}(j) \leq n - 1 < S_1(j) + \cdots + S_{V_j(n)}(j),
\]
where the left hand side is equal to 0 when \( V_j(n) = 1 \). Indeed, the left hand side is the last passage time to state \( j \) before time \( n \) and the right hand side is the first passage time to state \( j \) after time \( n - 1 \). We then have:
\[
\frac{S_1(j) + \cdots + S_{V_j(n)-1}(j)}{V_j(n)} < \frac{n}{V_j(n)} \leq \frac{S_1(j) + \cdots + S_{V_j(n)}(j)}{V_j(n)}.
\] [1.16]

From theorem 1.25 on the strong law of large numbers, we have:
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} S_k(j) = m_j, \quad \mathbb{P}_j\text{-a.s.},
\]
therefore, taking the limit in [1.16], we obtain:
\[
\lim_{n \to \infty} \frac{n}{V_j(n)} = m_j, \quad \mathbb{P}_j\text{-a.s.}
\]
that is:
\[
\lim_{n \to \infty} \frac{V_j(n)}{n} = \frac{1}{m_j}, \quad \mathbb{P}_j\text{-a.s.}
\]
In total, we have shown, when $j$ is recurrent, that:

$$\mathbb{P}\left\{ \lim_{n \to \infty} \frac{V_j(n)}{n} = \frac{1_{\{\tau(j) < \infty\}}}{m_j} \mid \tau(j) = \infty \right\} = 1$$

and that, for all $i \in S$,

$$\mathbb{P}\left\{ \lim_{n \to \infty} \frac{V_j(n)}{n} = \frac{1_{\{\tau(j) < \infty\}}}{m_j} \mid X_0 = i, \tau(j) < \infty \right\} = 1,$$

that is, by deconditioning,

$$\mathbb{P}\left\{ \lim_{n \to \infty} \frac{V_j(n)}{n} = \frac{1_{\{\tau(j) < \infty\}}}{m_j} \right\} = 1,$$

which completes the proof.

**THEOREM 1.27.**– Let $X$ be an irreducible Markov chain. For all $j \in S$, we have:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k = j\}} = \frac{1_{\{\tau(j) < \infty\}}}{m_j}, \quad \mathbb{P}\text{-a.s.},$$

where $m_j = \mathbb{E}\{\tau(j) \mid X_0 = j\}$ is the expected return time to state $j$, with the convention $1/\infty = 0$. If, moreover, $X$ is positive recurrent then, for every bounded function $r : S \to \mathbb{R}$, we have:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} r_{X_k} = \sum_{j \in S} r_j \pi_j, \quad \mathbb{P}\text{-a.s.},$$

where $\pi = (\pi_j, j \in S)$ is the unique invariant probability of $X$.

**PROOF.**– For the first part, it is sufficient to apply theorem 1.26. Indeed, the chain $X$ being irreducible, it is either transient or recurrent. If it is transient then, for all $j \in S$, we have $m_j = \infty$, thus $1/m_j = 0$ and $1_{\{\tau(j) < \infty\}}/m_j = 0$. If it is recurrent then, from theorem 1.16, we have $\tau(j) < \infty$ with probability 1, and so $1_{\{\tau(j) < \infty\}} = 1$ with probability 1.

Moreover, let us now assume that $X$ is positive recurrent. From corollary 1.4, it has a unique invariant probability $\pi = (\pi_j, j \in S)$ that is positive and satisfies
\[ \pi_j = 1/m_j. \] Let \( r \) be a bounded function from \( S \) to \( \mathbb{R} \). Let us define \( c = \sup_{j \in S} |r_j| \). Let \( F \subset S \) be a subset of states of \( S \). We have:

\[
\left| \frac{1}{n} \sum_{k=0}^{n-1} r_{X_k} - \sum_{j \in S} r_j \pi_j \right| = \left| \frac{1}{n} \sum_{j \in S} r_j \sum_{k=0}^{n-1} 1\{X_k = j\} - \sum_{j \in S} r_j \pi_j \right|
\]

\[
= \left| \sum_{j \in S} r_j \left( \frac{V_j(n)}{n} - \pi_j \right) \right|
\]

\[
\leq \sum_{j \in S} |r_j| \left| \frac{V_j(n)}{n} - \pi_j \right|
\]

\[
\leq c \sum_{j \in S} \left| \frac{V_j(n)}{n} - \pi_j \right|
\]

\[
= c \sum_{j \in F} \left| \frac{V_j(n)}{n} - \pi_j \right| + c \sum_{j \in S \setminus F} \left| \frac{V_j(n)}{n} - \pi_j \right|
\]

\[
\leq c \sum_{j \in F} \left| \frac{V_j(n)}{n} - \pi_j \right| + c \sum_{j \in S \setminus F} \left( \frac{V_j(n)}{n} + \pi_j \right)
\]

\[
= c \sum_{j \in F} \left| \frac{V_j(n)}{n} - \pi_j \right| + c \sum_{j \in S \setminus F} \left( \frac{V_j(n)}{n} - \pi_j \right) + 2c \sum_{j \in S \setminus F} \pi_j.
\]

Noting that:

\[
\sum_{j \in S} V_j(n) = n \quad \text{and} \quad \sum_{j \in S} \pi_j = 1,
\]
we obtain:

\[
\sum_{j \in S \setminus F} \left( \frac{V_j(n)}{n} - \pi_j \right) = -\sum_{j \in F} \left( \frac{V_j(n)}{n} - \pi_j \right) \leq \left| \sum_{j \in F} \left( \frac{V_j(n)}{n} - \pi_j \right) \right|,
\]

which leads to:

\[
\left| \frac{1}{n} \sum_{k=0}^{n-1} r_{X_k} - \sum_{j \in S} r_j \pi_j \right| \leq 2c \sum_{j \in F} \left| \frac{V_j(n)}{n} - \pi_j \right| + 2c \sum_{j \in S \setminus F} \pi_j.
\]
We have seen that, for all \( j \in S \), we have:

\[
\lim_{n \to \infty} \frac{V_j(n)}{n} = \pi_j, \quad \mathbb{P}\text{-a.s.}
\]

Let \( \varepsilon > 0 \), we choose the subset of states \( F \) finite and such that:

\[
\sum_{j \in S \setminus F} \pi_j \leq \frac{\varepsilon}{4c}
\]

and we choose \( N = N(\omega) \) such that, for \( n \geq N \), we have:

\[
\left| \frac{V_j(n)}{n} - \pi_j \right| \leq \frac{\varepsilon}{4c|F|}, \quad \text{for all } j \in F,
\]

where \( |F| \) denotes the number of states of \( F \). This ensures that, for all \( n \geq N \), we have:

\[
2c \sum_{j \in F} \left| \frac{V_j(n)}{n} - \pi_j \right| + 2c \sum_{j \in S \setminus F} \pi_j \leq \varepsilon,
\]

which completes the proof.

\[ \blacksquare \]

**Corollary 1.5.**— **For all** \( j \in S \) **and for every initial distribution** \( \alpha \), **we have:**

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (\alpha P^k)_j = \frac{\mathbb{P} \{ \tau(j) < \infty \}}{m_j}.
\]

**In particular, for all** \( i, j \in S \), **we have:**

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,j} = \frac{f_{i,j}}{m_j}.
\]

**Proof.**— From theorem 1.26, we have, with probability 1,

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k = j\}} = \frac{1_{\{\tau(j) < \infty\}}}{m_j}.
\]
and since:

\[ 0 \leq \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k=j\}} \leq 1, \]

we obtain, by the dominated convergence theorem,

\[ \lim_{n \to \infty} \mathbb{E} \left\{ \frac{1}{n} \sum_{k=0}^{n-1} 1_{\{X_k=j\}} \right\} = \mathbb{E} \left\{ \frac{1_{\{\tau(j)\leq\infty\}}}{m_j} \right\}, \]

that is:

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbb{P}\{X_k=j\} = \frac{\mathbb{P}\{\tau(j)\leq\infty\}}{m_j}. \]

The second result is obtained simply by choosing \( \alpha = \delta_i \). Indeed, in this case, we have \( \mathbb{P}\{\tau(j) < \infty\} = \sum_{k \in S} \alpha_k f_{k,j} = f_{i,j}. \)

**Corollary 1.6.**– Let \( X \) be an irreducible Markov chain. For all \( j \in S \) and for every initial distribution \( \alpha \), we have:

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (\alpha P^k)_j = \frac{1}{m_j}. \]

In particular, for all \( i, j \in S \), we have:

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,j} = \frac{1}{m_j}. \]

**Proof.**– If \( X \) is transient then we have, for all \( j \in S, m_j = \infty \). If \( X \) is recurrent then, from theorem 1.16, we have:

\[ \mathbb{P}\{\tau(j) < \infty\} = 1. \]

We then conclude, in both cases, using corollary 1.5.
**THEOREM 1.28**.– For all $i, j \in S$, we have:

\[ i \rightarrow j \text{ and } i \text{ positive recurrent} \implies j \text{ positive recurrent}. \]

**PROOF.**– If $i = j$ the result is trivial. Let $i$ and $j$ be two states such that $i \neq j$ and let us assume that $i \rightarrow j$ and that $i$ is positive recurrent. From theorem 1.13, we know that $j$ is recurrent and that $f_{i,j} = f_{j,i} = 1$. Therefore, from lemma 1.1, we have $j \rightarrow i$. Since $i \neq j$, there exist two integers $\ell \geq 1$ and $m \geq 1$ such that:

\[ (P^\ell)_{i,j} > 0 \quad \text{and} \quad (P^m)_{j,i} > 0. \]

It then follows that, for all $k \geq 0$, we have:

\[ (P^{m+k+\ell})_{j,j} \geq (P^m)_{j,i}(P^\ell)_{i,j}. \]

Summing over $k$ from 0 to $n - 1$ with $n \geq 1$ and dividing by $n$, we obtain:

\[ \frac{1}{n} \sum_{k=0}^{n-1} (P^{m+k+\ell})_{j,j} \geq (P^m)_{j,i}(P^\ell)_{i,j} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,i} \]

and after a change of variable in the left hand side:

\[ \frac{1}{n} \sum_{k=0}^{m+\ell+n-1} (P^k)_{j,j} - \frac{1}{n} \sum_{k=0}^{m+\ell-1} (P^k)_{j,j} \geq (P^m)_{j,i}(P^\ell)_{i,j} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,i}. \]

From corollary 1.5 and since $i$ and $j$ are recurrent, that is $f_{i,i} = f_{j,j} = 1$, we have, by taking the limit when $n$ tends to infinity in this last inequality and since $m_i < \infty$, \[ \frac{1}{m_j} \geq \frac{(P^m)_{j,i}(P^\ell)_{i,j}}{m_i} > 0. \]

It follows that $m_j < \infty$, that is $j$ is positive recurrent.

**COROLLARY 1.7**.– For all $i, j \in S$, we have:

\[ i \rightarrow j \text{ and } i \text{ null recurrent} \implies j \text{ null recurrent}. \]
52 Markov Chains – Theory, Algorithms and Applications

PROOF.– If \( i = j \) the result is trivial. Let \( i \) and \( j \) be two states such that \( i \neq j \) and let us assume that \( i \rightarrow j \) and that \( i \) is null recurrent. From theorem 1.13, we know that \( j \) is recurrent and \( f_{i,j} = f_{j,i} = 1 \). It then follows, from Lemma 1.1, that \( j \rightarrow i \). If \( j \) is positive recurrent then, from theorem 1.28, \( i \) is also positive recurrent, which contradicts the hypothesis, therefore, \( j \) is null recurrent.

Theorem 1.28 and corollary 1.7 show that, like transience, recurrence and periodicity, positive recurrence and null recurrence are class properties, which means that if a state of an equivalence class is positive recurrent (respectively null recurrent) then every state of this class is positive recurrent (respectively null recurrent). Therefore, an irreducible Markov chain is either transient, null recurrent or positive recurrent.

DEFINITION 1.18.– A non-empty subset \( C \) of states of \( S \) is said to be positive recurrent (respectively null recurrent) if all its states are positive recurrent (respectively null recurrent).

COROLLARY 1.8.– If an equivalence class \( C \) is finite and closed then it is positive recurrent.

PROOF.– Let \( C \) be a finite and closed equivalence class. By definition it is irreducible, therefore, it is either transient, null recurrent or positive recurrent. Since \( C \) is finite, if it is transient then it is non-closed from theorem 1.15. It is, therefore, recurrent.

Let us assume that \( C \) is null recurrent. In this case, we have, for all \( j \in C \), \( m_j = \infty \) and from theorem 1.13, that \( f_{i,j} = 1 \), for all \( i, j \in C \). Corollary 1.5 then states that

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,j} = 0.
\]

The class \( C \) being closed, we can use lemma 1.2. Since, in addition, \( C \) is finite, we obtain, by summing this relation over \( j \in C \),

\[
0 = \sum_{j \in C} \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} (P^k)_{i,j} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} \sum_{j \in C} (P^k)_{i,j} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} 1 = 1,
\]

which is a contradiction, therefore, \( C \) is positive recurrent. ■
1.11. First passage times and number of visits

1.11.1. First passage time to a state

Let us recall that the random variable $\tau(j)$, which counts the number of transitions necessary to reach state $j$, is defined by:

$$\tau(j) = \inf\{n \geq 1 \mid X_n = j\},$$

where $\tau(j) = \infty$ if this set is empty. The time $\tau(j)$ is also called the first passage time to state $j$.

For all $i, j \in S$ and, for every integer $n \geq 1$, we have defined:

$$f_{i,j}^{(n)} = \mathbb{P}\{\tau(j) = n \mid X_0 = i\} = \mathbb{P}\{X_n = j, X_k \neq j, 1 \leq k \leq n - 1 \mid X_0 = i\}.$$

Thus $f_{i,i}^{(n)}$ denotes the probability, starting from $i$, that the first return to state $i$ occurs at time $n$ and, for $i \neq j$, $f_{i,j}^{(n)}$ denotes the probability, starting from $i$, that the first visit to state $j$ occurs at time $n$. We have also introduced the quantities $f_{i,j}$ defined by:

$$f_{i,j} = \mathbb{P}\{\tau(j) < \infty \mid X_0 = i\} = \sum_{n=1}^{\infty} f_{i,j}^{(n)}.$$

$f_{i,i}$ is, therefore, the probability, starting from $i$, that the first return to state $i$ occurs in a finite time and, for $i \neq j$, $f_{i,j}$ is the probability, starting from $i$, that the first visit to state $j$ occurs in a finite time.

It has been shown in theorem 1.7 that, for all $i, j \in S$ and, for all $n \geq 1$, we have:

$$f_{i,j}^{(n)} = \begin{cases} P_{i,j} & \text{if } n = 1 \\ \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} f_{\ell,j}^{(n-1)} & \text{if } n \geq 2 \end{cases} \tag{1.17}$$

and

$$f_{i,j} = P_{i,j} + \sum_{\ell \in S \setminus \{j\}} P_{i,\ell} f_{\ell,j}. \tag{1.18}$$
Let us fix a state $j_0 \in S$ and define, for all $i \in S$ and $n \geq 0$,

$$v_i(n) = \mathbb{P}\{\tau(j_0) > n \mid X_0 = i\}.$$ 

For all $i \in S$, the sequence $(v_i(n))_{n \geq 0}$ is decreasing and lower bounded, therefore, it is convergent. If we denote by $v_i$ the limit of $v_i(n)$ when $n$ tends to infinity then we have:

$$v_i = \lim_{n \to \infty} v_i(n) = \mathbb{P}\{\tau(j_0) = \infty \mid X_0 = i\} = 1 - f_{i,j_0}.$$ 

Note that, from theorem 1.13, if $i \to j_0$ and if $i$ is recurrent then we have $f_{i,j_0} = 1$, that is $v_i = 0$.

We denote by $S_0$ the state space $S$ without state $j_0$, in other words:

$$S_0 = S \setminus \{j_0\}.$$ 

We also denote by $v(n)$ the column vector defined by:

$$v(n) = (v_i(n), i \in S_0)$$

and by $M$ the matrix obtained from $P$ by removing the row and the column corresponding to state $j_0$, that is:

$$M = (P_{i,j})_{i,j \in S_0}.$$ 

**Theorem 1.29.**– For all $n \geq 0$, we have:

$$v(n) = M^n \mathbb{1}. \quad [1.19]$$ 

**Proof.**– Clearly, since $\tau(j_0) > 0$, we have, for all $i \in S$, $v_i(0) = 1$, that is, in particular,

$$v(0) = \mathbb{1}.$$
Using recurrence relation \([1.17]\) on the \(f^{(n)}_{i,j}\), we obtain, for all \(n \geq 1\) and \(i \in S\), using Fubini’s theorem,

\[
v_i(n) = 1 - \sum_{k=1}^{n} f_{i,j_0}^{(k)}
\]

\[
= 1 - P_{i,j_0} - \sum_{k=2}^{n} \sum_{\ell \in S_0} P_{i,\ell} f^{(k-1)}_{\ell,j_0}
\]

\[
= 1 - P_{i,j_0} - \sum_{\ell \in S_0} P_{i,\ell} \sum_{k=2}^{n} f^{(k-1)}_{\ell,j_0}
\]

\[
= 1 - P_{i,j_0} - \sum_{\ell \in S_0} P_{i,\ell} \sum_{k=1}^{n-1} f^{(k)}_{\ell,j_0}
\]

\[
= 1 - P_{i,j_0} - \sum_{\ell \in S_0} P_{i,\ell} (1 - v_\ell(n - 1))
\]

\[
= \sum_{\ell \in S_0} P_{i,\ell} v_\ell(n - 1),
\]

which gives in matrix notation, by taking \(i \in S_0\), \(v(n) = Mv(n - 1)\), that is:

\[
v(n) = M^n v(0) = M^n \mathbb{1},
\]

which completes the proof.

We denote by \(v\) the column vector of the limits \(v_i\) for \(i \in S_0\), that is:

\[
v = (v_i, \ i \in S_0).
\]

If \(x\) and \(y\) are two vectors of the same size then we say that \(x \leq y\) if, for all \(i\), we have \(x_i \leq y_i\).

**Theorem 1.30.**– The vector \(v\) is a solution to equation:

\[
z = Mz
\]

and we have \(v \leq \mathbb{1}\). Moreover, \(v\) is the largest solution upper bounded by \(\mathbb{1}\), that is if \(z\) is another solution such that \(z \leq \mathbb{1}\) then \(z \leq v\).
PROOF. We have just seen that \( v_i = 1 - f_{i,j_0} \), therefore, we have \( v \leq 1 \). From equation [1.19], we have \( v(n) = Mv(n-1) \), that is for all \( i \in S_0 \),

\[
v_i(n) = \sum_{j \in S_0} M_{i,j}v_j(n-1).
\]

Taking the limit, by the dominated convergence theorem, we obtain:

\[
v_i = \sum_{j \in S_0} M_{i,j}v_j,
\]

that is \( v = Mv \). Let us note that this result can also be obtained directly from relation [1.18] by using the fact that \( v_i = 1 - f_{i,j_0} \). To show that \( v \) is the largest solution to [1.21] upper bounded by 1, let us assume that \( z \) is another solution to [1.21] such that \( z \leq 1 \). We proceed by induction. For \( n = 0 \), we have:

\[
v(0) = 1 \geq z.
\]

Let us assume, for an integer \( n \geq 0 \), that we have \( v(n) \geq z \). Using equation [1.19], we have, since the coefficients of the matrix \( M \) are non-negative,

\[
v(n+1) = Mv(n) \geq Mz = z.
\]

Thus we have shown, by induction, that for all \( n \geq 0 \), we have:

\[
v(n) \geq z.
\]

Taking the limit when \( n \) tends to infinity, we obtain \( v \geq z \), which completes the proof.

In order to calculate \( f_{j_0,j_0} = 1 - v_{j_0} \), we use relation [1.18] and we obtain:

\[
f_{j_0,j_0} = 1 - \sum_{\ell \in S_0} P_{j_0,\ell} v_{\ell}.
\] [1.22]

In particular, this calculation allows us to determine whether a Markov chain is recurrent or transient. We can also calculate the expected value of \( \tau(j_0) \) in the following way. For all \( i \in S \), we define \( w_i \) by:

\[
w_i = \mathbb{E}(\tau(j_0) \mid X_0 = i).
\]
We have, in particular, \( w_i = m_{i,j} \) and \( w_{j0} = m_{j0,j0} = m_{j0} \), where the expected hitting times \( m_{i,j} \) are given by relation [1.9]. We denote by \( w \) the column vector containing the coefficients \( w_i \) except \( w_{j0} \), that is \( w = (w_i, \ i \in S_0) \). The first part of the following theorem is identical to theorem 1.18, but here we write it in matrix notation and we propose a different proof.

**Theorem 1.31.**—Let \( X \) be a Markov chain. We have:

1) \( w = \mathbb{1} + Mw \).

2) \( w = \sum_{n=0}^{\infty} M^n \mathbb{1} \).

3) \( w \) is the smallest non-negative solution to equation \( z = \mathbb{1} + Mz \).

**Proof.**—For all \( i \in S \), we have:

\[
    w_i = \mathbb{P}\{\tau(j_0) \mid X_0 = i\} = \sum_{n=0}^{\infty} \mathbb{P}\{\tau(j_0) > n \mid X_0 = i\} = \sum_{n=0}^{\infty} v_i(n).
\]

Using relation [1.20], we obtain using Fubini’s theorem, for all \( i \in S \),

\[
    w_i = 1 + \sum_{n=1}^{\infty} v_i(n) = 1 + \sum_{n=1}^{\infty} \sum_{\ell \in S_0} P_{i,\ell} v_{\ell}(n-1) = 1 + \sum_{\ell \in S_0} P_{i,\ell} \sum_{n=1}^{\infty} v_{\ell}(n-1) = 1 + \sum_{\ell \in S_0} P_{i,\ell} w_{\ell},
\]

which is exactly relation [1.9]. Taking \( i \in S_0 \), this relation can be written in matrix notation, as:

\[
    w = \mathbb{1} + Mw.
\]

From theorem 1.29, we have:

\[
    w = \sum_{n=0}^{\infty} v(n) = \sum_{n=0}^{\infty} M^n \mathbb{1}.
\]
Let \( z \) be a solution to \( z = I + Mz \) such that \( z \geq 0 \). By induction, we have, for all \( n \geq 0 \),

\[
z = \sum_{k=0}^{n} M^k I + M^{n+1} z.\]

Since \( z \geq 0 \), we obtain:

\[
z \geq \sum_{k=0}^{n} M^k I,
\]

which gives, by taking the limit when \( n \) tends to infinity,

\[
z \geq w,
\]

which completes this proof.

To calculate \( m_{j_0} = w_{j_0} \), we use relation [1.23] that gives:

\[
m_{j_0} = 1 + \sum_{\ell \in S_0} P_{j_0,\ell} w_{\ell}. \tag{1.24}
\]

### 1.11.2. First passage time to a subset of states

The results obtained in the previous section are easily generalized to a subset of states. Let us consider a partition of the state space \( S \) in two subsets \( B \) and \( B^c \). By definition of a partition, \( B \) and \( B^c \) are non-empty and we have \( B \cap B^c = \emptyset \) and \( B \cup B^c = S \). We then define the random variable \( \tau(B^c) \), which counts the number of transitions necessary to reach set \( B^c \), by:

\[
\tau(B^c) = \inf\{n \geq 1 \mid X_n \in B^c\},
\]

where \( \tau(B^c) = \infty \) if this set is empty. The time \( \tau(B^c) \) is also called the first passage time to the subset \( B^c \).

For all \( i \in S \) and for every integer \( n \geq 1 \), we define:

\[
f_{i,B^c}^{(n)} = \mathbb{P}\{\tau(B^c) = n \mid X_0 = i\}.\]
Thus, if \( i \in B^c \), \( f_{i,B^c}^{(n)} \) is the probability, starting from \( i \), that the first return to the subset \( B^c \), without necessarily exiting \( B^c \), occurs at time \( n \) and, if \( i \in B \), \( f_{i,B^c}^{(n)} \) is the probability, starting from \( i \), that the first visit to the subset \( B^c \) occurs at time \( n \). We also define the quantities \( f_{i,B^c} \) by:

\[
f_{i,B^c} = \mathbb{P}\{\tau(B^c) < \infty \mid X_0 = i\} = \sum_{n=1}^{\infty} f_{i,B^c}^{(n)}.
\]

If \( i \in B^c \), \( f_{i,B^c} \) is the probability, starting from \( i \), that the first return to subset \( B^c \), without necessarily exiting \( B^c \), occurs in a finite time and, if \( i \in B \), \( f_{i,B^c} \) is the probability, starting from \( i \), that the first visit to the subset \( B^c \) occurs in a finite time.

By definition of \( f_{i,B^c}^{(n)} \), we have, for \( n = 1 \),

\[
f_{i,B^c}^{(1)} = \mathbb{P}\{X_1 \in B^c \mid X_0 = i\} = \sum_{j \in B^c} P_{i,j}.
\]

and, for \( n \geq 2 \),

\[
f_{i,B^c}^{(n)} = \mathbb{P}\{X_n \in B^c, X_k \in B, 1 \leq k \leq n - 1 \mid X_0 = i\}.
\]

Following the same steps as those used in the proof of theorem 1.7, it is easy to check that for all \( i \in S \) and for all \( n \geq 1 \), we have:

\[
f_{i,B^c}^{(n)} = \begin{cases} 
\sum_{j \in B^c} P_{i,j} & \text{if } n = 1 \\
\sum_{\ell \in B} P_{i,\ell} f_{\ell,B^c}^{(n-1)} & \text{if } n \geq 2
\end{cases} \tag{1.25}
\]

and

\[
f_{i,B^c} = \sum_{j \in B^c} P_{i,j} + \sum_{\ell \in B} P_{i,\ell} f_{\ell,B^c} \tag{1.26}
\]

For all \( i \in S \) and for all \( n \geq 0 \), we define:

\[
v_i(n) = \mathbb{P}\{\tau(B^c) > n \mid X_0 = i\}.
\]
For all $i \in S$, the sequence $(v_i(n))_{n \geq 0}$ is decreasing and lower bounded, therefore, it converges. If we denote by $v_i$ the limit of $v_i(n)$ when $n$ approaches infinity, we have, by definition,

$$v_i = \lim_{n \to \infty} v_i(n) = \mathbb{P}\{\tau(B^c) = \infty \mid X_0 = i\} = 1 - f_{i,B^c}.$$

**Theorem 1.32.** For all $j \in B^c$, we have $\tau(B^c) \leq \tau(j)$ and, for all $i \in S$, we have $f_{i,j} \leq f_{i,B^c}$.

**Proof.** For all $j \in B^c$ and $n \geq 0$, we have $X_n = j \Rightarrow X_n \in B^c$, therefore,

$$\{n \geq 1 \mid X_n = j\} \subseteq \{n \geq 1 \mid X_n \in B^c\}.$$

By taking the infimum of each set, we obtain $\tau(B^c) \leq \tau(j)$. We thus have $\tau(j) < \infty \Rightarrow \tau(B^c) < \infty$, which means that for all $i \in S$, we have $f_{i,j} \leq f_{i,B^c}$. □

We decompose matrix $P$ following the partition $B, B^c$ by writing:

$$P = \left( \begin{array}{cc} P_B & P_{BB^c} \\ P_{B^cB} & P_{B^c} \end{array} \right).$$

The matrix $P_B$ (respectively $P_{B^c}$) contains the transition probabilities between states of $B$ (respectively $B^c$) and the matrix $P_{BB^c}$ (respectively $P_{B^cB}$) contains the transition probabilities from states of $B$ (respectively $B^c$) to states of $B^c$ (respectively $B$). We denote by $v_B(n)$ and $v_{B^c}(n)$ the column vectors defined by:

$$v_B(n) = (v_i(n), \ i \in B) \quad \text{and} \quad v_{B^c}(n) = (v_i(n), \ i \in B^c).$$

Let us recall that vector $\mathbb{1}$ is the column vector whose entries are all equal to 1 and whose dimension is determined by the context of its use.

**Theorem 1.33.** For all $n \geq 0$, we have:

$$v_B(n) = (P_B)^n \mathbb{1}, \quad [1.27]$$

$v_{B^c}(0) = \mathbb{1}$ and, for all $n \geq 1$,

$$v_{B^c}(n) = P_{B^cB}(P_B)^{n-1} \mathbb{1}.$$

**Proof.** Clearly, since $\tau(B^c) > 0$, we have, for all $i \in S$, $v_i(0) = 1$, that is:

$$v_B(0) = \mathbb{1} \text{ and } v_{B^c}(0) = \mathbb{1}.$$
Using recurrence relation [1.25] on the \( f_{i,B^c}^{(n)} \), we obtain, for all \( n \geq 1 \) and \( i \in S \), using Fubini’s theorem,

\[
v_i(n) = 1 - \sum_{k=1}^{n} f_{i,B^c}^{(k)}
= 1 - \sum_{j \in B^c} P_{i,j} - \sum_{k=2}^{n} \sum_{\ell \in B} P_{i,\ell} f_{\ell,B^c}^{(k-1)}
= 1 - \sum_{j \in B^c} P_{i,j} - \sum_{\ell \in B} P_{i,\ell} \sum_{k=2}^{n} f_{\ell,B^c}^{(k-1)}
= 1 - \sum_{j \in B^c} P_{i,j} - \sum_{\ell \in B} P_{i,\ell} \sum_{k=1}^{n-1} f_{\ell,B^c}^{(k)}
= 1 - \sum_{j \in B^c} P_{i,j} - \sum_{\ell \in B} P_{i,\ell} \left( 1 - v_\ell(n-1) \right)
= \sum_{\ell \in B} P_{i,\ell} v_\ell(n-1).
\]

[1.28]

For \( i \in B \), we obtain, in matrix notation \( v_B(n) = P_B v_B(n-1) \), that is, for all \( n \geq 0 \),

\[
v_B(n) = (P_B)^n v_B(0) = (P_B)^n 1.
\]

For \( i \in B^c \), we obtain, in matrix notation \( v_{B^c}(n) = P_{B^c B} v_B(n-1) \), that is, for all \( n \geq 1 \),

\[
v_{B^c}(n) = P_{B^c B} (P_B)^{n-1} 1,
\]

which completes this proof.

We denote by \( v_B \) the column vector of the limits \( v_i \) for \( i \in B \), that is:

\[
v_B = (v_i, i \in B).
\]

**Theorem 1.34.** The vector \( v_B \) is a solution to the equation:

\[
z = P_B z
\]

[1.29]
and we have $v_B \leq \mathbb{1}$. Moreover, $v_B$ is the largest solution upper bounded by $\mathbb{1}$, that is if $z$ is another solution such that $z \leq \mathbb{1}$ then $z \leq v_B$.

**Proof.**— By definition, we have $v_B \leq \mathbb{1}$. From equation [1.26], we have, for all $i \in B$, since $v_i = 1 - f_{i,B^c}$,

$$v_i = 1 - f_{i,B^c} = 1 - \sum_{j \in B^c} P_{i,j} - \sum_{\ell \in B} P_{i,\ell}(1 - v_\ell),$$

that is:

$$v_i = \sum_{\ell \in B} P_{i,\ell}v_\ell,$$

which can be written as $v_B = P_B v_B$.

To show that $v_B$ is the largest solution to [1.29] upper bounded by $\mathbb{1}$, let us assume that $z$ is another solution to [1.29] such that $z \leq \mathbb{1}$. We proceed by induction. For $n = 0$, we have:

$$v_B(0) = \mathbb{1} \geq z.$$

Let us assume that, for an integer $n \geq 0$, we have $v_B(n) \geq z$. Using relation [1.27], we have, since the coefficients of $P_B$ are non-negative,

$$v_B(n + 1) = P_B v_B(n) \geq P_B z = z.$$

Thus we have shown, by induction, that for all $n \geq 0$, we have:

$$v_B(n) \geq z.$$

Taking the limit when $n$ tends to infinity, we obtain $v_B \geq z$, which completes this proof.

To calculate the coefficients $v_i$, for $i \in B^c$, we use relation [1.26] and we obtain, as we have already seen, since $v_i = 1 - f_{i,B^c}$,

$$v_i = \sum_{\ell \in B} P_{i,\ell}v_\ell.$$
If we denote by $v_{B^c}$ the column vector defined by $v_{B^c} = (v_i, i \in B^c)$, we obtain, in matrix notation,

$$v_{B^c} = P_{B^c} B v_B.$$ 

The calculation of the expected value of $\tau(B^c)$ can be done in the following way.

For all $i \in S$, we define $w_i$ by:

$$w_i = \mathbb{E}\{\tau(B^c) \mid X_0 = i\}.$$ 

We denote by $w_B$ and $w_{B^c}$ the column vectors defined by:

$$w_B = (w_i, i \in B) \quad \text{and} \quad w_{B^c} = (w_i, i \in B^c).$$

**Theorem 1.35.** We have:

$$w_B = 1 + P_B w_B \quad \text{and} \quad w_{B^c} = 1 + P_{B^c} B w_B.$$ 

Moreover, $w_B$ is the smallest non-negative solution to the equation $z = 1 + P_B z$.

**Proof.** For all $i \in S$, we have:

$$w_i = \mathbb{E}\{\tau(B^c) \mid X_0 = i\} = \sum_{n=0}^{\infty} \mathbb{P}\{\tau(B^c) > n \mid X_0 = i\} = \sum_{n=0}^{\infty} v_i(n).$$ 

Using relation [1.28], we obtain using Fubini’s theorem, for all $i \in S$,

$$w_i = 1 + \sum_{n=1}^{\infty} v_i(n)$$

$$= 1 + \sum_{n=1}^{\infty} \sum_{\ell \in B} P_{i,\ell} v_\ell(n - 1)$$

$$= 1 + \sum_{\ell \in B} P_{i,\ell} \sum_{n=1}^{\infty} v_\ell(n - 1)$$

$$= 1 + \sum_{\ell \in B} P_{i,\ell} w_\ell,$$
which gives, in matrix notation, by taking $i \in B$,

$$w_B = \mathbb{1} + P_Bw_B$$

and, by taking $i \in B^c$,

$$w_{B^c} = \mathbb{1} + P_{B^c}w_B.$$

From theorem 1.33, we have:

$$w_B = \sum_{n=0}^{\infty} v_B(n) = \sum_{n=0}^{\infty} (P_B)^n \mathbb{1}. \quad [1.30]$$

Let $z$ be a solution to $z = \mathbb{1} + P_Bz$ such that $z \geq 0$. By induction, we have, for all $n \geq 0$,

$$z = \sum_{k=0}^{n} (P_B)^k \mathbb{1} + (P_B)^{n+1}z.$$

Since $z \geq 0$, we obtain:

$$z \geq \sum_{k=0}^{n} (P_B)^k \mathbb{1},$$

which gives, by taking the limit when $n$ tends to infinity,

$$z \geq w_B,$$

which completes this proof.

1.11.3. Expected number of visits

Let $B$ be a non-empty subset of transient states of a Markov chain $X$ with countable state space $S$ and with transition probability matrix $P$. We denote by $P_B$ the submatrix obtained from $P$ by removing the rows and the columns corresponding to the states which do not belong to $B$. $P_B$ is thus the matrix of the transition probabilities between states of $B$. Note that we can take $B = S$ if the chain is transient and in this case we have $P_B = P$. 
Let \( G_B \) be the matrix defined by:

\[
G_B = \sum_{n=0}^{\infty} (P_B)^n.
\]

When \( B \) is the set of all transient states of \( S \), the matrix \( G_B \) is referred to as the fundamental matrix of chain \( X \). For all \( i, j \in B \) and, for all \( n \geq 0 \), we have:

\[
((P_B)^n)_{i,j} \leq (P^n)_{i,j}.
\]

Indeed, for \( n = 0 \) and \( n = 1 \), it is easy to see that this inequality is an equality. By assuming that this inequality holds for integer \( n - 1 \), and, in other words, by assuming that for all \( i, j \in B \), we have \(((P_B)^{n-1})_{i,j} \leq (P^{n-1})_{i,j}\), we obtain:

\[
((P_B)^n)_{i,j} = \sum_{k \in B} ((P_B)^{n-1})_{i,k}(P_B)_{k,j}
\leq \sum_{k \in B} (P^{n-1})_{i,k}P_{k,j}
\leq \sum_{k \in S} (P^{n-1})_{i,k}P_{k,j}
= (P^n)_{i,j}.
\]

The states of \( B \) being transient, it follows, from corollary 1.1, that for all \( i, j \in B \), we have:

\[
(G_B)_{i,j} = \sum_{n=0}^{\infty} (P_B)^n)_{i,j} \leq \sum_{n=0}^{\infty} (P^n)_{i,j} < \infty,
\]

that is \( G_B < \infty \).

**Theorem 1.36.**— We have:

\[
P_B G_B = G_B P_B < \infty \quad \text{and} \quad (I - P_B)G_B = G_B(I - P_B) = I.
\]

Moreover, if \( B \) is finite then the matrix \( I - P_B \) is invertible and we have:

\[
(I - P_B)^{-1} = G_B.
\]
PROOF.— For all \( n \geq 0 \), we define the matrix \( G_B(n) \) by:

\[
G_B(n) = \sum_{k=0}^{n} (P_B)^k.
\]

Using Fubini’s theorem, we have, for all \( i, j \in B \),

\[
\lim_{n \to \infty} (P_B G_B(n))_{i,j} = \lim_{n \to \infty} \sum_{\ell \in B} (P_B)_{i,\ell}(G_B(n))_{\ell,j} = \sum_{\ell \in B} (P_B)_{i,\ell}(G_B)_{\ell,j} = (P_B G_B)_{i,j},
\]

hence:

\[
\lim_{n \to \infty} P_B G_B(n) = P_B G_B.
\]

In the same way, we show that:

\[
\lim_{n \to \infty} G_B(n) P_B = G_B P_B.
\]

But, we also have, by definition of \( G_B \),

\[
P_B G_B(n) = G_B(n) P_B = G_B(n + 1) - I,
\]

hence, taking the limit when \( n \) tends to infinity, we get

\[
P_B G_B = G_B P_B = G_B - I.
\]

This shows that \( P_B G_B = G_B P_B < \infty \) and that:

\[
(I - P_B) G_B = G_B (I - P_B) = I.
\]

If the subset of states \( B \) is finite then we have:

\[
(I - P_B)^{-1} = G_B,
\]

which completes this proof.
THEOREM 1.37.– $G_B$ is the smallest non-negative solution to the matrix equation $H(I - P_B) = I$ and the smallest non-negative solution to the matrix equation $(I - P_B)H = I$.

PROOF.– We have already seen that $G_B$ satisfies both equations. Let $H \geq 0$ be a solution to the matrix equation $H(I - P_B) = I$. This equation can also be written as $H = I + HP_B$ and by iterating this equality we obtain:

$$H = I + P_B + \cdots + (P_B)^n + H(P_B)^{n+1}.$$  

Since $H \geq 0$, we have $H(P_B)^{n+1} \geq 0$, and

$$H \geq I + P_B + \cdots + (P_B)^n.$$  

Taking the limit when $n$ tends to infinity, we obtain:

$$H \geq \sum_{n=0}^{\infty} (P_B)^n = G_B.$$  

The same reasoning can be applied to the matrix equation $(I - P_B)H = I$. ■

For all $j \in S$, let us recall that $V_j$ denotes the total number of visits to state $j$, that is:

$$V_j = \sum_{n=0}^{\infty} 1_{\{X_n = j\}}.$$  

We then have, from theorem 1.12, for all $i, j \in B$,

$$\mathbb{E}\{V_j \mid X_0 = i\} = \sum_{n=0}^{\infty} (P^n)_{i,j}.$$  

If $B$ denotes the set of all transient states of $X$ then the matrix $P$ has the structure [1.3] and we then have, for all $i, j \in B$,

$$\mathbb{E}\{V_j \mid X_0 = i\} = \sum_{n=0}^{\infty} (P^n)_{i,j} = \sum_{n=0}^{\infty} ((P_B)^n)_{i,j} = (G_B)_{i,j}.$$
In this case, if we denote by $V_B$ the total number of visits to states of $B$, we have:

$$V_B = \sum_{j \in B} V_j.$$ 

Using Fubini’s theorem, we obtain, for all $i \in B$,

$$\mathbb{E}\{V_B \mid X_0 = i\} = \sum_{j \in B} \mathbb{E}\{V_j \mid X_0 = i\} = \sum_{j \in B} (G_B)_{i,j}.$$ 

The first passage time $\tau(B^c) = \inf\{n \geq 1 \mid X_n \in B^c\}$ to the subset $B^c$ has been studied in the previous section. It is easy to see that if $X_0 \in B$ then we have $\tau(B^c) = V_B$. In particular, the column vector $w_B = (w_i, \ i \in B)$ defined in the previous section by $w_i = \mathbb{E}\{\tau(B^c) \mid X_0 = i\}$ is given by $w_B = G_B \mathbb{1}$, which is consistent with relation [1.30].

**Remark 1.1.–** If the subset $B$ is finite then the equation $(I - P_B)H = H(I - P_B) = I$ has a unique solution which is $H = G_B = (I - P_B)^{-1}$. Indeed, if $H_1$ and $H_2$ are two solutions to this equation, we have:

$$H_1 = H_1I = H_1((I - P_B)H_2) = (H_1(I - P_B))H_2 = IH_2 = H_2.$$ 

In the case where $B$ is infinite, the third equality above does not always hold since the product of infinite matrices is generally not associative. The discrete and absorbing birth-and-death processes described in Chapter 3, section 3.2.2, give an example of solving this equation as well as the calculation of the matrix $G_B$ and the vector $w_B$ in the case where $B$ is infinite.

### 1.12. Finite Markov chains

In this section, we consider the particular case, widely used in practice, where the state space $S$ is finite. In this case, we prove that a Markov chain cannot be transient and an irreducible Markov chain is necessarily recurrent. Moreover, recurrent states are all positive recurrent or, in other words, there are no null recurrent states.

**Theorem 1.38.–** Let $X$ be a Markov chain on a finite state space $S$.

- $X$ cannot be transient.
- If $X$ is irreducible then $X$ is recurrent.
PROOF.– If $X$ is transient then from corollary 1.1, for all $i, j \in S$, we have:

$$\sum_{n=1}^{\infty} (P^n)_{i,j} < \infty$$

and, since $S$ is finite,

$$\sum_{j \in S} \sum_{n=1}^{\infty} (P^n)_{i,j} < \infty.$$

However, using Fubini’s theorem, we have:

$$\sum_{j \in S} \sum_{n=1}^{\infty} (P^n)_{i,j} = \sum_{n=1}^{\infty} \sum_{j \in S} (P^n)_{i,j} = \sum_{n=1}^{\infty} 1 = \infty,$$

which is a contradiction.

As for the second point, recurrence and transience being class properties, if $X$ is irreducible then it is necessarily transient or recurrent. However, we have shown in the first point that it cannot be transient, therefore, it is recurrent.

THEOREM 1.39.– Let $X$ be a Markov chain on a finite state space $S$. If $X$ is irreducible then $X$ has a unique invariant probability $\pi = (\pi_j, j \in S)$, positive and given by $\pi_j = 1/m_j$.

PROOF.– The chain $X$ being irreducible on a finite state space, we deduce from theorem 1.38 that $X$ is recurrent. Theorem 1.17 then states that $X$ has, up to a multiplicative constant, a unique, positive invariant measure. In the proof of this theorem, it is shown that this measure is denoted, for a fixed state $i$, by $\gamma_i = (\gamma_{ij}, j \in S)$, where $0 < \gamma_{ij} < \infty$, and is given by relation [1.4]. This relation gives us $\gamma_i^i = 1$ and, by definition,

$$m_i = \sum_{j \in S} \gamma_{ji},$$

which shows that $m_i < \infty$, since $S$ is finite. Since $\gamma_i$ is unique, up to a multiplicative constant, taking:

$$\pi_j = \frac{\gamma_{ij}}{m_i},$$
the measure \( \pi = (\pi_j, j \in S) \) is the unique invariant probability on \( S \) and it is positive. The probability \( \pi_j \) being independent of \( i \), taking \( i = j \), we have:

\[
\pi_j = \frac{\gamma_j}{m_j} = \frac{1}{m_j},
\]

which completes this proof.

**COROLLARY 1.9.** Let \( X \) be a Markov chain on a finite state space \( S \). If \( X \) is irreducible then \( X \) is positive recurrent.

**PROOF.** The chain \( X \) being irreducible on a finite state space, from theorem 1.39, it has a unique positive invariant probability and, from corollary 1.4, it is positive recurrent.

We could also have used corollary 1.8 to prove this result since \( S \) has only one equivalence class that is finite and closed.

**COROLLARY 1.10.** A Markov chain on a finite state space does not have null recurrent states.

**PROOF.** Let \( j \) be a recurrent state of a Markov chain on a finite state space. From theorem 1.14, its equivalence class \( C(j) \) is closed and since \( C(j) \) is finite, it is positive recurrent, from corollary 1.8, therefore, the state \( j \) is positive recurrent. Thus we have shown that every recurrent state is positive recurrent, or, in other words, there are no null recurrent states.

**1.13. Absorbing Markov chains**

We now consider a Markov chain \( X \) whose countable state space \( S \) contains a finite, non-empty set \( B \) of transient states and \( J \) recurrent classes \( C_1, \ldots, C_J \) if \( J < \infty \) and \( C_1, \ldots \) if \( J = \infty \). The transition probability matrix \( P \) of \( X \) can then be written as:

\[
P = \begin{pmatrix}
P_B & P_{B,C_1} & P_{B,C_2} & \cdots & P_{B,C_{J-1}} & P_{B,C_J} \\
0 & P_{C_1} & 0 & \cdots & 0 & 0 \\
0 & 0 & P_{C_2} & 0 & \cdots & 0 \\
0 & 0 & 0 & \ddots & \ddots & \ddots \\
0 & 0 & 0 & \cdots & P_{C_{J-1}} & 0 \\
0 & 0 & 0 & \cdots & 0 & P_{C_J}
\end{pmatrix},
\]

[1.31]
where $P_{C_j}$ (respectively $P_B$) contains the transition probabilities between states of $C_j$ (respectively $B$) and $P_{B,C_j}$ contains the transition probabilities from states of $B$ to states of $C_j$. If $J = \infty$ then the structure of matrix $P$ is identical to that of [1.3]. In the rest of this section, the notation $j = 1, \ldots, J$ means $j \geq 1$ if $J = \infty$.

We assume, without loss of generality that, for all $j = 1, \ldots, J$, we have $P_{B,C_j} \neq 0$. Indeed, the condition $P_{B,C_j} = 0$ means that the class $C_j$ is not accessible from set $B$. It can, therefore, be studied independently from other classes. Such a Markov chain is called an absorbing Markov chain.

The matrix $P$ being block triangular, it is easy to see, by induction, that we have, for all $n \geq 0$,

$$P^n = \begin{pmatrix} (P_B)^n & R_1(n) & R_2(n) & \cdots & R_{J-1}(n) & R_J(n) \\ 0 & (P_{C_1})^n & 0 & \cdots & 0 & 0 \\ 0 & 0 & (P_{C_2})^n & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & (P_{C_{J-1}})^n & 0 \\ 0 & 0 & 0 & \cdots & 0 & (P_{C_J})^n \end{pmatrix},$$

where the matrix $R_j(n)$, of size $(|B|, |C_j|)$, is given for all $j = 1, \ldots, J$ by $R_j(0) = 0$ and, for all $n \geq 1$, by:

$$R_j(n) = \sum_{\ell=0}^{n-1} (P_B)^\ell P_{B,C_j} (P_{C_j})^{n-1-\ell}.$$

We assume that $X_0 \in B$. The initial probability distribution $\alpha$ of $X$ can then be written as:

$$\alpha = (\alpha_B, 0, \ldots, 0),$$

where $\alpha_B$ is the row vector of dimension $|B|$ containing the initial probabilities corresponding to the states of $B$. The hypothesis $X_0 \in B$ ensures that $\alpha_B 1 = 1$. Introducing $B^c = C_1 \cup \cdots \cup C_J$ and since $X_0 \in B$, we define the total amount of time $\tau(B^c)$ spent in the transient states, by:

$$\tau(B^c) = \inf\{n \geq 1 \mid X_n \in B^c\}.$$
This random variable has already been studied in section 1.11.2, where we have shown, see relation [1.27] of theorem 1.33, that, for all \( i \in B \), we have:

\[
\mathbb{P}\{\tau(B^c) \leq n \mid X_0 = i\} = 1 - ((P_B)^n \mathbb{1})_i.
\]

The states of \( B \) being transient, we have, from corollary 1.1:

\[
\lim_{n \to \infty} (P_B)^n = 0. \tag{1.33}
\]

\( B \) being finite, it follows, by taking the limit when \( n \) tends to infinity, that:

\[
\mathbb{P}\{\tau(B^c) < \infty \mid X_0 = i\} = 1
\]

and, from theorem 1.36, that:

\[
\mathbb{E}\{\tau(B^c) \mid X_0 = i\} = \sum_{n=0}^{\infty} \mathbb{P}\{\tau(B^c) > n \mid X_0 = i\}
\]

\[
= \sum_{n=0}^{\infty} ((P_B)^n \mathbb{1})_i
\]

\[
= ((I - P_B)^{-1} \mathbb{1})_i.
\]

We then obtain, from [1.33] and since \( X_0 \in B \),

\[
\mathbb{P}\{\tau(B^c) \leq n\} = 1 - \alpha_B (P_B)^n, \quad \alpha_B = \mathbb{P}\{\tau(B^c) < \infty\} = 1
\]

and

\[
\mathbb{E}\{\tau(B^c)\} = \alpha_B (I - P_B)^{-1} \mathbb{1} < \infty.
\]

Thus we have handled the case where \( J = 1 \) since, in this case, we have \( B^c = C_1 \). We now assume that \( J \geq 2 \). The hitting time of class \( C_j \), defined by:

\[
\tau(C_j) = \inf\{n \geq 1 \mid X_n \in C_j\}
\]
satisfies, for all \( n \geq 0 \) and since \( X_0 \in B \),
\[
\tau(C_j) \leq n \iff X_n \in C_j.
\]

Therefore, we have, from theorem 1.2 and form [1.32] of the matrix \( P^n \), for all \( n \geq 0 \),
\[
\mathbb{P}\{\tau(C_j) \leq n\} = \mathbb{P}\{X_n \in C_j\}
= \sum_{i \in C_j} (\alpha P^n)_i
= \alpha_B R_j(n) \mathbb{1}
= \alpha_B \sum_{\ell=0}^{n-1} (P_B)^{\ell} P_{B,C_j}(P_{C_j})^{n-1-\ell} \mathbb{1},
\]
where the above sum is equal to 0 if \( n = 0 \). The matrices \( P_{C_j} \) being stochastic, we have \( P_{C_j} \mathbb{1} = \mathbb{1} \), which gives, for all \( n \geq 0 \), \((P_{C_j})^n \mathbb{1} = \mathbb{1}\) and, from theorem 1.36,
\[
\mathbb{P}\{\tau(C_j) \leq n\} = \alpha_B \sum_{\ell=0}^{n-1} (P_B)^{\ell} P_{B,C_j} \mathbb{1}
= \alpha_B (I - (P_B)^n)(I - P_B)^{-1} P_{B,C_j} \mathbb{1}.
\]

Taking the limit when \( n \) tends to infinity, we obtain, using relation [1.33] and since \( B \) is finite,
\[
\mathbb{P}\{\tau(C_j) < \infty\} = \alpha_B (I - P_B)^{-1} P_{B,C_j} \mathbb{1}. \tag{1.34}
\]

It also follows that, for all \( n \geq 1 \),
\[
\mathbb{P}\{\tau(C_j) = n\} = \alpha_B (P_B)^{n-1} P_{B,C_j} \mathbb{1}.
\]

Let \( C_j \) be a fixed class. There exists at least one state \( i \in B \) such that \( \mathbb{P}\{\tau(C_j) < \infty \mid X_0 = i\} < 1 \).

Indeed, since:
\[
\mathbb{P}\{\tau(C_j) < \infty \mid X_0 = i\} = ((I - P_B)^{-1} P_{B,C_j} \mathbb{1})_i,
\]
if, for all \(i \in B\), we have \(\mathbb{P}\{\tau(C_j) < \infty \mid X_0 = i\} = 1\) then we obtain \((I - P_B)^{-1}P_{B,C_j}1 = 1\). It follows that \(P_{B,C_j}1 = (I - P_B)1\), which means, following form \([1.31]\) of the matrix \(P\), that, for every class \(C_\ell, \ell \neq j\), we have \(P_{B,C_\ell} = 0\), which is contrary to the hypothesis, since we have \(J \geq 2\). For this state \(i\), which satisfies \(\mathbb{P}\{\tau(C_j) < \infty \mid X_0 = i\} < 1\), we have \(\mathbb{E}\{\tau(C_j) \mid X_0 = i\} = \infty\).

Let us now calculate the probability that absorption occurs through class \(C_j\). This probability, which we denote by \(p_j\), is defined by:

\[
p_j = \mathbb{P}\{X_{\tau(B^c)} \in C_j\}.
\]

We then have \(\mathbb{P}\{\tau(B^c) = 0, X_{\tau(B^c)} \in C_j\} = 0\) and, for all \(n \geq 1\),

\[
\begin{align*}
\mathbb{P}\{\tau(B^c) = n, X_{\tau(B^c)} \in C_j\} &= \mathbb{P}\{\tau(B^c) = n, X_n \in C_j\} \\
&= \mathbb{P}\{X_{n-1} \in B, X_n \in C_j\} \\
&= \sum_{\ell \in B} \mathbb{P}\{X_{n-1} = \ell, X_n \in C_j\} \\
&= \sum_{\ell \in B} \mathbb{P}\{X_{n-1} = \ell\}\mathbb{P}\{X_n \in C_j \mid X_{n-1} = \ell\} \\
&= \sum_{\ell \in B} (\alpha P_n^{\ell-1}) \sum_{h \in C_j} P_{\ell,h} \\
&= \alpha_B(P_B)^{n-1}P_{B,C_j}1 \\
&= \mathbb{P}\{\tau(C_j) = n\},
\end{align*}
\]

where the fifth equality is due to theorem 1.2 and the penultimate equality comes from the decomposition of matrix \(P^n\) given in \([1.32]\). Let us note, furthermore, that, since \(X_0 \in B\), we have, for all \(n \geq 0\),

\[
\tau(C_j) = n \iff \tau(B^c) = n \text{ and } X_{\tau(B^c)} \in C_j.
\]

Summing the last equality over \(n\), we obtain, from \([1.34]\) and since \(\mathbb{P}\{\tau(B^c) < \infty\} = 1\),

\[
p_j = \mathbb{P}\{\tau(C_j) < \infty\} = \alpha_B(I - P_B)^{-1}P_{B,C_j}1. \tag{1.36}
\]

It follows that:

\[
\sum_{j=1}^{J} p_j = \alpha_B(I - P_B)^{-1}\sum_{j=1}^{J} P_{B,C_j}1 = \alpha_B(I - P_B)^{-1}(I - P_B)1 = \alpha_B1 = 1.
\]
The expected absorption time in class $C_j$ is given, again from theorem 1.36, since $B$ is finite and since $P\{\tau(B^c) < \infty\} = 1$, by:

$$E\{\tau(B^c)1_{X_{\tau(B^c)} \in C_j}\} = \sum_{n=0}^{\infty} P\{\tau(B^c) > n, X_{\tau(B^c)} \in C_j\} = \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} \alpha_B(P_B)^{k-1}P_{B,C_j} = \sum_{n=0}^{\infty} \alpha_B(P_B)^n(I - P_B)^{-1}P_{B,C_j} = \alpha_B(I - P_B)^{-2}P_{B,C_j} < \infty.$$

Thanks to equivalence [1.35], this expected time can also be written as $E\{\tau(C_j)1_{\tau(C_j) < \infty}\}$. Note that, in this way, we return to the calculation of $E\{\tau(B^c)\}$, that is:

$$E\{\tau(B^c)\} = \sum_{j=1}^{J} E\{\tau(B^c)1_{X_{\tau(B^c)} \in C_j}\} = \alpha_B(I - P_B)^{-2}J \sum_{j=1}^{J} P_{B,C_j} = \alpha_B(I - P_B)^{-2}(I - P_B) \sum_{j=1}^{J} P_{B,C_j} = \alpha_B(I - P_B)^{-1}.$$  

Finally, the expected absorption time, given that absorption occurs in class $C_j$ can be written as $E\{\tau(B^c) \mid X_{\tau(B^c)} \in C_j\} = E\{\tau(C_j) \mid \tau(C_j) < \infty\}$ and is given, when $p_j \neq 0$, by:

$$E\{\tau(B^c) \mid X_{\tau(B^c)} \in C_j\} = \frac{E\{\tau(B^c)1_{X_{\tau(B^c)} \in C_j}\}}{p_j} = \frac{\alpha_B(I - P_B)^{-2}P_{B,C_j}}{\alpha_B(I - P_B)^{-1}P_{B,C_j}} = \frac{\alpha_B(I - P_B)^{-2}P_{B,C_j}}{\alpha_B(I - P_B)^{-1}P_{B,C_j}}.$$

Note also that we can have $p_j = 0$ in some particular cases, for example when the initial distribution is such that the class $C_j$ is only accessible from states of $B$ which are never visited by the chain.
1.14. Examples

1.14.1. Two-state chain

Let us consider the Markov chain $X$ with two states, $S = \{0, 1\}$, with transition probability matrix $P$ given by:

$$P = \begin{pmatrix} 1 - p & p \\ q & 1 - q \end{pmatrix}.$$

We first assume that $p, q \in (0, 1)$, in other words none of the four transition probabilities is null. The state diagram of $X$ is shown in Figure 1.1.

![State diagram of the Markov chain in the example of section 1.14.1](image)

Since $p, q \in (0, 1)$, this chain is clearly irreducible and aperiodic. The invariant probability $\pi = (\pi_0, \pi_1)$ is then given by $\pi = \pi P$ with $\pi_0 + \pi_1 = 1$, which leads to:

$$\pi_0 = \frac{q}{p + q} \quad \text{and} \quad \pi_1 = \frac{p}{p + q}.$$

This example, being quite simple, allows us to explicitly calculate $P^n$ and we have, for all $n \geq 0$,

$$P^n = \frac{1}{p + q} \begin{pmatrix} q & p \\ q & p \end{pmatrix} + \frac{(1 - (p + q))^n}{p + q} \begin{pmatrix} p & -p \\ -q & q \end{pmatrix}$$

and since $|p + q - 1| < 1$, we get, as expected:

$$\lim_{n \to \infty} P^n = \frac{1}{p + q} \begin{pmatrix} q & p \\ q & p \end{pmatrix}.$$

This result remains valid in the particular case where $p = 1$ or $q = 1$, with $p \neq q$. 
However, if \( p = q = 1 \), we have \( p + q - 1 = 1 \) and the sequence \( P^n \) does not converge. Indeed, in this case, \( P \) is the symmetric matrix:

\[
P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\]

and the Markov chain \( X \) is periodic, with period 2. We have \( P^2 = I \), where \( I \) is the identity matrix. Therefore, for all \( n \geq 0 \), we have \( P^{2n} = I \) and \( P^{2n+1} = P \). Nevertheless, the chain still has an invariant probability given by:

\[
\pi = (1/2, 1/2).
\]

Finally, if we denote by \( \alpha = (\alpha_0, \alpha_1) \) the initial distribution of \( X \), we have, still in the case where \( p = q = 1 \) and since \( \alpha_0 + \alpha_1 = 1 \),

\[
\mathbb{P}\{X_n = 0\} = (\alpha P^n)_0 = \frac{1}{2} (1 + (\alpha_0 - \alpha_1)(-1)^n),
\]

which gives, if \( \alpha_0 = \alpha_1 = 1/2 \),

\[
\mathbb{P}\{X_n = 0\} = \frac{1}{2}.
\]

Thus, in this case, the sequence \( (\alpha P^n)_{n \geq 0} \) converges to \((1/2, 1/2)\) while the sequence \( (P^n)_{n \geq 0} \) diverges.

The hitting time of state \( j_0 = 0 \) is easy to calculate. Indeed, if we define \( w_1 = \mathbb{E}\{\tau(0) \mid X_0 = 1\} \) then we have, from theorem 1.31,

\[
w_1 = 1 + (1 - q)w_1,
\]

since vector \( w \) has a single entry \( w_1 \) and since matrix \( M \) is reduced to \( 1 - q \). It follows that:

\[
w_1 = \frac{1}{q}.
\]

From relation [1.24], we have:

\[
w_0 = \mathbb{E}\{\tau(0) \mid X_0 = 0\} = 1 + pw_1,
\]
that is:

\[ w_0 = \frac{p + q}{q}. \]

Note that we could also have written \( w_0 = m_0 = 1/\pi_0 \).

### 1.14.2. Gambler’s ruin

Consider a gambler with an initial fortune of \( i \) euros. On each successive gamble, he/she has a probability \( p \in (0, 1) \) to increase his/her fortune by 1 euro and a probability \( q = 1 - p \) to decrease it by 1 euro. The game ends when the gambler either reaches a total fortune of \( N \) euros or is ruined. We assume that \( N \geq 2 \) and its initial fortune \( i \) is such that \( 1 \leq i \leq N - 1 \). We denote by \( X_n \) the fortune of the gambler after the \( n \)th gamble. We have \( X_0 = i \) and, for all \( n \geq 0 \), if \( X_n = 0 \) or if \( X_n = N \) then \( X_{n+1} = X_n \) and if \( 0 < X_n < N \) then:

\[
X_{n+1} = \begin{cases} 
X_n - 1 & \text{with probability } q \\
X_n + 1 & \text{with probability } p.
\end{cases}
\]

This shows that the process \( X = \{X_n, \ n \in \mathbb{N}\} \) is a discrete-time Markov chain on the state space \( S = \{0, 1, \ldots, N - 1, N\} \). The state diagram of \( X \) is shown in Figure 1.2.

![State diagram of the Markov chain in the example of section 1.14.2](image)

The transition probability matrix \( P \) of this chain is of form [1.31] with \( J = 2 \) recurrence classes \( C_1 = \{0\} \) and \( C_2 = \{N\} \), states 0 and \( N \) being absorbing here. We then have \( P_{C_1} = P_{C_2} = 1 \), the transient states being the states of \( B = \{1, \ldots, N - 1\} \). From relation [1.36], the probabilities \( p_1 \) and \( p_2 \) to be absorbed, respectively, by classes \( C_1 = \{0\} \) and \( C_2 = \{N\} \), satisfy:

\[
p_1 = \mathbb{P}\{\tau(0) < \infty\} \quad \text{and} \quad p_2 = \mathbb{P}\{\tau(N) < \infty\},
\]
with

\[ p_1 + p_2 = 1. \]

The probability \( p_1 \) of being absorbed in state 0 is given by relation [1.34],

\[ p_1 = \alpha_B (I - P_B)^{-1} P_{B,C_1}. \]

The column vector \( P_{B,C_1} \) is given by \( P_{B,C_1} = (q, 0, \ldots, 0) \). We define the column vector \( h = (h_i, i = 1, \ldots, N - 1) \) by \( h = (I - P_B)^{-1} P_{B,C_1} \). The entry \( h_i \) is the probability to be absorbed in state 0 when the initial state is state \( i \), for \( i = 1, \ldots, N - 1 \). We then have, from relation [1.36] and for \( i = 1, \ldots, N - 1 \),

\[ h_i = f_{i,0} = \mathbb{P} \{ \tau(0) < \infty \mid X_0 = i \} = 1 - f_{i,N}. \]

To calculate vector \( h \), we multiply on the left by matrix \( I - P_B \). We then obtain

\[ (I - P_B)h = P_{B,C_1}, \]

which can be written as \( h = P_Bh + P_{B,C_1} \) or

\[
\begin{aligned}
    h_1 &= q + ph_2 \\
    h_\ell &= qh_{\ell-1} + ph_{\ell+1}, \text{ for } \ell = 2, \ldots, N - 2 \\
    h_{N-1} &= qh_{N-2}.
\end{aligned}
\]

Defining \( h_0 = f_{0,0} = \mathbb{P} \{ \tau(0) < \infty \mid X_0 = 0 \} \) and \( h_N = f_{N,0} = \mathbb{P} \{ \tau(0) < \infty \mid X_0 = N \} \), we have \( h_0 = 1 \) and \( h_N = 0 \). The previous induction can then be written, for all \( \ell = 1, \ldots, N - 1 \), as:

\[ h_\ell = qh_{\ell-1} + ph_{\ell+1} \]

or, since \( p + q = 1 \),

\[ p(h_{\ell+1} - h_\ell) = q(h_\ell - h_{\ell-1}). \]

Defining, for \( \ell = 0, \ldots, N - 1 \), \( x_\ell = h_{\ell+1} - h_\ell \), we obtain:

\[ x_\ell = \left( \frac{q}{p} \right) \ell x_0 = (h_1 - 1) \left( \frac{q}{p} \right) \ell. \]
Summing these values $x_{\ell}$ for $\ell$ from 1 to $N - 1$, we obtain:

$$-h_1 = x_1 + \ldots + x_{N-1} = (h_1 - 1) \sum_{\ell=1}^{N-1} \left( \frac{q}{p} \right)^{\ell}.$$ 

Defining $s_0 = 0$ and, for all $i = 1, \ldots, N - 1$,

$$s_i = \sum_{\ell=1}^{i} \left( \frac{q}{p} \right)^{\ell} = \begin{cases} 
\frac{q}{p} \left( 1 - \left( \frac{q}{p} \right)^{i} \right) & \text{if } p \neq q \text{ (that is } p \neq 1/2) \\
i & \text{if } p = q \text{ (that is } p = 1/2),
\end{cases}$$

we obtain:

$$h_1 = \frac{s_{N-1}}{s_{N-1} + 1}.$$ 

Summing the values of $x_{\ell}$ for $\ell$ from 1 to $i - 1$, we have:

$$h_i - h_1 = x_1 + \ldots + x_{i-1} = (h_1 - 1) s_{i-1},$$

which gives, by replacing $h_1$ by its value, for all $i = 1, \ldots, N - 1$,

$$h_i = \frac{s_{N-1} - s_{i-1}}{s_{N-1} + 1}.$$ 

Replacing the sums $s_{i-1}$ by their values, we obtain, for $i = 1, \ldots, N - 1$,

$$h_i = \begin{cases} 
\frac{\left( \frac{q}{p} \right)^{i} - \left( \frac{q}{p} \right)^{N}}{1 - \left( \frac{q}{p} \right)^{N}} & \text{if } p \neq q \text{ (that is } p \neq 1/2) \\
1 - \frac{i}{N} & \text{if } p = q \text{ (that is } p = 1/2),
\end{cases}$$
Defining $B^c = \{0, N\}$, time $\tau(B^c) = \inf\{n \geq 1 \mid X_n \in B^c\}$ is the time at which the chain is absorbed. We then have $p_1 = \mathbb{P}\{X_{\tau(B^c)} = 0\} = \alpha_B h$ or, for all $i = 1, \ldots, N - 1$,

$$h_i = \mathbb{P}\{X_{\tau(B^c)} = 0 \mid X_0 = i\}.$$

The expected hitting time $w_i$ of $0$, starting from state $i$, for $i = 1, \ldots, N - 1$, is given by $w_i = \mathbb{E}\{\tau(0) \mid X_0 = i\}$ and since $h_i = f_{i,0} = \mathbb{P}\{\tau(0) < \infty \mid X_0 = i\} < 1$, we have $w_i = \infty$, for all $i = 1, \ldots, N - 1$. The same result holds for the expected hitting time of state $N$, starting from state $i$, for $i = 1, \ldots, N - 1$.

It is important to point out that this Markov chain has an infinity of invariant probabilities. Indeed, the system $\pi = \pi P$ with $\pi \mathbb{1} = 1$ has as solutions $\pi_0 = a$, $\pi_1 = \ldots = \pi_{N-1} = 0$, $\pi_N = 1 - a$, for all $a \in [0, 1]$.

Concerning convergence to equilibrium, since the states $1, \ldots, N - 1$ are transient, we have, from corollary 1.1, for all $i = 0, \ldots, N$ and for all $j = 1, \ldots, N - 1$,

$$\lim_{n \to \infty} (P^n)_{i,j} = 0.$$

States $0$ and $N$ being absorbing, we clearly have, for all $n \geq 0$,

$$(P^n)_{0,0} = (P^n)_{N,N} = 1.$$

It follows that $(P^n)_{0,N} = (P^n)_{N,0} = 0$. For all $i = 1, \ldots, N - 1$, using relation [1.2] of theorem 1.6, we have:

$$(P^n)_{i,0} = \sum_{k=1}^{n} f_{i,0}^{(k)} (P^{n-k})_{0,0} = \sum_{k=1}^{n} f_{i,0}^{(k)}.$$

It follows that:

$$\lim_{n \to \infty} (P^n)_{i,0} = f_{i,0} = h_i,$$

and thus:

$$\lim_{n \to \infty} (P^n)_{i,N} = f_{i,N} = 1 - h_i.$$
We have seen that the expected total time spent in transient states is given by:

\[ \mathbb{E}\{\tau(B^c)\} = \alpha_B (I - P_B)^{-1} \mathbb{1}. \]

This calculation can be done in a similar way to that of vector \( h \). Note that when \( p = q = 1/2 \), we obtain:

\[ \mathbb{E}\{\tau(B^c) \mid X_0 = i\} = i(N - i). \]

**1.14.3. Success runs**

Consider the game of heads and tails with independent trials. If the coin turns up tail, the player wins 1 euro and if it turns up head, he/she loses all his/her winnings. When the player has in his/her possession \( i \) euros, he/she uses a coin with a probability of tail equal to \( p_i \in (0, 1) \) and a probability of head equal to \( q_i = 1 - p_i \). We denote by \( X_n \) the player’s fortune after the \( n \)th toss. We then have, for all \( n \geq 0 \),

\[ X_{n+1} = \begin{cases} 
0 & \text{with probability } q_{X_n} \\
X_n + 1 & \text{with probability } p_{X_n}.
\end{cases} \]

This shows that the process \( X = \{X_n, \ n \in \mathbb{N}\} \) is a discrete-time Markov chain on the state space \( S = \mathbb{N} \), with transition probability matrix \( P \) whose non-null coefficients are given, for all \( i \geq 0 \), by:

\[ P_{i,i+1} = p_i \quad \text{and} \quad P_{i,0} = q_i. \]

The state diagram of \( X \) is shown in Figure 1.3.

![State diagram of the Markov chain in example of section 1.14.3](image)

**Figure 1.3. State diagram of the Markov chain in example of section 1.14.3**

The Markov chain is clearly irreducible and aperiodic since the probabilities \( p_i \) and \( q_i \) are positive for all \( i \geq 0 \).
We will need the following result to deal with this example.

**THEOREM 1.40.** If \( (u_n)_{n \geq 0} \) is a sequence of real numbers such that, for all \( n \geq 0 \), \( u_n \in (0, 1) \) then we have:

1) \( \sum_{n=0}^{\infty} u_n < \infty \iff -\sum_{n=0}^{\infty} \ln(1 - u_n) < \infty. \)

2) \( \prod_{n=0}^{\infty} (1 - u_n) = 0 \iff \sum_{n=0}^{\infty} u_n = \infty. \)

**PROOF.** If the series \( \sum_{n} u_n \) converges then we have \( \lim_{n \to \infty} u_n = 0 \), hence \( -\ln(1 - u_n) \sim u_n \), which means that the series \( -\sum_{n} \ln(1 - u_n) \) converges. Conversely, if the series \( -\sum_{n} \ln(1 - u_n) \) converges then we have \( \lim_{n \to \infty} \ln(1 - u_n) = 0 \), hence \( \lim_{n \to \infty} u_n = 0 \) and \( -\ln(1 - u_n) \sim u_n \), which means that the series \( \sum_{n} u_n \) converges. To prove equivalence 2, we define, for all \( N \geq 0 \),

\[
P_N = \prod_{n=0}^{N} (1 - u_n) \quad \text{and} \quad S_N = -\sum_{n=0}^{N} \ln(1 - u_n).
\]

\( P_N \) and \( S_N \) are well-defined and we have \( S_N = -\ln(P_N) \) and \( P_N = e^{-S_N}. \)

If \( \lim_{N \to \infty} P_N = 0 \) then \( \lim_{N \to \infty} S_N = \infty \), which implies, by equivalence 1, that \( \sum_{n=0}^{\infty} u_n = \infty \). Conversely, if \( \sum_{n=0}^{\infty} u_n = \infty \) then, by equivalence 1, we have \( -\sum_{n=0}^{\infty} \ln(1 - u_n) = \infty \), or, in other words, \( \lim_{N \to \infty} S_N = \infty \), and thus \( \lim_{N \to \infty} P_N = 0. \)

The first passage time \( \tau(0) \) to state 0 has been defined by:

\( \tau(0) = \inf \{ n \geq 1 \mid X_n = 0 \}, \)

and we define, for all \( n \geq 0 \) and \( i \geq 0 \),

\( v_i(n) = P\{ \tau(0) > n \mid X_0 = i \}. \)

We have \( v_i(0) = 1 \), for all \( i \geq 0 \) and, from the particular structure of the state diagram of \( X \), Figure 1.3, we have, for all \( i \geq 0 \) and \( n \geq 1 \),

\[
v_i(n) = P\{ \tau(0) > n \mid X_0 = i \} = P\{ X_1 = i + 1, X_2 = i + 2, \ldots, X_n = i + n \mid X_0 = i \} = p_i p_{i+1} \cdots p_{i+n-1}.
\]
Therefore, we have, for all $i \geq 0$,

$$v_i = \mathbb{P}\{\tau(0) = \infty \mid X_0 = i\} = \lim_{n \to \infty} \prod_{k=i}^{i+n-1} p_k.$$

Equivalence 2 of theorem 1.40 allows us to write, for all $i \geq 1$,

$$v_i = \prod_{k=i}^{\infty} p_k = \prod_{k=i}^{\infty} (1 - q_k) = 0 \iff \sum_{k=0}^{\infty} q_k = \infty$$

or alternatively:

$$v_i > 0 \iff \sum_{k=0}^{\infty} q_k < \infty.$$

Applying relation [1.22] with $j_0 = 0$, we obtain:

$$f_{0,0} = 1 - p_0 v_1.$$

It follows that state 0 is recurrent if and only if $v_1 = 0$ and thus state 0 is transient if and only if $v_1 > 0$. Since the chain $X$ is irreducible, it is either recurrent or transient, therefore,

$$X \text{ is recurrent } \iff \sum_{k=0}^{\infty} q_k = \infty$$

$$X \text{ is transient } \iff \sum_{k=0}^{\infty} q_k < \infty.$$

If the chain $X$ is recurrent then we have, for all $i \geq 1$,

$$w_i = \mathbb{E}\{\tau(0) \mid X_0 = i\} = \sum_{n=0}^{\infty} v_i(n) = 1 + \sum_{n=1}^{\infty} \prod_{k=i}^{i+n-1} p_k.$$

To calculate $m_0 = w_0$, we apply relation [1.24] with $j_0 = 0$, which gives:

$$m_0 = 1 + p_0 w_1 = 1 + \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k.$$
State 0 is thus positive recurrent if the latter sum is finite and null recurrent otherwise. Since $X$ is irreducible, it follows, from theorem 1.28 and corollary 1.7, that $X$ is positive recurrent if this latter sum is finite and null recurrent otherwise. We have shown, in summary, that:

1) $X$ is transient $\iff \sum_{k=0}^{\infty} q_k < \infty$.

2) $X$ is positive recurrent $\iff \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k < \infty$.

3) $X$ is null recurrent $\iff \sum_{k=0}^{\infty} q_k = \infty$ and $\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k = \infty$.

Note that, in the second case, the required condition is sufficient since, by equivalence 2 of theorem 1.40, we have:

$$\sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k < \infty \implies \prod_{k=0}^{\infty} p_k = \prod_{k=0}^{\infty} (1 - q_k) = 0 \iff \sum_{k=0}^{\infty} q_k = \infty.$$  

This comment has already been raised before, in the general case, just after definition 1.12.

Concerning convergence to equilibrium, since $P_{0,0} = q_0 > 0$ and since $X$ is irreducible, it follows that $X$ is aperiodic. Hence from corollary 1.1 and theorem 1.23, if $X$ is transient or null recurrent, we have, for every initial distribution of $X$ and for all $j \geq 0$,

$$\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = 0.$$  

If $X$ is positive recurrent then, from corollary 1.4, $X$ has a unique invariant probability $\pi = (\pi_j, j \geq 0)$ that is positive. To calculate it, we consider the system $\pi = \pi P$, which can also be written, for all $j \geq 1$, as:

$$\pi_j = p_{j-1} \pi_{j-1}.$$  

Thus we obtain, for all $j \geq 1$,

$$\pi_j = \pi_0 \prod_{k=0}^{j-1} p_k.$$
\( \pi \) being a probability, we have:

\[
1 = \sum_{n=0}^{\infty} \pi_n = \pi_0 \left( 1 + \sum_{n=1}^{\infty} \prod_{k=0}^{n-1} p_k \right) = \pi_0 \left( 1 + \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k \right).
\]

We have previously seen that this latter series converges since \( X \) is positive recurrent. Therefore, we obtain:

\[
\pi_0 = \frac{1}{1 + \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k} = \frac{1}{m_0}.
\]

More generally, corollary 1.4 states that the expected return time to state \( j \), which we have denoted by \( m_j \), is given, for all \( j \geq 0 \), by:

\[
m_j = \frac{1}{\pi_j} = \frac{1 + \sum_{n=0}^{\infty} \prod_{k=0}^{n} p_k}{\prod_{k=0}^{j-1} p_k},
\]

where the denominator is equal to 1 if \( j = 0 \). Finally, theorem 1.22 allows us to state that, for every initial distribution of \( X \) and for all \( j \geq 0 \), we have:

\[
\lim_{n \to \infty} \mathbb{P}\{X_n = j\} = \pi_j.
\]

In the case where, for all \( i \geq 0 \), we have \( p_i = p \in (0, 1) \), we obtain \( q_i = q = 1 - p \) and it is easy to see that \( X \) is positive recurrent. The stationary distribution is then given, for all \( j \geq 0 \), by:

\[
\pi_j = p^j (1 - p).
\]

This example being particularly simple, it was not necessary to use the main results of the previous section, however, this is not always the case as we will see in Chapter 3.
1.15. Bibliographical notes

There are numerous books dealing with discrete-time Markov chains on a finite or countably infinite state space. Among these volumes, some are less recent but nevertheless fundamental, such as [BHA 60], [COX 65], [CHU 67], [ÇIN 75], [FEL 57], [FEL 66], [FRE 83], or [HOE 72], [KAR 75], [KAR 81], [KEM 66], [KEM 76], [ROS 83]. Among the latest books on this subject, let us cite, in particular, [ASM 03], [BHA 90], [BOL 98], [BRÉ 98], [KIJ 97], [KUL 10], [NOR 97], [SHI 89], [TIJ 03] and [TRI 02], which present the main results and, for some, offer a number of applications related to dynamic systems modeling. Every one of these books handles in more or less detail the theoretical aspects associated with the analysis of discrete-time Markov chains. Some propose specific studies of particular interest, for example [KEM 76], in which the authors approach for the first time, to the best of our knowledge, the state aggregation problem. This problem consists of determining whether a process constructed from a Markov chain by aggregating, in some way, a number of states into a single state, still remains a Markov chain. This question has later been studied in papers like [RUB 89a] and [RUB 91].