# Lecture notes on Markov chains

Olivier Lévêque, olivier.leveque#epfl.ch
National University of Ireland, Maynooth, August 2-5, 2011

# 1 Discrete-time Markov chains

## 1.1 Basic definitions and Chapman-Kolmogorov equation

**(Very) short reminder on conditional probability.** Let $A$, $B$, $C$ be events.

* $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$ (well defined only if $\mathbb{P}(B) > 0$)

* $\mathbb{P}(A \cap B|C) = \dfrac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(C)} = \dfrac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} \cdot \dfrac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)} = \mathbb{P}(A|B \cap C) \cdot \mathbb{P}(B|C)$

Let now $X$ be a discrete random variable.

* $\sum_k \mathbb{P}(X = x_k|B) = 1$

* $\mathbb{P}(B) = \sum_k \mathbb{P}(B|X = x_k)\, \mathbb{P}(X = x_k)$

* $\sum_k \mathbb{P}(X = x_k, A|B) = \mathbb{P}(A|B)$

but watch out that

* $\sum_k \mathbb{P}(B|X = x_k) \neq 1$

**Definition 1.1.** A *Markov chain* is a discrete-time stochastic process $(X_n,\ n \geq 0)$ such that each random variable $X_n$ takes values in a discrete set $S$ ($S = \mathbb{N}$, typically) and

$$\mathbb{P}(X_{n+1} = j \,\big|\, X_n = i, X_{n-1} = i_{n-1}, ..., X_0 = i_0) = \mathbb{P}(X_{n+1} = j \big| X_n = i) \qquad \forall n \geq 0, j, i, i_{n-1}, ..., i_0 \in S$$

That is, as time goes by, the process loses the memory of the past.

If moreover $\mathbb{P}(X_{n+1} = j|X_n = i) = p_{ij}$ is independent of $n$, then $X$ is said to be a *time-homogeneous* Markov chain. We will focus on such chains during the course.

**Terminology.**
* The possible values taken by the random variables $X_n$ are called the *states* of the chain. $S$ is called the *state space*.
* The chain is said to be *finite-state* if the set $S$ is finite ($S = \{0, \dots, N\}$, typically).
* $P = (p_{ij})_{i,j \in S}$ is called the *transition matrix* of the chain.

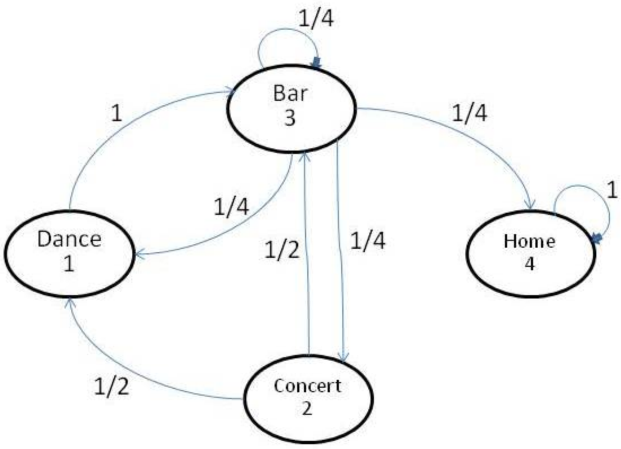
**Properties of the transition matrix.**
* $p_{ij} \geq 0$, $\forall i, j \in S$.
* $\sum_{j \in S} p_{ij} = 1$, $\forall i \in S$.

It is always possible to represent a time-homogeneous Markov chain by a transition graph.

**Example 1.2.** (music festival)
The four possible states of a student in a music festival are $S = \{$ "dancing", "at a concert", "at the bar", "back home"$\}$. Let us assume that the student changes state during the festival according to the following transition matrix:

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

This Markov chain can be represented by the following transition graph:




**Example 1.3.** (simple symmetric random walk)
Let $(X_n, n \geq 1)$ be i.i.d. random variables such that $\mathbb{P}(X_n = +1) = \mathbb{P}(X_n = -1) = \frac{1}{2}$, and let $(S_n, n \geq 0)$ be defined as $S_0 = 0, S_n = X_1 + ... + X_n, n \geq 1$. Then $(S_n,\, n \in \mathbb{N})$ a Markov chain with state space $S = \mathbb{Z}$. Indeed:

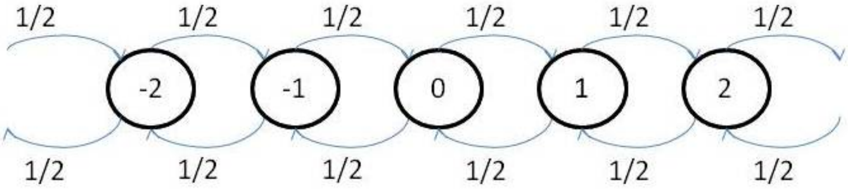
$$\begin{aligned} &\mathbb{P}(S_{n+1} = j | S_n = i, S_{n-1} = i_{n-1}, ..., S_0 = i_0) \\ &= \mathbb{P}(X_{n+1} = j - i | S_n = i, S_{n-1} = i_{n-1}, ..., S_0 = i_0) = \mathbb{P}(X_{n+1} = j - i) \end{aligned}$$

by the assumption that the variables $X_n$ are independent. The chain is moreover time-homogeneous, as

$$\mathbb{P}(X_{n+1} = j - i) = \begin{cases} \frac{1}{2} & \text{if } |j - i| = 1 \\ 0 & \text{otherwise} \end{cases}$$

does not depend on $n$.

Here is the transition graph of the chain:



The *distribution at time* $n$ of the Markov chain $X$ is given by:

$$\pi_i^{(n)} = \mathbb{P}(X_n = i), \quad i \in S$$

We know that $\pi_i^{(n)} \geq 0$ for all $i \in S$ and that $\sum_{i \in S} \pi_i^{(n)} = 1$.

The *initial distribution* of the chain is given by

$$\pi_i^{(0)} = \mathbb{P}(X_0 = i), \quad i \in S$$

It must be specified together with the transition matrix $P = (p_{ij}), i, j \in S$ in order to characterize the chain completely. Indeed, by repeatedly using the Markov property, we obtain:

$$\begin{aligned}
&\mathbb{P}(X_n = i_n, X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0) \\
&= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0) \cdot \mathbb{P}(X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0) \\
&= p_{i_{n-1}, i_n} \, \mathbb{P}(X_{n-1} = i_{n-1}, ..., X_1 = i_1, X_0 = i_0) \\
&= \ldots = p_{i_{n-1}, i_n} \, p_{i_{n-2}, i_{n-1}} \cdots p_{i_1, i_2} \, p_{i_0, i_1} \, \pi_{i_0}^{(0)}
\end{aligned}$$

so knowing $P$ and knowing $\pi^{(0)}$ allows to compute all the above probabilities, which give a compete description of the process.

The $n$*-step transition probabilities* of the chain are given by

$$p_{ij}^{(n)} = \mathbb{P}(X_{m+n} = j | X_m = i), \quad n, m \geq 0, \quad i, j \in S$$

Let us compute:

$$\begin{aligned}
p_{ij}^{(2)} &= \mathbb{P}(X_{n+2} = j | X_n = i) = \sum_{k \in S} \mathbb{P}(X_{n+2} = j, X_{n+1} = k | X_n = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+2} = j | X_{n+1} = k, X_m = i) \cdot \mathbb{P}(X_{n+1} = k | X_n = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{n+2} = j | X_{n+1} = k) \cdot \mathbb{P}(X_{n+1} = k | X_n = i) = \sum_{k \in S} p_{ik} \, p_{kj} \qquad (1)
\end{aligned}$$

where the Markov property property was used in (1). In a similar manner, we obtain the *Chapman-Kolmogorov equation* for generic values of $m$ and $n$:

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} \, p_{kj}^{(m)}, \quad i, j \in S, \quad n, m \geq 0$$

where we define by convention $p_{ij}^{(0)} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases}$

Notice that in terms of the transition matrix $P$, this equation simply reads:

$$(P^{n+m})_{ij} = (P^n \, P^m)_{ij} = \sum_{k \in S} (P^n)_{ik} \, (P^m)_{kj}$$

where, again by convention, $P^0 = I$, the identity matrix.

Notice also that

$$\pi_j^{(n)} = \mathbb{P}(X_n = j) = \sum_{i \in S} \mathbb{P}(X_n = j | X_{n-1} = i)\, \mathbb{P}(X_{n-1} = i) = \sum_{i \in S} p_{ij}\, \pi_i^{(n-1)}$$

In matrix form (considering $\pi^{(n)}$ as a row vector), this equation reads $\pi^{(n)} = \pi^{(n-1)} P$, from which we deduce that $\pi^{(n)} = \pi^{(n-2)} P^2 = \ldots = \pi^{(0)} P^n$, i.e.

$$\pi_j^{(n)} = \sum_{i \in S} p_{ij}^{(n)}\, \pi_i^{(0)}$$

## 1.2 Classification of states

We list here a set of basic definitions.

* A state $j$ is *accessible* from state $i$ if $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

* State $i$ and $j$ *communicate* if both $j$ is accessible from $i$ and $i$ is accessible from $j$. Notation: $i \longleftrightarrow j$.

"To communicate" is an equivalence relation:

reflexivity: $i$ always communicates with $i$ (by definition).
symmetry: if $i$ communicates with $j$, then $j$ communicates with $i$ (also by definition).
transitivity: if $i$ communicates with $j$ and $j$ communicates with $k$, then $i$ communicates with $k$ (proof in the exercises)

* Two states that communicate are said to belong to the same *equivalence class*, and the state space S is divided into a certain number of such classes.

In Example 1.2, the state space $S$ is divided into two classes: {"dancing", "at a concert", "at the bar"} and {"back home"}. In Example 1.3, there is only one class $S = \mathbb{Z}$.

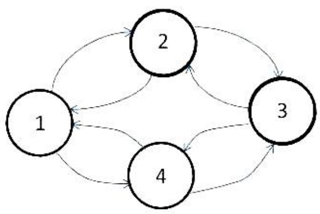
* The Markov chain is said to be *irreducible* if there is only one equivalence class (i.e. all states communicate with each other).

* A state $i$ is *absorbing* if $p_{ii} = 1$.

* A state $i$ is *periodic with period* $d$ if $d$ is the smallest integer such that $p_{ii}^{(n)} = 0$ for all $n$ which are not multiples of $d$. In case $d = 1$, the state is said to be *aperiodic*.

* It can be shown that if a state $i$ is periodic with period $d$, then all states in the same class are periodic with the same period $d$, in which case the whole class is periodic with period $d$.

**Example 1.4.** The Markov chain whose transition graph is given by




is an irreducible Markov chain, periodic with period 2.

### 1.2.1 Recurrent and transient states

Let us recall here that $p_{ii}^{(n)} = \mathbb{P}(X_n = i|X_0 = i)$ is the probability, starting from state $i$, to come back to state $i$ after $n$ steps. Let us also define $f_i = \mathbb{P}(X \text{ever returns to } i|X_0 = i)$.

**Definition 1.5.** A state $i$ is said to be *recurrent* if $f_i = 1$ or *transient* if $f_i < 1$.

It can be shown that all states in a given class are either recurrent or transient. In Example 1.2, the class {"dancing", "at a concert", "at the bar"} is transient (as there is a positive probability to leave the class and never come back) and the class {"back home"} is obviously recurrent. The random walk example 1.3 is more involved and requires the use of the following proposition.

**Proposition 1.6.**
* State $i$ is recurrent if and only if $\sum_{n\geq 1} p_{ii}^{(n)} = \infty$.
* State $i$ is transient if and only if $\sum_{n\geq 1} p_{ii}^{(n)} < \infty$.

Notice that the two lines of the above proposition are redundant, as a state is transient if and only if it is not recurrent.

*Proof.* Let $T_i$ be the first time the chain X returns to state $i$. Therefore, $f_i = \mathbb{P}(T_i < \infty|X_0 = i)$. Let also $N_i$ be the the number of times the chain X returns to state $i$ and let us compute

$$\begin{aligned}
\mathbb{P}(N_i < \infty|X_0 = i) &= \mathbb{P}(\exists n \geq 1 : X_n = i \ \& \ X_m \neq i, \ \forall m > n|X_0 = i) \\
&= \sum_{n\geq 1} \mathbb{P}(X_n = i \ \& \ X_m \neq i, \ \forall m > n|X_0 = i) \\
&= \sum_{n\geq 1} \mathbb{P}(X_m \neq i, \ \forall m > n|X_n = i, X_0 = i) \ \mathbb{P}(X_n = i|X_0 = i)
\end{aligned}$$

As $X$ is a time-homogeneous Markov chain, we have

$$\mathbb{P}(X_m \neq i, \ \forall m > n|X_n = i, X_0 = i) = \mathbb{P}(X_m \neq i, \ \forall m > n|X_n = i) = \mathbb{P}(X_m \neq i, \ \forall m > 0|X_0 = i)$$

Therefore,

$$\begin{aligned}
\mathbb{P}(N_i < \infty|X_0 = i) &= \mathbb{P}(X_m \neq i, \ \forall m > 0|X_0 = i) \sum_{n\geq 1} \mathbb{P}(X_n = i|X_0 = i) \\
&= \mathbb{P}(T_i = \infty|X_0 = i) \sum_{n\geq 1} p_{ii}^{(n)} = (1 - f_i) \sum_{n\geq 1} p_{ii}^{(n)} \qquad (2)
\end{aligned}$$

This implies that

- If $i$ is recurrent, then $1 - f_i = 0$, so by (2), $\mathbb{P}(N_i < \infty|X_0 = i) = 0$, whatever the value of $\sum_{n\geq 1} p_{ii}^{(n)}$. This in turn implies

$$\mathbb{P}(N_i = \infty|X_0 = i) = 1, \quad \text{so} \quad \mathbb{E}(N_i|X_0 = i) = \infty$$

As $N_i = \sum_{n\geq 1} 1_{\{X_n=i\}}$, we also have

$$\infty = \mathbb{E}(N_i|X_0 = i) = \sum_{n\geq 1} \mathbb{E}(1_{\{X_1=i\}}|X_0 = i) = \sum_{n\geq 1} \mathbb{P}(X_n = i|X_0 = i) = \sum_{n\geq 1} p_{ii}^{(n)}$$

which proves the claim in this case.

- If on the contrary $i$ is transient, then $1 - f_i > 0$ and as $\mathbb{P}(N_i = \infty|X_0 = i) \leq 1$ , we obtain, using (2)

$$\sum_{n\geq 1} p_{ii}^{(n)}\,(1 - f_i) \leq 1 \quad \text{i.e.} \quad \sum_{n\geq 0} p_{ii}^{(n)} \leq \frac{1}{1 - f_i} < \infty$$

which completes the proof. □

Notice that as a by-product, we showed in this proof that if a state of a Markov chain is recurrent, then it is visited infinitely often by the chain, with probability 1 (therefore the name "recurrent").

**Application.** (simple random walk, symmetric or asymmetric)
Let us consider the simple random walk $(S_n,\, n \in \mathbb{N})$, with the following transition probabilities:

$$S_0 = 0, \quad \mathbb{P}(S_{n+1} = S_n + 1) = p = 1 - \mathbb{P}(S_{n+1} = S_n - 1) \quad \text{where } 0 < p < 1$$

Starting from 0, the probability to reach 0 after $2n$ steps is given by

$$p_{00}^{(2n)} = \mathbb{P}(S_{2n} = 0|S_0 = 0) = \binom{2n}{n} p^n\,(1 - p)^n, \quad \text{where} \quad \binom{2n}{n} = \frac{(2n)!}{(n!)^2}$$

Notice that $p_{00}^{(2n+1)} = 0$ for all $n$ and $p$, as an even number of steps is required to come back to 0. Using Stirling's approximation formula $n! \simeq n^n e^{-n}\sqrt{2\pi n}$, we obtain

$$p_{00}^{(2n)} \simeq \frac{(4p(1 - p))^n}{\sqrt{\pi n}}$$

From this expression, we see that if $p = 1/2$, then

$$\sum_{n\geq 1} p_{00}^{(n)} = \sum_{n\geq 1} p_{00}^{(2n)} = \sum_{n\geq 1} \frac{1}{\sqrt{\pi n}} = \infty$$

so by Proposition 1.6, state 0 is recurrent, and as the chain is irreducible, the whole chain is recurrent. If on the contrary $p \neq 1/2$, then $4p(1 - p) < 1$, so

$$\sum_{n\geq 1} p_{00}^{(n)} = \sum_{n\geq 1} p_{00}^{(2n)} = \sum_{n\geq 1} \frac{(4p(1 - p))^n}{\sqrt{\pi n}} < \infty$$

so state 0, and therefore the whole chain, is transient (in this case, the chain "escapes" to either $+\infty$ or $-\infty$, depending on the value of $p$).

Among recurrent states, we further make the following distinction (the justification for this distinction will come later).

**Definition 1.7.** Let $i$ be a recurrent state and $T_i$ be the first return time to state $i$.
* $i$ is *positive recurrent* if $\mathbb{E}(T_i|X_0 = i) < \infty$
* $i$ is *null recurrent* if $\mathbb{E}(T_i|X_0 = i) = \infty$

That is, if state $i$ is null recurrent, then the chain comes back infinitely often to $i$, because the state is recurrent, but the time between two consecutive visits to $i$ is on average infinite!

Notice that even if $f_i = \mathbb{P}(T_i < \infty|X_0 = i) = 1$, this does *not* imply that $\mathbb{E}(T_i|X_0 = i) < \infty$, as

$$\mathbb{E}(T_i|X_0 = i) = \sum_{n\geq 1} n\,\mathbb{P}(T_i = n|X_0 = i)$$

can be arbitrarily large.

**Remarks.**
* In a given class, all states are either positive recurrent, null recurrent or transient.
* In a finite state Markov chain, all recurrent states are actually positive recurrent.
* The simple symmetric random walk turns out to be null recurrent.

## 1.3 Stationary and limiting distributions

Let us first remember that a time-homogeneous Markov chain at time $n$ is characterized by its distribution $\pi^{(n)} = (\pi_i^{(n)},\, i \in S)$, where $\pi_i^{(n)} = \mathbb{P}(X_n = i)$, and that

$$\pi^{(n+1)} = \pi^{(n)} P, \quad \text{i.e.} \quad \pi_j^{(n+1)} = \sum_{i\in S} \pi_i^{(n)}\, p_{ij}, \quad \forall j \in S$$

**Definition 1.8.** A distribution $\pi^* = (\pi_i^*,\, i \in S)$ is said to be a *stationary distribution* for the Markov chain $(X_n,\, n \geq 0)$ if

$$\pi^* = \pi^* P, \quad \text{i.e.} \quad \pi_j^* = \sum_{i\in S} \pi_i^*\, p_{ij}, \quad \forall j \in S \tag{3}$$

**Remarks.**
* $\pi^*$ does not necessarily exist, nor is it necessarily unique.
* As we will see, if $\pi^*$ exists and is unique, then $\pi_i^*$ can always be interpreted as the average proportion of time spent by the chain $X$ in state $i$. It also turns out in this case that

$$\mathbb{E}(T_i|X_0 = i) = \frac{1}{\pi_i^*}$$

where $T_i = \inf\{n \geq 0 : X_n = i\}$ is the first time the chain comes back to state $i$.
* If $\pi^{(0)} = \pi^*$, then $\pi^{(1)} = \pi^*\, P = \pi^*$; likewise, $\pi^{(n)} = \pi^*\, P^n = \ldots = \pi^*$, $\forall n \geq 0$, that is, if the initial distribution of the chain is stationary (we also say the chain is "in stationary state", by abuse of language), then it remains stationary over time.

**Trivial example.**
If $(X_n,\, n \geq 0)$ is a sequence of i.i.d. random variables, then $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_{n+1} = j)$ does actually depend neither on $i$ nor on $n$ , so $\pi_i^* = \mathbb{P}(X_n = i)$ (which is also independent of $n$) is a stationary distribution for the chain. Indeed,

$$\sum_{i \in S} \pi_i^* p_{ij} = \left( \sum_{i \in S} \pi_i^* \right) \mathbb{P}(X_n = j) = 1 \cdot \mathbb{P}(X_n = j) = \pi_j^*$$

Moreover, notice that in this example, $\pi^{(0)} = \pi^*$, so the chain is in stationary state from the beginning.

**Definition 1.9.** A distribution $\pi^* = (\pi_i^*,\, i \in S)$ is said to be a *limiting distribution* for the Markov chain $(X_n,\, n \geq 0)$ if for every initial distribution $\pi^{(0)}$ of the chain, we have

$$\lim_{n \to \infty} \pi_i^{(n)} = \pi_i^*, \quad \forall i \in S \tag{4}$$

**Remarks.**
* If $\pi^*$ is a limiting distribution, then it is stationary. Indeed, for all $n \geq 0$, we always have $\pi^{(n+1)} = \pi^{(n)} P$. If $\lim_{n \to \infty} \pi^{(n)} = \pi^*$, then from the previous equation (and modulo a technical detail), we deduce that $\pi^* = \pi^* P$.
* A limiting distribution $\pi^*$ does not necessarily exist, but if it exists, then it is unique.

The following theorem is central to the theory of Markov chains.

**Theorem 1.10.** Let $(X_n,\, n \geq 0)$ be an irreducible and aperiodic Markov chain. Let us moreover assume that it admits a stationary distribution $\pi^*$. Then $\pi^*$ is a limiting distribution, i.e. for any initial distribution $\pi^{(0)}$, $\lim_{n \to \infty} \pi_i^{(n)} = \pi_i^*$, $\forall i \in S$.

We sketch the proof of this theorem below.

*Proof.* (sketch)
The idea behind the proof is the following *coupling argument*. Let $(X_n, n \geq 0)$ be the Markov chain decribed above and let $(Y_n,\, n \geq 0)$ be an independent replica of this one, except for the fact that $Y$ starts with initial distribution $\pi^*$ (so $\mathbb{P}(Y_n = i) = \pi_i^*$ for all $n \geq 0$ and all $i \in S$).

Let us now look at the bivariate process $(Z_n = (X_n, Y_n),\, n \geq 0)$. It can be shown that $Z$ is also a Markov chain, with state space $S \times S$ and transition matrix

$$\mathbb{P}(Z_{n+1} = (j, l) | Z_n = (i, k)) = p_{ij}\, p_{kl}$$

As $X$ and $Y$ are both irreducible and aperiodic, $Z$ is also irreducible and aperiodic. It also admits the following joint stationary distribution: $\Pi^*_{(i,k)} = \pi_i^*\, \pi_k^*$. We now use the following fact:

If a Markov chain is irreducible and admits a stationary distribution, then it is recurrent.

(This fact can be shown by contradiction: if an irreducible Markov chain is transient, then it cannot admit a stationary distribution.)

So $Z$ is recurrent, which implies the following: let $\tau = \inf\{n \geq 0 : X_n = Y_n\}$ be the first time that the two chains $X$ and $Y$ meet. One can show that for all $n \geq 0$ and $i \in S$,

$$\mathbb{P}(X_n = i, \tau \leq n) = \mathbb{P}(Y_n = i, \tau \leq n)$$

But as $Z$ is recurrent, we also have that $\mathbb{P}(\tau < \infty) = 1$, whatever the initial distribution $\pi^{(0)}$ of the chain $X$. So we obtain for $i \in S$:

$$\begin{aligned} |\pi_i^{(n)} - \pi_i^*| &= |\mathbb{P}(X_n = i) - \mathbb{P}(Y_n = i)| \\ &= |\mathbb{P}(X_n = i, \tau \leq n) - \mathbb{P}(Y_n = i, \tau \leq n)| + |\mathbb{P}(X_n = i, \tau > n) - \mathbb{P}(Y_n = i, \tau > n)| \\ &\leq 0 + \mathbb{P}(\tau > n) \underset{n\to\infty}{\to} 0 \end{aligned}$$

as $\mathbb{P}(\tau < \infty) = 1$. So $\pi^*$ is a limiting distribution. □

Another equally important theorem is the following, However, its proof is more involved and will be skipped.

**Theorem 1.11.** Let $(X_n,\, n \geq 0)$ be an irreducible and positive recurrent Markov chain. Then $X$ admits a unique stationary distribution $\pi^*$.

**Remark.**
An irreducible finite-state Markov chain is always positive recurrent. So by the above theorem, it always admits a unique stationary distribution.

**Definition 1.12.** A (time-homogeneous) Markov chain $(X_n,\, n \geq 0)$ is said to be *ergodic* if it is irreducible, aperiodic and positive recurrent.

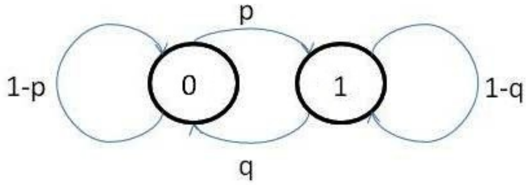
With this definition in hand, we obtain the following corollary of Theorems 1.10 and 1.11.

**Corollary 1.13.** An ergodic Markov chain $(X_n,\, n \geq 0)$ admits a unique stationary distribution $\pi^*$. Moreover, this distribution is also a limiting distribution, i.e.

$$\lim_{n\to\infty} \pi_i^{(n)} = \pi_i^*, \quad \forall i \in S$$

We give below a list of examples illustrating the previous theorems.

**Example 1.14.** (two-state Markov chain)
Let us consider a two-state Markov chain with the following transition graph (where $0 < p, q \leq 1$):



As both $p, q > 0$, this chain is clearly irreducible, and as it is finite-state, it is also positive recurrent. So by Theorem 1.11, it admits a stationary distribution. Writing down the equation for the stationary distribution $\pi = \pi P$, we obtain

$$\pi_0 = \pi_0 (1-p) + \pi_1 q, \quad \pi_1 = \pi_1 q + \pi_1 (1-p) \tag{5}$$

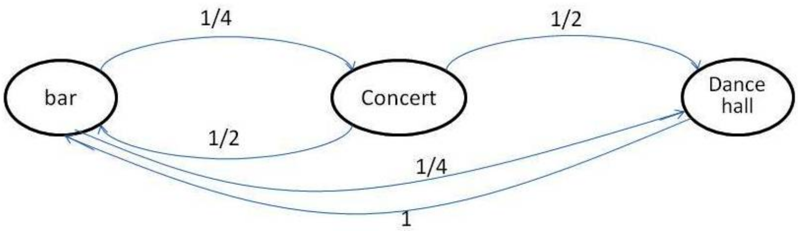
Remember also that as $\pi$ is a distribution, we must have $\pi_0 + \pi_1 = 1$ and $\pi_0, \pi_1 \geq 0$. Solving these equations (and noticing that the two equations in (5) are actually redundant), we obtain

$$\pi_0 = \pi_0 (1-p) + (1-\pi_0) q \quad \Rightarrow \quad \pi_0 (p+q) = q, \quad \text{i.e.} \quad \pi_0 = \frac{q}{p+q}$$

so $\pi^* = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ is the stationary distribution. Moreover, if $p+q < 2$ (i.e. if it is not the case that both $p = q = 1$), then the chain is also aperiodic and therefore ergodic, so by Corollary 1.13, $\pi^* = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$ is also a limiting distribution.

Notice that when both $p = q = 1$, then $\pi^* = \left(\frac{1}{2}, \frac{1}{2}\right)$ is the unique stationary distribution of the chain, but in this case, the chain is periodic (with period $d = 2$) and $\pi^*$ is *not* a limiting distribution. If for example the chain starts in state 0, then the distribution of the chain will switch from $\pi^{(n)} = (1, 0)$ at even times to $\pi^{(n)} = (0, 1)$ at odd times, and reciprocally, but it will never converge to the stationary distribution $\pi^* = \left(\frac{1}{2}, \frac{1}{2}\right)$.

**Example 1.15.** (music festival: modified version)
Let us consider the chain with the following transition graph:




It has the corresponding transition matrix:

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

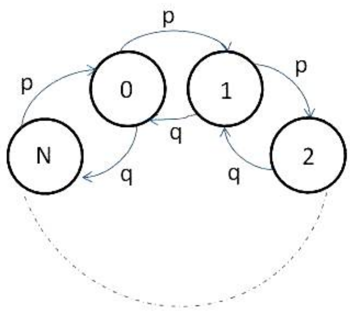
We can again easily see that the chain is ergodic. The computation of its stationary and limiting distribution gives

$$\pi^* = \left(\frac{8}{13}, \frac{3}{13}, \frac{2}{13}\right)$$

Quite unexpectedly, the student spends most of the time at the bar...

**Example 1.16.** (simple symmetric random walk)
Let us consider the simple symmetric random walk of Example 1.3. This chain is irreducible, periodic with period 2 and all states are null recurrent. There does not exist a stationary distribution here (NB: it should be the uniform distribution on $\mathbb{Z}$, which does not exist).

**Example 1.17.** (cyclic random walk on $\{0, 1, \dots, N\}$)
Let us consider the chain with the following transition graph (with $0 < p, q < 1$ and $p + q = 1$):




It has the corresponding transition matrix:

$$P = \begin{pmatrix} 0 & p & 0 & 0 & q \\ q & 0 & p & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & q & 0 & p \\ p & 0 & 0 & q & 0 \end{pmatrix}$$

This chain is irreducible and finite-state, so it is also positive recurrent, but its periodicity depends on the value of $N$: if $N$ is odd (that is, the number of states is even), then the chain is periodic with period 2; if on the contrary $N$ is even (that is, the number of states is odd), then the chain is aperiodic. In order to find its stationary distribution, observe that for all $j \in S$, $\sum_{i \in S} p_{ij} = p + q = 1$, so we can use Proposition 1.18 below to conclude that $\pi^* = \left(\frac{1}{N+1}, \dots, \frac{1}{N+1}\right)$, In case $N$ is even, this distribution is also a limiting distribution.

**Proposition 1.18.** Let $(X_n,\ n \geq 0)$ be a finite-state irreducible Markov chain with state space $S = \{0, \dots, N\}$ and let $\pi^*$ be its unique stationary distribution (whose existence is ensured by Theorem 1.11 and the remark following it). Then $\pi^*$ is the uniform distribution if and only if the transition matrix $P$ of the chain satisfies:

$$\sum_{i \in S} p_{ij} = 1, \quad \forall j \in S$$

**Remark.**
Notice that the above condition is saying that the *columns* of the matrix $P$ should sum up to 1, which is different from the condition seen at the beginning that the *rows* of the matrix $P$ should sum up to 1 (satisfied by *any* transition matrix).

*Proof.* * If $\pi^* = \left(\frac{1}{N+1}, \dots, \frac{1}{N+1}\right)$ is a stationary distribution, then

$$\frac{1}{N+1} = \sum_{i \in S} \frac{1}{N+1}\, p_{ij}, \quad \forall j \in S, \quad \text{i.e.} \quad \sum_{i \in S} p_{ij} = 1, \quad \forall j \in S$$

* If $\sum_{i \in S} p_{ij} = 1, \forall j \in S$, then one can simply check that $\pi^* = \left(\frac{1}{N+1}, \dots, \frac{1}{N+1}\right)$ satisfies the equation $\pi^* = \pi^* P$. □

**What happens without the aperiodicity assumption?**

**Theorem 1.19.** Let $(X_n,\, n \geq 0)$ be an irreducible and positive recurrent Markov chain, and let $\pi^*$ be its unique stationary distribution (whose existence is ensured by Theorem 1.11). Then for any initial distribution $\pi^{(0)}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \pi_i^{(k)} = \pi_i^*, \quad \forall i \in S$$

In this sense, $\pi_i^*$ can still be interpreted as the average proportion of time spent by the chain in state $i$, and it also holds that

$$\mathbb{E}(T_i | X_0 = i) = \frac{1}{\pi_i^*}, \quad \forall i \in S$$

## 1.4 Reversible Markov chains and detailed balance

Let $(X_n,\, n \geq 0)$ be a time-homogeneous Markov chain. Let us now consider this chain *backwards*, i.e. consider the process $(X_n, X_{n-1}, X_{n-2}, \ldots, X_1, X_0)$: this process turns out to be also a Markov chain (but not necessarily time-homogeneous). Indeed:

$$\begin{aligned}
&\mathbb{P}(X_n = j | X_{n+1} = i, X_{n+2} = k, X_{n+3} = l, \ldots) \\
&\quad = \frac{\mathbb{P}(X_n = j, X_{n+1} = i, X_{n+2} = k, X_{n+3} = l, \ldots)}{\mathbb{P}(X_{n+1} = i, X_{n+2} = k, X_{n+3} = l, \ldots)} \\
&\quad = \frac{\mathbb{P}(X_{n+2} = k, X_{n+3} = l, ..., | X_{n+1} = i, X_n = j)}{\mathbb{P}(X_{n+2} = k, X_{n+3} = l, ..., | X_{n+1} = i)} \, \frac{\mathbb{P}(X_{n+1} = j, X_n = j)}{\mathbb{P}(X_{n+1} = i)} \\
&\quad = \frac{\mathbb{P}(X_{n+2} = k, X_{n+3} = l, ..., | X_{n+1} = i)}{\mathbb{P}(X_{n+2} = k, X_{n+3} = l, ..., | X_{n+1} = i)} \, \mathbb{P}(X_n = j | X_{n+1} = i) \qquad (6) \\
&\quad = \mathbb{P}(X_n = j | X_{n+1} = i)
\end{aligned}$$

where (6) follows from the Markov property of the forward chain $X$.

Let us now compute the transition probabilities:

$$\mathbb{P}(X_n = j | X_{n+1} = i) = \frac{\mathbb{P}(X_n = j, X_{n+1} = i)}{\mathbb{P}(X_{n+1} = i)} = \frac{\mathbb{P}(X_{n+1} = i | X_n = j)\, \mathbb{P}(X_n = j)}{\mathbb{P}(X_{n+1} = i)} = \frac{p_{ji}\, \pi_j^{(n)}}{\pi_i^{(n+1)}}$$

We observe that these transition probabilities may depend on $n$, so the backward chain is not necessarily time-homogeneous, as mentioned above.

Let us now assume that the chain is irreducible and positive recurrent. Then by Theorem 1.11, it admits a unique stationary distribution $\pi^*$. Let us moreover assume that the initial distribution of the chain is the stationary distribution (so the chain is in stationary state: $\pi^{(n)} = \pi^*$, $\forall n \geq 0$, i.e. $\mathbb{P}(X_n = i) = \pi_i^*$, $\forall n \geq 0$, $\forall i \in S$). In this case,

$$\mathbb{P}(X_n = j | X_{n+1} = i) = \frac{p_{ji}\, \pi_j^*}{\pi_i^*} = \widetilde{p}_{ij}$$

i.e. the backward chain is time-homogeneous with transition probabilities $\widetilde{p}_{ij}$.

**Definition 1.20.** The chain $X$ is said to be *reversible* if $\widetilde{p}_{ij} = p_{ij}$, i.e. the transition probabilities of the forward and the backward chains are equal. In this case, the following *detailed balance equation* is satisfied:

$$\pi_i^* \, p_{ij} = \pi_j^* \, p_{ji}, \quad \forall i, j \in S \tag{7}$$

**Remarks.**
* If a distribution $\pi^*$ satisfies the above detailed balance equation, then it is a stationary distribution. Indeed, if $\pi^*$ satisfies (7), then

$$\sum_{i \in S} \pi_i^* \, p_{ij} = \sum_{i \in S} \pi_j^* \, p_{ji} = \pi_j^* \sum_{i \in S} p_{ji} = \pi_j^*, \quad \forall j \in S$$

* In order to find the stationary distribution of a chain, solving the detailed balance equation (7) is easier than solving the stationary distribution equation (3), but this works of course only if the chain is reversible.
* Equation (7) has the following interpretation: it says that in the Markov chain, the flow from state $i$ to state $j$ is equal to that from state $j$ to state $i$.
* If equation (7) is satisfied, then $\pi^*$ is the uniform distribution if and only if $P$ is a symmetric matrix.

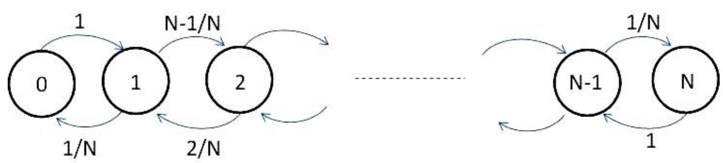
**Example 1.21.** (cyclic random walk on $\{0, 1, \ldots, N\}$)
Let us consider the cyclic random walk on $\{0, 1, \ldots, N\}$ of Example 1.17 with right and left transition probabilities $p$ and $q$ ($p + q = 1$). We have seen that the unique stationary distribution of this chain is the uniform distribution $\pi^* = \left(\frac{1}{N+1}, \ldots, \frac{1}{N+1}\right)$. Is it the case that the detailed balance equation is satisfied here? By the above remark, this happens only when the transition matrix $P$ is symmetric, i.e. when $p = q = 1/2$. Otherwise, we see that the flow of the Markov chain is more important in one direction than in the other.

**Example 1.22.** (Ehrenfest urns)
Let us consider the following process: there are 2 urns and $N$ numbered balls. At each time step, one ball is picked uniformly at random among the $N$ balls, and transferred from the urn where it lies to the other urn.

Let now $X_n$ be the number of balls located in the first urn at time $n$. The process $X$ describes a Markov chain on $\{0, \ldots, N\}$, whose transition probabilities are given by

$$p_{i,i+1} = \frac{N-i}{N} \quad p_{i,i-1} = \frac{i}{N} \quad \forall 1 \le i \le N-1 \quad \text{and} \quad p_{01} = 1, \quad p_{N,N-1} = 1$$

The corresponding transition graph is



This chain is clearly irreducible, periodic with period 2 and positive recurrent, so it admits a unique stationary distribution $\pi^*$. A priori, we are not sure that the chain is reversible (although it is a reasonable guess in the present case), but we can still try solving the detailed balance equation and see where it leads:

$$\pi_i^* \, p_{i,i+1} = \pi_{j+1}^* p_{i+1,i} \quad \text{i.e.} \quad \pi_i^* \, \frac{N-i}{N} = \pi_{i+1}^* \, \frac{i+1}{N} \quad \Rightarrow \quad \pi_{i+1}^* = \frac{N-i}{i+1} \, \pi_i^*$$

So by induction, we obtain

$$\pi_i^* = \frac{(N-i+1)\cdots N}{i\cdots 1} \, \pi_0^* = \frac{N!}{(N-i)!\, i!} \, \pi_0^* = \left( \begin{array}{c} N \\ i \end{array} \right) \pi_0^*$$

Writing down the normalization condition $\sum_{i=0}^N \pi_i^* = 1$, we obtain

$$1 = \pi_0^* \sum_{i=0}^n \left( \begin{array}{c} N \\ i \end{array} \right) = \pi_0^* \, 2^N \quad \text{so} \quad \pi_i^* = 2^{-N} \left( \begin{array}{c} N \\ i \end{array} \right), \quad i = 0, 1, ..., N$$

**Remark.**
In physics, this process models the diffusion of particles across a porous membrane. It leads to the following paradox: assume the chain starts in state $X_0 = 0$ (that is, all the particles are on one side of the membrane), and let then the chain evolve over time. As the chain is recurrent, it will come back infinitely often to its initial state $0$. This seems a priori in contradiction with the second principle of thermodynamics, which states that the entropy of a physical system should not decrease. Here, the entropy of the state $0$ is much less than that of any other state in the middle, so the chain should not come back to $0$ after having visited states in the middle. The paradox has been resolved by observing that for macroscopic systems (that is, $N \sim 6,022 \times 10^{23}$, the Avogadro number), the recurrence to state $0$ is never observed in practice, as $\pi_0^* = 2^{-N}$.

## 1.5 Hitting times and absorption probabilities

Let $(X_n,\, n \geq 0)$ be a Markov chain with state space $S$ and transition matrix $P$ and let $A$ be a subset of the state space $S$ (notice that $A$ need not be a class). In this section, we are interested in knowing what is the probability that the Markov chain $X$ reaches a state in $A$. For this purpose, we introduce the following definitions.

**Definition 1.23.**
* *Hitting time*: $H_A = \inf\{n \geq 0 : X_n \in A\}$ = the first time the chain $X$ "hits" the subset $A$.
* *Hitting probability*: $h_{iA} = \mathbb{P}(H_A < \infty | X_0 = i) = \mathbb{P}(\exists n \geq 0$ such that $X_n \in A | X_0 = i)$, $i \in S$.

**Remarks.**
* The time $H_A$ to hit a given set $A$ might be infinity (if the chain never hits $A$).
* On the contrary, we say by convention that if $X_0 = i$ and $i \in A$, then $H_A = 0$ and $h_{iA} = 1$.
* If $A$ is an absorbing set of states (i.e. there is no way for the chain to leave the set $A$ once it has entered it), then the probability $h_{iA}$ is called an *absorption probability*. A particular case that will be of interest to us is when $A$ is a single absorbing state.

The following theorem allows to compute the vector of hitting probabilities $h_A = (h_{iA},\, i \in S)$.

**Theorem 1.24.** The vector $h_A = (h_{iA},\, i \in S)$ is the *minimal non-negative* solution of the following equation:

$$\begin{cases} h_{iA} = 1 & \forall i \in A \\ h_{iA} = \sum_{j \in S} p_{ij}\, h_{jA} & \forall i \notin A \end{cases} \tag{8}$$

By *minimal* solution, we mean that if $g_A = (g_{iA},\, i \in S)$ is another solution of (8), then $g_{iA} \geq h_{iA}, \forall i \in S$.

**Remarks.**
* The vector $h_A$ is *not* a probability distribution, i.e. we do *not* have $\sum_{i \in S} h_{iA} = 1$.
* This theorem is nice, but notice that in order to compute a single hitting probability $h_{iA}$, one needs a priori to solve the equation for the entire vector $h_A$. It turns out however in many situations that solving the equation is much easier than computing directly hitting probabilities.

*Proof.* * Let us first prove that $h_A$ is a solution of (8). If $i \in A$, then $h_{iA} = 1$, as $H_A = 0$ in this case. If $i \notin A$, then

$$\begin{aligned} h_{iA} &= \mathbb{P}(\exists n \geq 0 : X_n \in A | X_0 = i) = \mathbb{P}(\exists n \geq 1 : X_n \in A | X_0 = i) \\ &= \sum_{j \in S} \mathbb{P}(\exists n \geq 1 : X_n \in A | X_1 = j, X_0 = i)\, \mathbb{P}(X_1 = j | X_0 = i) \\ &= \sum_{j \in S} \mathbb{P}(\exists n \geq 1 : X_n \in A | X_1 = j)\, \mathbb{P}(X_1 = j | X_0 = i) && (9) \\ &= \sum_{j \in S} \mathbb{P}(\exists n \geq 0 : X_n \in A | X_0 = j)\, p_{ij} = \sum_{j \in S} h_{jA}\, p_{ij} && (10) \end{aligned}$$

where (9) follows from the Markov property and (10) follows from time-homogeneity.

Notice that if the state space $S$ is finite, then it can be proved that there is a unique solution to equation (8), so the proof stops here.

* In general however, there might exist multiple solutions to equation (8). Let us then prove that $h_A$ is minimal among these. For this purpose, assume $g_A$ is another solution of (8). We want to prove that $g_{iA} \geq h_{iA}, \forall i \in S$. As $g_A$ is a solution, we obtain the following.

If $i \in A$, then $g_{iA} = 1 = h_{iA}$. If $i \notin A$, then

$$\begin{aligned} g_{iA} &= \sum_{j \in S} p_{ij} g_{jA} = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} g_{jA} = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \left( \sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk}\, g_{kA} \right) \\ &= \mathbb{P}(X_1 \in A | X_0 = i) + \mathbb{P}(X_2 \in A, X_1 \notin A | X_0 = i) + \sum_{j \notin A} \sum_{k \notin A} p_{ij}\, p_{jk}\, g_{kA} \\ &= \mathbb{P}(X_1 \in A \text{ or } X_2 \in A | X_0 = i) + \sum_{j \notin A} \sum_{k \notin A} p_{ij}\, p_{jk}\, g_{kA} \end{aligned}$$

Observe that the last term on the right-hand side is non-negative, so

$$g_{iA} \geq \mathbb{P}(X_1 \in A \text{ or } X_2 \in A | X_0 = i)$$

This procedure can be iterated further and gives, for any $n \geq 1$:

$$g_{iA} \geq \mathbb{P}(X_1 \in A \text{ or } X_2 \in A \text{ or ... or } X_n \in A | X_0 = i)$$

So finally, we obtain

$$g_{iA} \geq \mathbb{P}(\exists n \geq 1 : X_n \in A | X_0 = i) = \mathbb{P}(\exists n \geq 0 : X_n \in A | X_0 = i) = h_{iA}$$

which completes the proof. □

We are also interested in knowing how long does the Markov chain $X$ need to reach a state in $A$ on average. For this purpose, let us introduce the following definition.

**Definition 1.25.** The *average hitting time* of a set $A$ from a state $i \in S$ is defined as

$$\mu_{iA} = \mathbb{E}(H_A | X_0 = i) = \sum_{n \geq 0} n \, \mathbb{P}(H_A = n | X_0 = i)$$

Notice that this average hitting time might be $\infty$. The following theorem allows to compute the vector of average hitting times $\mu_A = (\mu_{iA}, \, i \in S)$. As its proof follows closely the one of Theorem 1.24, we do not repeat it here.

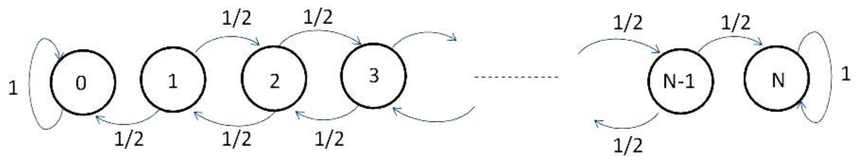
**Theorem 1.26.** The vector $\mu_A = (\mu_{iA}, i \in S)$ is the minimal non-negative solution of the following equation:

$$\begin{cases} \mu_{iA} = 0 & \forall i \in A \\ \mu_{iA} = 1 + \sum_{j \notin A} p_{ij} \, \mu_{jA} & \forall i \notin A \end{cases} \tag{11}$$

Please pay attention that this equation is similar to equation (8), but of course not identical.

We list below a series of examples where the above two theorems can be used.

**Example 1.27.** (gambler's ruin on $\{0, 1, 2, ..., N\}$)
Let us consider the time-homogeneous Markov chain with the following transition graph:




This chain models the following situation: a gambler plays "heads or tails" repeatedly, and each time wins or loses one euro with equal probability $1/2$; he plays until he either loses everything or wins a total amount of $N$ euros. Assuming that he starts with $i$ euros (with $1 \leq i \leq N-1$), what is the probability that he loses everything?

The answer is $h_{i0} = \mathbb{P}(H_0 < \infty | X_0 = i)$ (indeed, the only alternative is $H_N < \infty$). Let us therefore try solving equation (8):

$$\begin{cases} i = 0 : & h_{00} = 1 \\ 1 \leq i \leq N-1 : & h_{i0} = \frac{1}{2}(h_{i-1,0} + h_{i+1,0}) \quad \text{i.e.} \quad h_{i+1,0} = 2h_{io} - h_{i-1,0} \\ i = N : & h_{N0} = 0 \end{cases} \tag{12}$$

Notice that there is actually no equation for $h_{N0}$; we therefore choose the smallest non-negative value, i.e. $0$ (another view on this is that we know that $h_{N0} = 0$, as $0$ is not accessilble from $N$). This gives

$$h_{20} = 2h_{10} - 1, \quad h_{30} = 2h_{20} - h_{10} = 3h_{10} - 2, \quad \ldots$$

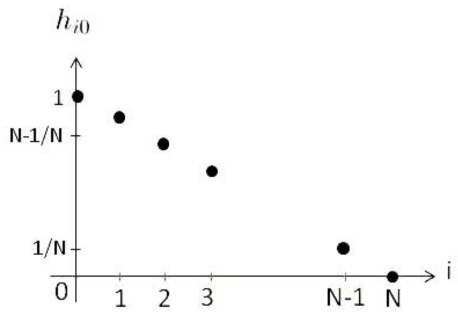
By induction, we obtain $h_{i0} = ih_{10} - (i-1), \forall 1 \leq i \leq N-1$.

Writing down equation (12) for $i = N-1$, we further obatin

$$0 = h_{N0} = 2h_{N-1,0} - h_{N-2,0} = 2(N-1)h_{10} - 2(N-2) - (N-2)h_{10} + (N-3)$$

Therefore $Nh_{10} - N + 1 = 0$, i.e.

$$h_{10} = \frac{N-1}{N} \quad \text{and} \quad h_{i0} = i\,\frac{N-1}{N} - (i-1) = \frac{iN - i - Ni + N}{N} = \frac{N-i}{N}$$




Here is now a second question: how long will the game last on average (until the gambler either loses everything or wins $N$ euros), assuming again the gambler starts with $i$ euros ($1 \leq i \leq N-1$)?

The answer is the following: let us consider the subset $A = \{0, N\}$; the average duration of the game is $\mu_{iA} = \mathbb{E}(H_A | X_0 = i)$. Notice that $h_{iA} = 1$ (as there is no other alternative than to end in $0$ or $N$) and also that the chain has a finite number of states, so $\mu_{iA} < \infty$ (whereas it can be checked that both $\mu_{i0} = \mu_{iN} = \infty$). The equation (11) for the vector $\mu_A$ reads in this case:

$$\begin{cases} i = 0 : & \mu_{0A} = 0 \\ 1 \leq i \leq N-1 : & \mu_{iA} = 1 + \frac{1}{2}(\mu_{i-1,A} + \mu_{i+1,A}) \quad \text{i.e.} \quad \mu_{i+1,A} = 2\mu_{iA} - 2 - \mu_{i-1,A} \\ i = N : & \mu_{NA} = 0 \end{cases} \tag{13}$$

The solution of this equation is obtained similarly to the previous one:

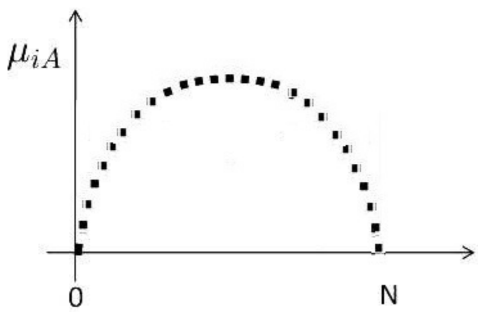
$$\mu_{2A} = 2\mu_{1A} - 2, \quad \mu_{3A} = 2\mu_{2A} - 2 - 2\mu_{1A} = 3\mu_{1A} - 6, \quad \ldots$$

so by induction, we obtain: $\mu_{iA} = i\mu_{1A} - i(i-1)$

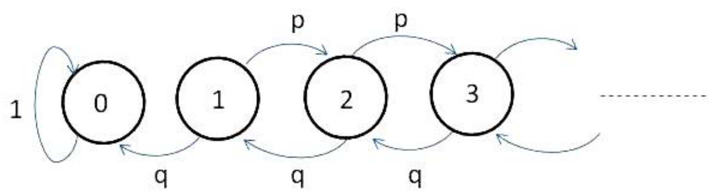
Writing down equation (13) for $i = N-1$, we further obtain

$$\begin{aligned} 0 = \mu_{NA} &= 2\mu_{N-1,A} - 2 - \mu_{N-2,A} \\ &= 2(N-1)\mu_{1A} - 2(N-1)(N-2) - 2 - (N-2)\mu_{1A} + (N-2)(N-3) \\ &= N\mu_{1A} - (N-2)(2(N-1) - (N-3)) - 2 \\ &= N\mu_{1A} - (N^2 - N - 2) - 2 = N\mu_{1A} - N^2 + N \end{aligned}$$

So $\mu_{1A} = \frac{N^2-N}{N} = N-1$ and $\mu_{iA} = i(N-1) - i(i-1) = i(N-i)$.




**Example 1.28.** (gambler's ruin on $\mathbb{N}$)
Let us consider the time-homogeneous Markov chain with the following transition graph:




This Markov chain describes a gambler playing repeatedly until he loses everything (there is no more upper limit $N$), winning each game with probability $0 < p < 1$ and losing with probability $q = 1 - p$. Starting from a fortune of $i$ euros, what is the probability that the gambler loses everything?

The answer is again $h_{i0}$, so let us try solving equation (8) (first assuming that $p \neq 1/2$):

$$h_{00} = 1, \quad h_{i0} = p h_{i+1,0} + q h_{i-1,0} \quad i \geq 1$$

The general solution of this difference equation is given by

$$h_{i0} = \alpha\, y_+^i + \beta\, y_-^i$$

where $y_\pm$ are the two solutions of the quadratic equation $y = py^2 + q$, i.e. $y_+ = 1$, $y_- = q/p$. Therefore,

$$h_{i0} = \alpha + \beta\, (q/p)^i$$

Using the boundary condition $h_{10} = ph_{20} + q$, we moreover obtain that $\alpha + \beta = 1$, i.e.

$$h_{i0} = \alpha + (1-\alpha)\, (q/p)^i$$

For any $\alpha \in [0, 1]$, the above expression is a non-negative solution of equation (8). The parameter $\alpha$ remains to be determined, using the fact that we are looking for the *minimal* solution.

* if $p < q$, then the minimal solution is given by $h_{i0} = 1$, $\forall i$ (i.e. $\alpha = 1$)

* If $p > q$, then the minimal solution is given by $h_{i0} = (q/p)^i$, $\forall i$ (i.e. $\alpha = 0$)

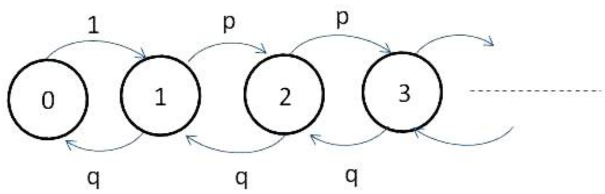
In the borderline case where $p = q = 1/2$, following what has been done in the previous example leads to

$$h_{i0} = ih_{10} - (i - 1), \quad \forall i$$

and we see that the minimal *non-negative* solution is actually given by $h_{10} = 1$, leading to $h_{i0} = 1$ for all $i$.

In conclusion, as soon as $p \leq 1/2$, the gambler is guaranteed to lose everything with probability 1, whatever his initial fortune.

**Remarks.**
* These absorption probabilities we just computed are also the hitting probabilities of the Markov chain with the following transition graph:




And for this chain, the probabilities $h_{10} = \mathbb{P}(H_0 < \infty | X_0 = 1)$ and $f_0 = \mathbb{P}(T_0 < \infty | X_0 = 0)$ are equal! So this new chain is recurrent if and only if $h_{10} = 1$, that is, if and only if $p \leq 1/2$.

* In the case $p = 1/2$, we can also compute the average hitting times $\mu_{i0}$, following what has been done in the Example 1.27. We obtain:

$$\mu_{i0} = i\mu_{10} - i(i - 1), \quad \forall i$$

As $i(i - 1)$ increases faster to $\infty$ than $i$, we see that the vector $\mu_0$ can be non-negative only if $\mu_{10} = \infty$ itself, i.e. $\mu_{i0} = \infty$ for all $i$. This is saying that in this case, the average time to reach $0$ from any starting point $i$ is actually infinite!

* Making now the connection between these two remarks, we see that for the chain described above, we have

$$\infty = \mu_{10} = \mathbb{E}(H_0 | X_0 = 1) = \mathbb{E}(T_0 | X_0 = 0)$$

i.e. the expected return time to state $0$ is infinite, so the chain is null recurrent when $p = 1/2$ (similarly, it can be argued that the chain is positive recurrent when $p < 1/2$).

### 1.5.1 Application: branching processes

Here is a simple (not to say simplistic) model of evolution of the number of individuals in a population over the generations. Let first $(p_j,\, j \geq 0)$ be a given probability distribution.

Let now $X_n$ describe the number of individuals in the population at generation $n$. At each generation $n$, each individual $i \in \{1, ..., X_n\}$ has $C_i^n$ children, where $(C_i^n,\, i \geq 1,\, n \geq 0)$ are i.i.d. random variables with distribution $\mathbb{P}(C_i^n = j) = p_j$, $j \geq 0$. The number of individuals at generation $n+1$ is therefore:

$$X_{n+1} = C_1^n + ... + C_{X_n}^n$$

Because the random variable $C_i^n$ are i.i.d., the process $(X_n,\, n \geq 0)$ is a time-homogeneous Markov chain (what happens to generation $n+1$ only depends on the value $X_n$, not on what happened before). Let us moreover assume that the population starts with $i > 0$ individuals.

We are interested in computing the extinction probability of this population, namely:

$$h_{i0} = \mathbb{P}(X_n = 0 \text{ for some } n \geq 1 | X_0 = i).$$

This model was originally introduced by Galton and Watson in the 19th century in order to study the extinction of surnames in noble families. It nowadays has found numerous applications in biology, and numerous variants of the model exist also.

**Remarks.**
* If $p_0 = \mathbb{P}(C_i^n = 0) = 0$, then the extinction probability $h_{i0} = 0$, trivially; let us therefore assume that $p_0 > 0$. In this case, 0 is an absorbing state and all the other states are transient.
* If a population starts with $i$ individuals, then if extinction occurs, it has to occur for the family tree each of the $i$ ancestors. So because of the i.i.d. assumption, the total extinction probability is the product of the extinction probabilities of each subtree, i.e. $h_{i0} = (h_{10})^i$.
* As a corollary, the fact that extinction occurs with probability 1 or not does not depend on the initial number of individuals in the population.
* For $i = 1$, the transition probabilities have the following simple expression:

$$p_{1j} = \mathbb{P}(X_{n+1} = j | X_n = 1) = \mathbb{P}(C_1^n = j) = p_j.$$

From Theorem 1.24, we know that the vector $h_0 = (h_{i0},\, i \geq 0)$ is the minimal non negative solution of

$$h_{00} = 1, \quad h_{i0} = \sum_{j \geq 0} p_{ij}\, h_{j0}, \quad i \geq 1$$

In particular, we obtain the following closed equation for $h_{10}$:

$$h_{10} = \sum_{j \geq 0} p_{1j}\, h_{j0} = \sum_{j \geq 0} p_j\, (h_{10})^j \tag{14}$$

In order to solve this equation for $h_{10}$ (remembering that we are looking for the minimal non-negative solution), let us define the generating function

$$g(z) = \sum_{j \geq 0} p_j z^j, \quad z \in [0, 1]$$

Equation (14) can therefore be rewritten as the fixed point equation $h_{10} = g(h_{10})$. Its minimal non-negative solution is given by the following proposition.

**Proposition 1.29.** Let $\mu_c = \sum_{j\geq 1} p_j\, j$ be the average number of children of a given individual.

* If $\mu_c \leq 1$, then $h_{10} = 1$, i.e. extinction occurs with probability $1$.

* If $\mu_c > 1$, then the minimal solution of $h_{10} = g(h_{10})$ is a number strictly between $0$ and $1$, so both extinction and survival occur with positive probability.

From this proposition, we see that slightly more than one child per individual is needed on average in order for the population to survive. But of course, there is always a positive probability that at some generation, no individual has a child and the population gets extinct.

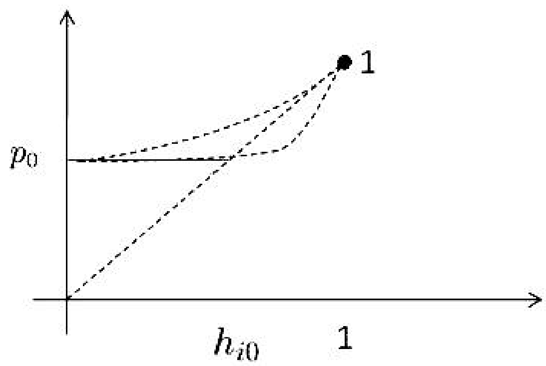
*Proof.* Let us analyze the properties of the generating function $g$:

* $g(0) = p_0 \in ]0, 1]$, $g(1) = \sum_{j\geq 0} p_j = 1$

* $g'(z) = \sum_{j\geq 1} p_j\, j\, z^{i-1}$, so $g'(1) = \sum_{j\geq 1} p_j\, j = \mu_c$.

* $g''(z) = \sum_{j\geq 2} p_j\, j(j-1)\, z^{j-2} \geq 0$, so $g$ is a convex function.

Given these properties, we see that only two things can happen:




* Either $\mu_c \leq 1$ (top curve), and then the unique solution to equation $h_{10} = g(h_{10})$ is $h_{10}^* = 1$.

* Or $\mu_c > 1$ (bottom curve), and then equation $h_{10} = g(h_{10})$ admits two solutions, the minimal of which is a number $h_{10}^* \in ]0, 1[$. □

# 2 Continuous-time Markov chains

## 2.1 The Poisson process

**Preliminary.** (convergence of the binomial distribution towards the Poisson distribution)
Let $c > 0$ and $X_1, \ldots, X_M$ be i.i.d. random variables such that $\mathbb{P}(X_i = +1) = c/M$ and $\mathbb{P}(X_i = 0) = 1 - (c/M)$, for $1 \leq i \leq M$.

Let also $Z_M = X_1 + \ldots + X_M$. Then $Z_M \sim \text{Bi}(M, c/M)$, i.e.

$$\mathbb{P}(Z_M = k) = \left( \begin{array}{c} M \\ k \end{array} \right) (c/M)^k \, (1 - (c/M))^{M-k}, \quad 0 \leq k \leq M$$

**Proposition 2.1.** As $M \to \infty$, the distribution of $Z_M$ converges to that of a Poisson random variable with parameter $c > 0$, i.e.

$$\mathbb{P}(Z_M = k) \underset{M \to \infty}{\to} \frac{c^k}{k!} \, e^{-k}, \quad \forall k \geq 0$$

*Proof.* Let us compute

$$\begin{aligned} \mathbb{P}(Z_M = k) &= \left( \begin{array}{c} M \\ k \end{array} \right) (c/M)^k \, (1 - (c/M))^{M-k} \\ &= \frac{M(M-1) \cdot (M-k+1)}{k!} \, \frac{c^k}{M^k} \, (1 - (c/M))^M \, (1 - (c/M))^{-k} \underset{M \to \infty}{\to} \frac{c^k}{k!} \, e^{-c} \end{aligned}$$

□

### 2.1.1 Definition and basic properties

The Poisson process is a continuous-time process counting events taking place at random times, such as e.g. customers arriving at a desk. Its definition follows.

**Definition 2.2.** A continuous-time stochastic process $(N_t, \, t \in \mathbb{R}_+)$ is a *Poisson process with intensity* $\lambda > 0$ if:

* $N$ is integer-valued: $N_t \in \mathbb{N}$, $\forall t \in \mathbb{R}_+$.
* $N_0 = 0$ and $N$ is increasing: $N_s \leq N_t$ if $s \leq t$.
* $N$ has *independent and stationary increments*: for all $0 \leq t_1 \leq \ldots \leq t_m$ and $n_1, \ldots, n_m \in \mathbb{N}$,

$$\begin{aligned} &\mathbb{P}(N_{t_1} = n_1, N_{t_2} - N_{t_1} = n_2, \ldots, N_{t_m} - N_{t_{m-1}} = n_m) \\ &\quad = \mathbb{P}(N_{t_1} = n_1) \, \mathbb{P}(N_{t_2} - N_{t_1} = n_2) \cdots \mathbb{P}(N_{t_m} - N_{t_{m-1}} = n_m) \end{aligned}$$

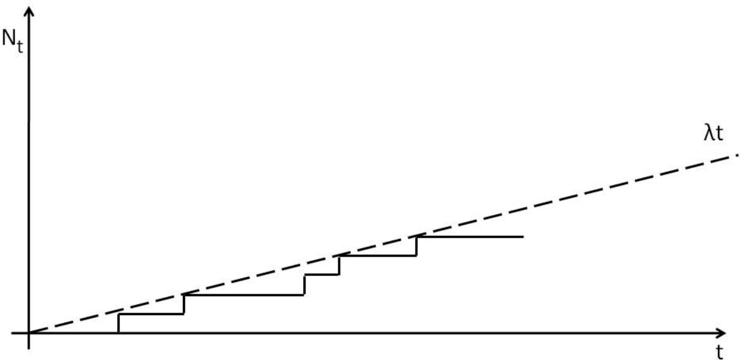
and for all $0 \leq s \leq t$ and $n, m \in \mathbb{N}$,

$$\mathbb{P}(N_t - N_s = n) = \mathbb{P}(N_{t-s} = n)$$

* $\mathbb{P}(N_{\Delta t} = 1) = \lambda \, \Delta t + o(\Delta t)$, $\mathbb{P}(N_{\Delta t} \geq 2) = o(\Delta t)$, where by definition $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$.

NB: As a consequence of this definition, we see that $\mathbb{P}(N_{\Delta t} = 0) = (1 - \lambda \Delta t) + o(\Delta t)$.

**Illustration.** Here is a graphical representation of the time evolution of a Poisson process.




From the above definition, we deduce in the proposition below the distribution of the Poisson process at a given time instant.

**Proposition 2.3.** At time $t \in \mathbb{R}_+$, $N_t$ is a Poisson random variable with parameter $\lambda t$, i.e.

$$\mathbb{P}(N_t = k) = \frac{(\lambda t)^k}{k!}\, e^{-\lambda t}, \quad k \geq 0$$

*Proof.* (sketch)
Let $t \in \mathbb{R}_+$, $M \geq 1$ and define $\Delta t = t/M$. We can write

$$N_t = \sum_{i=1}^{M} X_i, \quad \text{where} \quad X_i = N_{i\Delta t} - N_{(i-1)\Delta t}$$

From the last line of Definition 2.2 and the stationarity property, we deduce that

$$\begin{aligned} \mathbb{P}(X_i = 1) &= \mathbb{P}(N_{\Delta t} = 1) \simeq \lambda \Delta t = \frac{\lambda t}{M} \\ \mathbb{P}(X_i \geq 2) &= \mathbb{P}(N_{\Delta t} \geq 2) \simeq 0 \\ \mathbb{P}(X_i = 0) &= \mathbb{P}(N_{\Delta t} = 0) \simeq 1 - \lambda \Delta t = 1 - \frac{\lambda t}{M} \end{aligned}$$

The random variables $X_i$ can therefore be considered as (nearly) Bernoulli random variables with parameter $\frac{\lambda t}{M}$. Therefore,

$$\begin{aligned} \mathbb{P}(N_t = k) = \mathbb{P}(X_1 + ... + X_M = k) &\simeq \begin{pmatrix} M \\ k \end{pmatrix} \left(\frac{\lambda t}{M}\right)^k \left(1 - \frac{\lambda t}{M}\right)^{M-k} \\ &\underset{M \to \infty}{\to} \frac{(\lambda t)^k}{k!}\, e^{-\lambda t} \end{aligned}$$

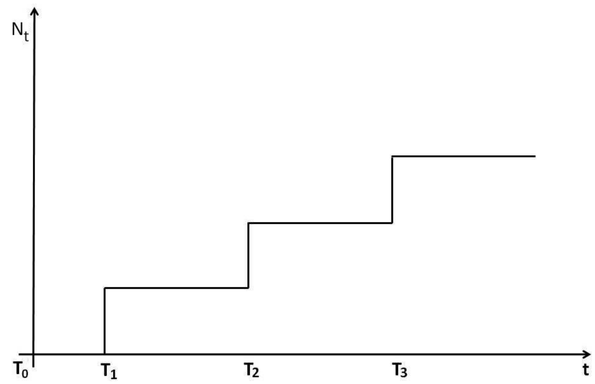
by Proposition 2.1. □

**Corollary 2.4.** For $t \in \mathbb{R}_+$, $\mathbb{E}(N_t) = \lambda t$ (so $\lambda$ is the average number of events per unit time).

### 2.1.2 Joint distribution of the arrival times and inter-arrival times

**Definition 2.5.** The *arrival times* of a Poisson process are defined as

$$T_0 = 0, \quad T_n = \inf\{t \in \mathbb{R}_+ : N_t = n\}, \quad n \geq 1$$




The cumulative distribution function (cdf) of a given arrival time can be computed easily:

$$\mathbb{P}(T_n \leq t) = \mathbb{P}(N_t \geq n) = \sum_{k \geq n} \mathbb{P}(N_t = k) = \sum_{k \geq n} \frac{(\lambda t)^k}{k!} e^{-\lambda t}$$

and its corresponding probability density function (pdf) is given by

$$\begin{aligned} p_{T_n}(t) &= \frac{d}{dt} \mathbb{P}(T_n \leq t) = \sum_{k \geq n} \frac{\lambda^k t^{k-1}}{(k-1)!} e^{-\lambda t} - \sum_{k \geq n} \frac{(\lambda t)^k}{k!} \lambda e^{-\lambda t} \\ &= \lambda e^{-\lambda t} \left( \sum_{k \geq n} \frac{(\lambda t)^{k-1}}{(k-1)!} - \sum_{k \geq n} \frac{(\lambda t)^k}{k!} \right) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} \end{aligned}$$

i.e. $T_n \sim \text{Gamma}(n, \lambda)$ (and remember that such a Gamma random variable can be written as the sum of $n$ i.i.d. exponential random variables, each of parameter $\lambda > 0$),

We now would like to compute the *joint* distribution of the arrival times $T_1, \ldots, T_n$. For this, let us recall the following.

* If $T$ is a non-negative random variable, then for $0 \leq a < b$, $\mathbb{P}(a < T \leq b) = \int_a^b dt \, p_T(t)$, where $p_T$ is the pdf of $T$.

* Similarly, if $T_n \geq \ldots \geq T_2 \geq T_1$ are non-negative random variables, then for $0 \leq a_1 < b_1 < a_2 < b_2 < \ldots < a_n < b_n$, we have

$$\begin{aligned} &\mathbb{P}(a_1 < T_1 \leq b_1, \, a_2 < T_2 \leq b_2, \ldots, \, a_n < T_n \leq b_n) \\ &= \int_{a_1}^{b_1} dt_1 \int_{a_2}^{b_2} dt_2 \cdots \int_{a_n}^{b_n} dt_n \, p_{T_1,\ldots,T_n}(t_1, \ldots, t_n) \end{aligned}$$

where $p_{T_1,\ldots,T_n}$ is the joint pdf of $T_1, \ldots, T_n$.

Let us therefore compute

$$\mathbb{P}(a_1 < T_1 \le b_1,\, a_2 < T_2 \le b_2, \ldots, a_n < T_n \le b_n)$$

$$= \mathbb{P}(N_{a_1} = 0,\, N_{b_1} - N_{a_1} = 1,\, N_{a_2} - N_{b_1} = 0, \ldots,\, N_{a_n} - N_{b_{n-1}} = 0,\, N_{b_n} - N_{a_n} \ge 1)$$

$$= \mathbb{P}(N_{a_1} = 0)\, \mathbb{P}(N_{b_1} - N_{a_1} = 1)\, \mathbb{P}(N_{a_2} - N_{b_1} = 0) \cdots \mathbb{P}(N_{a_n} - N_{b_{n-1}} = 0)\, \mathbb{P}(N_{b_n} - N_{a_n} \ge 1)$$

$$= e^{-\lambda a_1}\, \lambda(b_1 - a_1)\, e^{\lambda(b_1 - a_1)}\, e^{-\lambda(a_2 - b_1)} \cdots e^{-\lambda(a_n - b_{n-1})}\, \big(1 - e^{-\lambda(b_n - a_n)}\big)$$

$$= \lambda^{n-1} \prod_{i=1}^{n-1} (b_i - a_i)\, (e^{-\lambda a_n} - e^{-\lambda\, b_n}) = \int_{a_1}^{b_1} dt_1 \cdots \int_{a_n}^{b_n} dt_n\, \lambda^n\, e^{-\lambda t_n}$$

So the joinf pdf of $T_1, \ldots, T_n$ is given by

$$p_{T_1,\ldots,T_n}(t_1, \ldots, t_n) = \lambda^n\, e^{-\lambda t_n}\, 1_{\{0 \le t_1 \le \ldots \le t_n\}}$$

In particular,

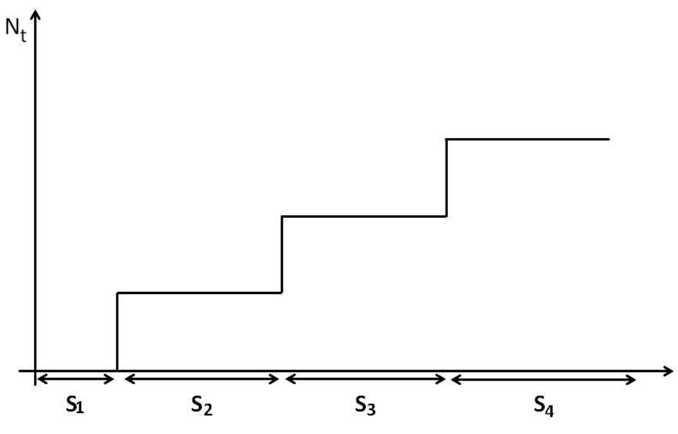
$$p_{T_1,\ldots,T_{n-1}|T_n}(t_1, \ldots, t_{n-1}|t_n) = \frac{p_{T_1,\ldots,T_n}(t_1, \ldots, t_n)}{p_{T_n}(t_n)}$$
$$= \frac{\lambda^n e^{-\lambda t_n}}{\lambda^n (t_n)^{n-1}\, e^{-\lambda t_n}/(n-1)!}\, 1_{\{0 \le t_1 \le \ldots \le t_n\}} = \frac{(n-1)!}{(t_n)^{n-1}}\, 1_{\{0 \le t_1 \le \ldots \le t_n\}}$$

i.e., given that $T_n = t_n$, the random variables $T_1, \ldots, T_{n-1}$ have the same distribution as the order statistics of $n-1$ random variables uniformly distributed on $[0, t_n]$.

**Definition 2.6.** The *inter-arrival times* of a Poisson process are defined as

$$S_n = T_n - T_{n-1}, \quad n \ge 1$$

Equivalently, $T_n = S_1 + S_2 + ... + S_n$, for $n \ge 1$.



The joint pdf of $S_1, \ldots, S_n$ can be easily computed from the joint pdf of $T_1, \ldots, T_n$:

$$p_{S_1,\ldots,S_n}(s_1, \ldots, s_n) = p_{T_1,\ldots,T_n}(s_1, s_1 + s_2, \ldots, s_1 + \ldots + s_n)$$
$$= \lambda^n e^{-\lambda(s_1+\ldots+s_n)} 1_{\{s_1 \geq 0, s_1+s_2 \geq s_1, \ldots, s_1+\ldots+s_n \geq s_1+\ldots+s_{n-1}\}} = \prod_{i=1}^n \lambda e^{-\lambda s_i} 1_{s_i \geq 0}$$

i.e. $S_1, \ldots, S_n$ are $n$ i.i.d. exponential random variables with parameter $\lambda$ (and as already observed above, $T_n$ is the sum of them). This gives rise to the following proposition (which can also be taken as an alternate definition of the Poisson process).

**Proposition 2.7.** Let $(S_n,\, n \geq 1)$ be i.i.d. exponential random variables with parameter $\lambda > 0$. Then the process defined as

$$N_t = \max\{n \geq 0 : S_1 + \ldots + S_n \leq t\}, \quad t \in \mathbb{R}_+$$

is a Poisson process of intensity $\lambda > 0$.

**Remark.**
The exponential distribution of the inter-arrival times leads to the following consequence:
* Let $t_0 \in \mathbb{R}_+$ be a fixed time, chosen independently of the process $N$. Then by stationarity,

$$\mathbb{P}(N_{t_0+t} - N_{t_0} \geq 1) = \mathbb{P}(N_t \geq 1) = 1 - \mathbb{P}(N_t = 0) = 1 - e^{-\lambda t}$$

* Let us now replace $t_0$ by an arrival time of the process $T_n$. Then again,

$\mathbb{P}(N_{T_n+t} - N_{T_n} \geq 1) = \mathbb{P}(S_{n+1} \leq t) = 1 - e^{-\lambda t}$, i.e. the probability is the same as before!

So the probability that an event takes place $t$ seconds after a given time does not depend on whether this given time is an arrival time of the process or not.

### 2.1.3 Additional properties

We prove below two useful propositions.

**Proposition 2.8.** (superposition of two independent Poisson processes)
Let $N^{(1)}$, $N^{(2)}$ be two independent Poisson processes with intensity $\lambda_1$ and $\lambda_2$, respectively. Then the process $N$ defined as

$$N_t = N_t^{(1)} + N_t^{(2)}, \quad t \in \mathbb{R}_+$$

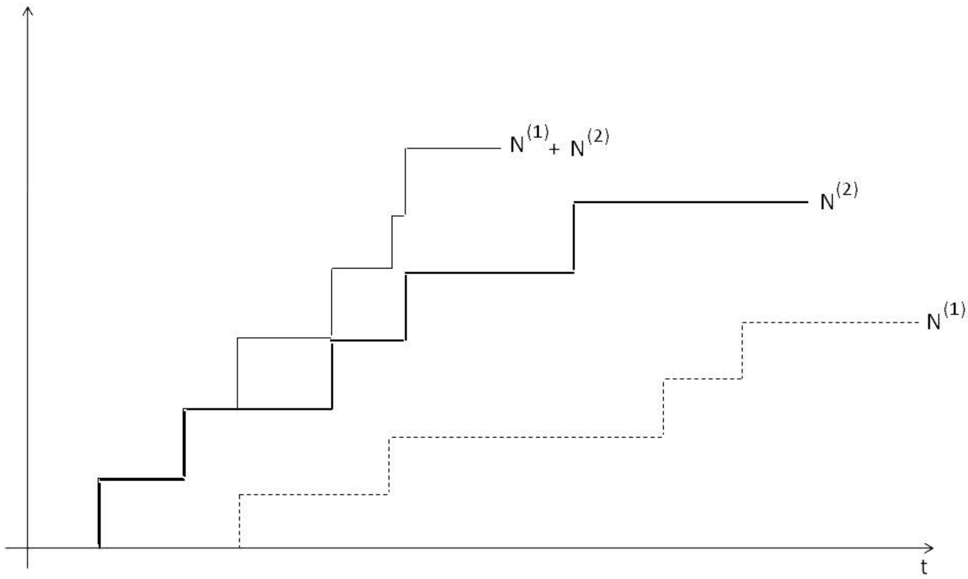
is again a Poisson process, with intensity $\lambda_1 + \lambda_2$.

*Proof.* (sketch)
We only prove here that for $t \in \mathbb{R}_+$, $N_t$ is a Poisson random variable with parameter $(\lambda_1 + \lambda_2)\, t$:

$$\mathbb{P}(N_t = n) = \mathbb{P}(N_t^{(1)} + N_t^{(2)} = n) = \sum_{k=0}^n \mathbb{P}(N_t^{(1)} = k, N_t^{(2)} = n - k)$$
$$= \sum_{k=0}^n \mathbb{P}(N_t^{(1)} = k)\, \mathbb{P}(N_t^{(2)} = n - k) = \sum_{k=0}^n \frac{(\lambda_1 t)^k}{k!} e^{-\lambda_1 t} \frac{(\lambda_2 t)^{n-k}}{(n-k)!} e^{-\lambda_2 t}$$
$$= \sum_{k=0}^n \frac{n!}{k!(n-k)!} \lambda_1^k \lambda_2^{n-k} \frac{t^n}{n!} e^{-(\lambda_1+\lambda_2)t} = \frac{((\lambda_1 + \lambda_2)t)^n}{n!} e^{-(\lambda_1+\lambda_2)t}$$

□

The superposition of two Poisson processes is illustrated on the figure below.




The next proposition is in some sense the reciprocal of the former one.

**Proposition 2.9.** (thinning of a Poisson process)
Let $N$ be a Poisson process with intensity $\lambda$ and let $(X_n,\, n \geq 1)$ be a sequence of i.i.d. random variables independent of $N$ and such that $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = 0)$, with $0 < p < 1$. Let us then associate to each arrival time $T_n$ of the original process $N$ a random variable $X_n$ and let $N^{(1)}$ be the process whose arrival times are those of the process $N$ for which $X_n = 1$. Then $N^{(1)}$ is again a Poisson process, with intensity $p\lambda$.

*Proof.* (sketch)
We only prove again that for $t \in \mathbb{R}_+$, $N_t^{(1)}$ is a Poisson random variable with parameter $p\lambda t$:

$$\begin{aligned}
\mathbb{P}(N_t^{(1)} = k) &= \sum_{n \geq k} \mathbb{P}(N_t = n, X_1 + \ldots + X_n = k) \\
&= \sum_{n \geq k} \mathbb{P}(N_t = n)\, \mathbb{P}(X_1 + \ldots + X_n = k) \qquad (15) \\
&= \sum_{n \geq k} \frac{(\lambda t)^n}{n!}\, e^{-\lambda t} \begin{pmatrix} n \\ k \end{pmatrix} p^k\, (1-p)^{n-k} \\
&= \frac{(p\lambda t)^k}{k!}\, e^{-\lambda t} \sum_{n \geq k} \frac{(\lambda t)^{n-k}}{(n-k)!}\, (1-p)^{n-k} = \frac{(p\lambda t)^k}{k!}\, e^{-p\lambda t}
\end{aligned}$$

where (15) follows from the assumed independence between the process $N$ and the random variables $(X_n,\, n \geq 1)$. □

## 2.2 Continuous-time Markov chains

### 2.2.1 Definition and basic properties

**Definition 2.10.** A *continuous-time Markov chain* is a stochastic process $(X_t,\, t \in \mathbb{R}_+)$ with values in a discrete state space $S$ such that

$$\mathbb{P}(X_{t_{n+1}} = j | X_{t_n} = i, X_{t_{n-1}} = i_{n-1}, \ldots, X_{t_0} = i_0) = \mathbb{P}(X_{t_{n+1}} = j | X_{t_n} = i),$$
$$\forall j, i, i_{n-1}, \ldots, i_0 \in S \quad \text{and} \quad \forall t_n > t_{n-1} > \ldots > t_0 \geq 0$$

If moreover

$$\mathbb{P}(X_{t+s} = j | X_s = i) = \mathbb{P}(X_t = j | X_0 = i) = p_{ij}(t), \quad \forall i, j \in S \quad \text{and} \quad \forall t, s \geq 0$$

then the chain is said to be *time-homogeneous* (we only consider such chains in the following).

Notice that we do not have anymore a single transition matrix $P = (p_{ij})_{i,j \in S}$, but a *collection* of transition matrices $P(t) = (p_{ij}(t))_{i,j \in S}$, indexed by $t \in \mathbb{R}_+$.

**Example 2.11.** The Poisson process with intensity $\lambda > 0$ is a continuous-time Markov chain, Indeed, for $j \geq i \geq i_{n-1} \geq \ldots \geq i_0 \in \mathbb{N}$, we have:

$$\begin{aligned}
&\mathbb{P}(N_{t_{n+1}} = j | N_{t_n} = i, N_{t_{n-1}} = i_{n-1}, \ldots, N_{t_0} = i_0) \\
&\quad = \mathbb{P}(N_{t_{n+1}} - N_{t_n} = j - i | N_{t_n} = i, N_{t_{n-1}} = i_{n-1}, \ldots, N_{t_0} = i_0) \\
&\quad = \mathbb{P}(N_{t_{n+1}} - N_{t_n} = j - i) = \mathbb{P}(N_{t_{n+1} - t_n} = j - i)
\end{aligned}$$

where the last line follows from the independence and the stationarity of increments. Similarly, we obtain

$$\mathbb{P}(N_{t_{n+1}} = j | N_{t_n} = i) = \mathbb{P}(N_{t_{n+1} - t_n} = j - i)$$

proving therefore the Markov property. Furthermore, the transition probabilities are given by

$$p_{ij}(t) = \mathbb{P}(N_t = j - i) = \frac{(\lambda t)^{j-i}}{(j-i)!}\, e^{-\lambda t}$$

**Notations.**
* $\pi(t) = (\pi_i(t), i \in S)$ is the *distribution of the Markov chain at time* $t \in \mathbb{R}_+$.
i.e. $\pi_i(t) = \mathbb{P}(X_t = i)$. Again, we have $\pi_i(t) \geq 0$ for all $i \in S$ and $\sum_{i \in S} \pi_i(t) = 1$, $\forall t \in \mathbb{R}_+$.
* $\pi(0) = (\pi_i(0), i \in S)$ is the *initial distribution* of the Markov chain.

One can check that $\pi_j(t) = \sum_{i \in S} \pi_i(0)\, p_{ij}(t)$ and $\pi_i(t+s) = \sum_{i \in S} \pi_i(t)\, p_{ij}(s)$.

The *Chapman-Kolmogorov equation* reads in the continuous-time case as

$$p_{ij}(t+s) = \sum_{k \in S} p_{ik}(t)\, p_{kj}(s), \quad \forall i, j \in S, \quad t, s \in \mathbb{R}_+$$

*Proof.* Along the lines of what has been shown in the discrete-time case, we obtain

$$\begin{aligned}
p_{ij}(t+s) &= \mathbb{P}(X_{t+s} = j | X_0 = i) = \sum_{k \in S} \mathbb{P}(X_{t+s} = j, X_t = k | X_0 = i) \\
&= \sum_{k \in S} \mathbb{P}(X_{t+s} = j | X_t = k, X_0 = i)\, \mathbb{P}(X_t = k | X_0 = i) = \sum_{k \in S} p_{ik}(t)\, p_{kj}(s)
\end{aligned}$$

□

### 2.2.2 Transition and sojourn times

**Disclaimer.** In this section and the following ones, rigorous proofs are often missing!

**Definition 2.12.** The *transition and sojourn times* of a continuous-time Markov chain are defined respectively as

$$T_0 = 0, \quad T_{n+1} = \inf\{t > T_n : X_t \neq X_{T_n}\}, \quad n \geq 0$$

and

$$S_n = T_n - T_{n-1}, \quad n \geq 1$$

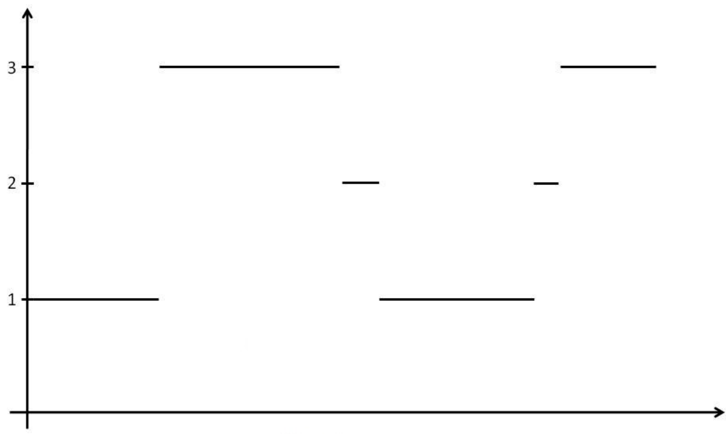
Equivalently, $T_n = S_1 + \ldots + S_n$.

The following fact is essentially a consequence of the Markov property.

**Proposition 2.13.** (without proof)
The sojourn times $(S_n,\, n \geq 1)$ are independent exponential random variables.

**Remarks.**
* In general, the sojourn times $(S_n,\, n \geq 1)$ are not identically distributed random variables.
* Also, the parameter of the exponential random variable $S_{n+1}$ depends on the state of the Markov chain at time $T_n$. That is, given that $X_{T_n} = i$, $S_{n+1}$ is an exponential random variable with some parameter $\nu_i$, so that the average waiting time in state $i$ is $\mathbb{E}(S_{n+1}|X_{T_n} = i) = 1/\nu_i$. The parameter $\nu_i$ is the rate at which the chain leaves state $i$.

Here is the graphical representation of the time evolution of a Markov chain with 3 states and $\nu_1 \sim \nu_3 < \nu_2$:




$\nu_1, \nu_3 < \nu_2$

In order to avoid strange behaviors (such as e.g. processes with infinitely many transitions during a fixed period of time), we make the following *additional assumption*:

$$\sum_{i=1}^{n} S_i = T_n \underset{n \to \infty}{\to} \infty$$

### 2.2.3 Embedded discrete-time Markov chain

Let us define $\widehat{X}_n = X_{T_n}$ and $\widehat{q}_{ij} = \mathbb{P}(X_{T_{n+1}} = j | X_{T_n} = i) = \mathbb{P}(\widehat{X}_{n+1} = j | \widehat{X}_n = i)$, $i, j \in S$.

**Fact.** (without proof)
The process $(\widehat{X}_n,\, n \geq 0)$ is a discrete-time Markov chain with transition probabilities $(\widehat{q}_{ij})_{i,j \in S}$. It is said to be *embedded* in the continuous-time Markov chain $(X_t,\, t \in \mathbb{R}_+)$.

**Remark.**
* Notice indeed that $\widehat{q}_{ij} \geq 0$, $\forall i, j \in S$ and $\sum_{j \in S} \widehat{q}_{ij} = 1$, $\forall i \in S$, as in the discrete-time case.
* Here, in addition, $\widehat{q}_{ii} = 0$, $\forall i \in S$, i.e. the embedded discrete-time Markov chain never has self-loops in its transition graph.
* The embedded chain does not "see" the time elapsed between any two transitions.

**Fact.** (again without proof)
The continuous-time Markov chain $(X_t,\, t \in \mathbb{R}_+)$ is completely characterized by the parameters $(\nu_i)_{i \in S}$ (= the rates at which the chain leaves states) and $(\widehat{q}_{ij})_{i,j \in S}$ (= the transition probabilities of the embedded discrete-time chain).

From this, we also deduce the following (by an approximate reasoning):

$$p_{ii}(\Delta t) = \mathbb{P}(X_{\Delta t} = i | X_0 = i) = \mathbb{P}(T_1 > \Delta t | X_0 = i) = e^{-\nu_i \Delta t} = 1 - \nu_i \Delta t + o(\Delta t) \quad (16)$$

$$\begin{aligned} p_{ij}(\Delta t) &= \mathbb{P}(X_{\Delta t} = j | X_0 = i) \simeq \mathbb{P}(X_{T_1} = j, T_1 \leq \Delta t | X_c = i) \simeq \widehat{q}_{ij}\,(1 - e^{-\nu_i \Delta t}) \\ &= \widehat{q}_{ij} \nu_i \Delta t + o(\Delta t) \end{aligned} \quad (17)$$

Let us therefore define a new matrix $Q$ as follows:

$$q_{ii} = -\nu_i \quad \text{and} \quad q_{ij} = \nu_i\, \widehat{q}_{ij}, \quad j \neq i$$

Then $|q_{ii}| = \nu_i$ represents the rate at which the chain leaves state $i$ and $q_{ij} = \nu_i\, \widehat{q}_{ij}$ represents the rate at which the chain transits from state $i$ to state $j$. Notice also that

$$\sum_{j \in S} q_{ij} = q_{ii} + \sum_{j \neq i} q_{ij} = -\nu_i + \nu_i \left( \sum_{j \neq i} \widehat{q}_{ij} \right) = 0, \quad \forall i \in S$$

Finally, we deduce from equations (16) and (17) that

$$P(\Delta t) = I + Q \Delta t + o(\Delta t)$$

The matrix $Q$ characterizes therefore the short-term behavior of the continuous-time Markov chain $X$. It is therefore called the *infinitesimal generator* of $X$.

### 2.2.4 Kolmogorov equations

**Proposition 2.14.** (Kolmogorov equation: version 1)

$$\frac{d\pi_j(t)}{dt} = \sum_{i \in S} \pi_i(t)\, q_{ij}, \quad \forall i, j \in S, \quad \forall t \in \mathbb{R}_+$$

or in matrix form: $\dfrac{d\pi}{dt}(t) = \pi(t)\, Q$.

*Proof.*

$$\pi_j(t+\Delta t) = \sum_{i\in S} \pi_i(t)\, p_{ij}\,(\Delta t) = \pi_j(t)\, p_{jj}(\Delta t) + \sum_{i\neq j} \pi_i(t)\, p_{ij}(\Delta t)$$
$$= \pi_j(t) + \sum_{i\in S} \pi_j(t)\, (q_{ij}\Delta t + o(\Delta t))$$

where the last equality is obtained using equations (16) and (17). Therefore,

$$\frac{\pi_j(t+\Delta t) - \pi_j(t)}{\Delta t} = \sum_{i\in S} \pi_i(t)\, q_{ij} + \frac{o(\Delta t)}{\Delta t}$$

so taking the limit $\Delta t \to 0$, we obtain (watch out that a technical detail is missing here)

$$\frac{d\pi_j}{dt}(t) = \lim_{\Delta t\to 0} \frac{\pi_j(t+\Delta t) - \pi_j(t)}{\Delta t} = \sum_{i\in S} \pi_i(t)\, q_{ij}$$

□

**Proposition 2.15.** (Kolmogorov equation: version 2, "forward" and "backward")

$$\frac{dp_{ij}}{dt}(t) = \sum_{k\in S} p_{ik}(t)\, q_{kj} = \sum_{k\in S} q_{ik}\, p_{kj}(t), \quad \forall i,j\in S, \quad \forall t\in\mathbb{R}_+$$

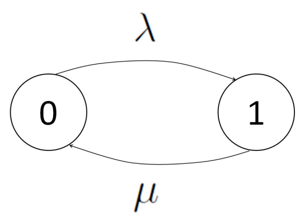
or in matrix form: $\frac{dP}{dt}(t) = P(t)\, Q = Q\, P(t)$.

We skip the proof: it follows the same lines as above. The only difference is that the Chapman-Kolomogorov equation is used here:

$$p_{ij}(t+\Delta t) = \sum_{k\in S} p_{ik}(t)\, p_{kj}(\Delta t) = \sum_{k\in S} p_{ik}(\Delta t)\, p_{kj}(t)$$

**Example 2.16.** (two-state continuous-time Markov chain)
Let us consider the continuous-time Markov chain with state space $S = \{0,1\}$ and infinitesimal generator

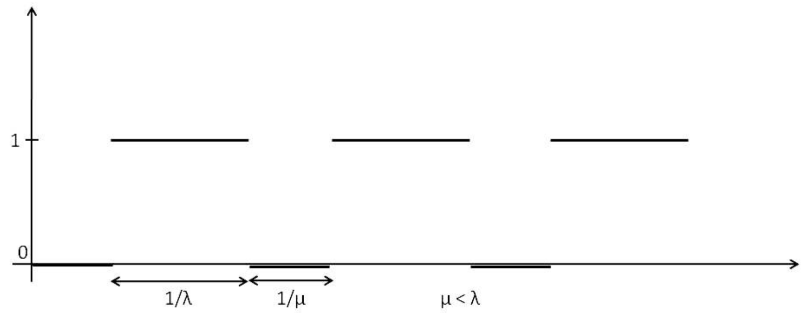
$$Q = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$



The embedded discrete-time Markov chain has the following transition matrix:

$$\widehat{Q} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

and the average waiting times are $\frac{1}{\lambda}$ in state $0$ and $\frac{1}{\mu}$ in state $1$.




The Kolmogorov equation (version 1) reads in this case

$$\left(\frac{d\pi_0}{dt}, \frac{d\pi_1}{dt}\right) = (\pi_0, \pi_1) \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$$

Solving this ordinary differential equation, we obtain

$$\pi_0(t) = \frac{\mu}{\lambda + \mu} + \left(\pi_0(0) - \frac{\mu}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}$$

and

$$\pi_1(t) = \frac{\lambda}{\lambda + \mu} + \left(\pi_1(0) - \frac{\lambda}{\lambda + \mu}\right) e^{-(\lambda+\mu)t}$$

### 2.2.5 Classification of states

As in the discrete-time case, let us introduce some definitions.

* Two states $i$ and $j$ *communicate* if $p_{ij}(t) > 0$ and $p_{ji}(t) > 0$ for some $t \geq 0$.

* The chain is said to be *irreducible* if all states communicate.

* Fact: if $p_{ij}(t) > 0$ for *some* $t > 0$, then $p_{ij}(t) > 0$ for *all* $t > 0$, so there is no notion of periodicity here.

* Let $R_i$ be the *first return time* to state $i$: $R_i = \inf\{t > T_1 : X_t = i\}$.

* A state $i$ is said to be *recurrent* if $f_i = \mathbb{P}(R_i < \infty | X_0 = i) = 1$ and *transient* otherwise.

* Moreover, if a state $i$ is recurrent, then it is *positive recurrent* if $\mathbb{E}(R_i | X_0 = i) < \infty$ and *null recurrent* otherwise.

**Remarks.**
* The continuous-time Markov chain $X$ and its embedded discrete-time Markov chain $\widehat{X}$ share all the above properties (except for periodicity).
* $X$ is said to be *ergodic* if it is irreducible and positive recurrent (and as before, in a finite-state chain, irreducible implies positive recurrent).

### 2.2.6 Stationary and limiting distributions

The following theorem is the equivalent of Corollary 1.13 for discrete-time Markov chains.

**Theorem 2.17.** Let $X$ be an ergodic continuous-time Markov chain. Then it admits a unique stationary distribution, i.e. a distribution $\pi^*$ such that

$$\pi^* P(t) = \pi^*, \quad \forall t \in \mathbb{R}_+ \tag{18}$$

Moreover, this distribution is a limiting distribution, i.e. for any initial distribution $\pi(0)$, we have

$$\lim_{t\to\infty} \pi(t) = \pi^*$$

**Remark.**
Equation (18) is not so easy to solve in general, but it can be shown to be equivalent (modulo a technical assumption) to the much nicer equation

$$\pi^* Q = 0, \quad \text{i.e.} \quad \sum_{i\in S} \pi_i^*\, q_{ij} = 0, \quad \forall j \in S \tag{19}$$

Here is the main proof idea in one direction: if $\pi^*$ satisfies (18), then $\pi^*(P(t) - P(0)) = 0$, $\forall t \in \mathbb{R}_+$. So

$$\lim_{t\to 0} \pi^* \left( \frac{P(t) - P(0)}{t} \right) = 0$$

which in turn implies (and here comes the technical detail that we skip) that $\pi^* \dfrac{dP}{dt}(0) = 0$, i.e. $\pi^* Q = 0$.

**Example 2.18.** (two-state continuous-time Markov chain)
Turning back to the two-state continuous-time Markov chain of Example 2.16 (which is ergodic), we need to solve

$$(\pi_0^*, \pi_1^*) \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix} = 0$$

which leads to $(\pi_0^*, \pi_1^*) = \left(\frac{\mu}{\lambda+\mu}, \frac{\lambda}{\lambda+\mu}\right)$. Notice that this result could also have been obtained by taking the limit $t \to \infty$ in the expression obtained for $(\pi_0(t), \pi_1(t))$.

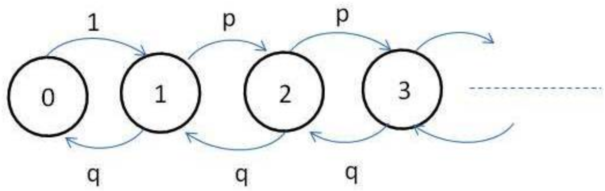
**Example 2.19.** (birth and death process)
Let $(X_t,\, t \in \mathbb{R}_+)$ be continuous-time Markov chain with state space $S = \mathbb{N}$ and infinitesimal generator

$$q_{0j} = \begin{cases} -\lambda & \text{if } j = 0 \\ \lambda & \text{if } j = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad q_{ij} = \begin{cases} \mu & \text{if } j = i-1 \\ -\lambda-\mu & \text{if } j = i \\ \lambda & \text{if } j = i+1 \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i \geq 1$$

So $\nu_0 = \lambda$, and $\nu_i = \lambda + \mu$ for $i \geq 1$. Moreover, the embedded discrete-time Markov chain has the following transition probabilities:

$$\widehat{q}_{01} = \frac{q_{01}}{\nu_0} = 1, \quad \widehat{q}_{i,i+1} = \frac{q_{i,i+1}}{\nu_i} = \frac{\lambda}{\lambda+\mu}, \quad \widehat{q}_{i,i-1} = \frac{q_{i,i-1}}{\nu_i} = \frac{\mu}{\lambda+\mu}$$

corresponding to the transition graph




where $p = \frac{\lambda}{\lambda+\mu}$ and $q = 1 - p = \frac{\mu}{\lambda+\mu}$, i.e. this chain is a random walk on $\mathbb{N}$.

Let us now look for the stationary distribution of the continuous-time Markov chain, if it exists.

* If $\lambda, \mu > 0$, then the chain is irreducible.
* If $\lambda \geq \mu$, i.e. $p \geq q$, then the chain is either transient or null recurrent. so there does not exist a stationary distribution.
* If on the contrary $\lambda < \mu$, i.e. $p < q$, then the chain is positive recurrent, and solving the equation $\pi^* Q = 0$ in this case leads to

$$\begin{aligned} \pi_0\, q_{00} + \pi_1\, q_{10} = 0, &\quad \pi_{i-1}\, q_{i,i-j} + \pi_i\, q_{ii} + \pi_{i+1}\, q_{i+1,i} = 0 \\ -\lambda\, \pi_0 + \mu\, \pi_1 = 0, &\quad \lambda\, \pi_{i-1} - (\lambda+\mu)\, \pi_i + \mu\, \pi_{j+1} = 0 \end{aligned}$$

So

$$\pi_1 = \frac{\lambda}{\mu}\, \pi_0, \quad \pi_2 = \frac{1}{\mu}\left((\lambda+\mu)\pi_1 - \lambda\pi_0\right) = \left(\frac{\lambda}{\mu}\right)^2 \pi_0$$

and by induction,

$$\pi_k = \left(\frac{\lambda}{\mu}\right)^k \pi_0, \quad \pi_0 \left(1 + \frac{\lambda}{\mu} + \frac{\lambda^2}{\mu^2} + \ldots\right) = 1$$

i.e., finally,

$$\pi_k^* = \left(\frac{\lambda}{\mu}\right)^k \left(1 - \frac{\lambda}{\mu}\right), \quad k \in \mathbb{N}$$

This concludes these short lecture notes on Markov chains.