Bachelor in Mathematics

Applied Stochastic Processes

Spring Semester, 2018

Sebastian Engelke and Sophie Hautphenne,
with edits by Tobias Hurth

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Disclaimer: This document is a draft version of the lecture notes that is not intended for distribution. Please report mistakes to tobias.hurth@epfl.ch or sophie.hautphenne@epfl.ch.

Chapter 1

Review of Probability Theory

Remark 1.0.1 *In these notes, we adopt the convention* $\mathbb{N} = \{0, 1, 2, \ldots\}$*, i.e.* 0 *is included in the set* $\mathbb{N}$*.*

1.1 Some notions in probability theory

Probability theory is explicitly based on mathematical analysis and measure theory. Measure theory is general and powerful (measurable functions, integration, ...) and underpins probability theory.
Let Ω be a nonempty set, the so-called *sample space*, let $\mathcal{A}$ be a σ-algebra on Ω, the so-called *event set*, and let $\mathbb{P} : \mathcal{A} \to [0, 1]$ be a *probability measure*. We call the tuple $(\Omega, \mathcal{A})$ a *measurable space*, and we call the triple $(\Omega, \mathcal{A}, \mathbb{P})$ a *probability space*. Elements of the event set $\mathcal{A}$ are called *events*. Recall the following definitions.

Definition 1.1.1 *A* σ*-algebra on a set* Ω *is a collection of subsets of* Ω *such that the following holds.*

1. $\Omega \in \mathcal{A}$
2. *If* $A \in \mathcal{A}$*, then we also have* $A^c \in \mathcal{A}$*. Here,* $A^c = \Omega \setminus A$ *denotes the complement of* A*. Another common notation for the complement of* A *is* $\bar{A}$*.*
3. *If* $(A_i)_{i \in \mathbb{N}}$ *is a countable collection of sets such that* $A_i \in \mathcal{A}$ *for all* $i \in \mathbb{N}$*, then we also have* $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$*.*

1. and 2. imply that any σ-algebra $\mathcal{A}$ also satisfies $\emptyset \in \mathcal{A}$. 3. implies that if $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$. By de Morgan's law, 2. and 3. imply that if $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$. Some people use the term *σ-field* instead of σ-algebra.

Definition 1.1.2 *A probability measure* $\mathbb{P}$ *on a measurable space* $(\Omega, \mathcal{A})$ *is a measure on* $(\Omega, \mathcal{A})$ *such that* $\mathbb{P}(\Omega) = 1$*. In order for* $\mathbb{P}$ *to be a measure on* $(\Omega, \mathcal{A})$*, we need*

1. $\mathbb{P}(A) \geq 0$ *for all* $A \in \mathcal{A}$.

2. If $(A_i)_{i \in \mathbb{N}}$ *is a countable collection of pairwise disjoint sets in* $\mathcal{A}$*, then*

$$\mathbb{P}(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i).$$

Property 2. is called *σ-additivity.*
To be able to do interesting things in probability, we next define the notion of "random variable", X.

Random variables

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $\mathcal{E}$ be a nonempty set and let ξ be a σ-algebra on $\mathcal{E}$. A *random variable* is a measurable function

$$X : \Omega \to \mathcal{E}.$$

Recall that X is measurable if for all $B \in \xi$, $X^{-1}(B) \in \mathcal{A}$, where $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$. The *distribution* of X is defined as the push-forward measure of $\mathbb{P}$ under the mapping X. This is a probability measure on $(\mathcal{E}, \xi)$ and we denote it by $\mathbb{P}_X$. That is we have

$$\mathbb{P}_X(B) = \mathbb{P}(X^{-1}(B)) = \mathbb{P}(X \in B), \quad B \in \xi.$$

Example 1.1.3 *Suppose that* $\mathcal{E} = \mathbb{R}$ *and that* ξ *is the Borel* σ*-algebra on* $\mathbb{R}$*, i.e. the smallest* σ*-algebra on* $\mathbb{R}$ *that contains all open subsets of* $\mathbb{R}$*. Let* $X : \Omega \to \mathbb{R}$ *be a random variable. Then, for any two real numbers* $a \leq b$*, the probability that* $X \in [a, b]$ *is* $\mathbb{P}[X^{-1}([a, b])] = \mathbb{P}_X([a, b])$.

For the remainder of this chapter, unless otherwise specified, random variables will always map to $\mathbb{R}$ equipped with the Borel σ-algebra. The *cumulative distribution function* (or c.d.f. in short) of a random variable X is defined as

$$F(x) = \mathbb{P}(X \leq x) = \mathbb{P}_X((-\infty, x]), \quad x \in \mathbb{R}.$$

We always have $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$. Moreover, if X and Y are two random variables with the same distribution (if X and Y are identically distributed), then they also have the same c.d.f.

1.2 Conditional probability

Definition 1.2.1 *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and let* A *and* B *be two events. Assume further that* $\mathbb{P}(B) > 0$*. Then, the conditional probability of* A *given* B *is defined as*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

If $\mathbb{P}(B) > 0$, the concept of conditional probability lets us define a new probability measure $\mathbb{P}_B$:

$$\mathbb{P}_B[\bullet] = \mathbb{P}[\bullet \mid B] = \frac{\mathbb{P}(\bullet \cap B)}{\mathbb{P}(B)}$$

Dividing the term $\mathbb{P}(\bullet \cap B)$ by $\mathbb{P}(B)$ can be interpreted as a "normalization", i.e. it ensures that $\mathbb{P}_B(\Omega) = 1$ and that $\mathbb{P}_B$ is thus truly a probability measure.
Conditional probabilities allow many problems to be simplified. A particularly helpful tool is the law of total probability that we state below.

Theorem 1.2.2 (Law of total probability) *Let $(B_i)_{i \in I}$ be a finite or countably infinite partition of Ω such that $B_i \in \mathcal{A}$ for all $i \in I$. This means that the sets $(B_i)_{i \in I}$ are pairwise disjoint and that $\bigcup_{i \in I} B_i = \Omega$. Then, for any event $A \in \mathcal{A}$, we have*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

Here, one should interpret $\mathbb{P}(A \mid B_i)\mathbb{P}(B_i)$ as 0 if $\mathbb{P}(B_i) = 0$.

The law of total probability follows immediately from σ-additivity of $\mathbb{P}$ and the definition of conditional probability.

1.3 Independence

Independence of events

Two events A and B are called *independent* if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. If $\mathbb{P}(B) > 0$, this is equivalently to

$$\mathbb{P}(A \mid B) = \mathbb{P}(A),$$

which has the following interpretation: If $\mathbb{P}(B) > 0$, then A and B are independent if and only if knowing that B occurred has no effect on the likelihood of A occurring. We use the shorthand $A \perp\!\!\!\perp B$ to indicate that A and B are independent events. A finite collection of events $A_1, \ldots, A_n$ is called mutually independent (or just independent) if

$$\mathbb{P}(A_1 \cap \ldots \cap A_n) = \mathbb{P}(A_1) \cdot \ldots \cdot \mathbb{P}(A_n).$$

Finally, an infinite collection of events is called (mutually) independent if every finite subcollection is independent.

Independence of random variables

Let $X_1, \ldots, X_n$ be finitely many random variables. Then $X_1, \ldots, X_n$ are called mutually independent (or just independent) if for any $a_1, \ldots, a_n \in \mathbb{R}$, we have

$$\mathbb{P}(X_1 \leq a_1, \ldots, X_n \leq a_n) = \mathbb{P}(X_1 \leq a_1) \cdot \ldots \cdot \mathbb{P}(X_n \leq a_n).$$

If the random variables $X_1, \dots, X_n$ are independent, then for every $i \in \{1, \dots, n\}$ and for every $a_1, \dots, a_n \in \mathbb{R}$ such that $\mathbb{P}(X_1 \le a_1, \dots, X_{i-1} \le a_{i-1}, X_{i+1} \le a_{i+1}, \dots, X_n \le a_n) > 0$, we have

$$\mathbb{P}(X_i \le a_i \mid X_0 \le a_0, \dots, X_{i-1} \le a_{i-1}, X_{i+1} \le a_{i+1}, \dots, X_n \le a_n) = \mathbb{P}(X_i \le a_i).$$

An infinite collection of random variables is called (mutually) independent if every finite subcollection is independent.

Conditional independence

Let A, B, C be three events such that $\mathbb{P}(C) > 0$. We say that A and B are *conditionally independent* given C if

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C).$$

If $\mathbb{P}(B \cap C) > 0$, this is equivalent to

$$\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C).$$

Using the definition of conditional probability, we can rewrite $\mathbb{P}(A \mid B \cap C) = \mathbb{P}(A \mid C)$ as

$$\frac{\mathbb{P}(A \cap B \cap C)}{\mathbb{P}(B \cap C)} = \frac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)}.$$

Thus,

$$\mathbb{P}(A \cap B \cap C) = \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(C)}\mathbb{P}(A \cap C) = \mathbb{P}(B \mid C)\mathbb{P}(A \cap C).$$

Dividing both sides by $\mathbb{P}(C)$ yields

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(B \mid C)\mathbb{P}(A \mid C).$$

Here is another way to state conditional independence: Two events A and B are conditionally independent given C if they are independent with respect to the probability measure $\mathbb{P}_C$ introduced in Section 1.2. In words, conditional independence given C means that A and B are independent if the event C occurs.

Two random variables X and Y are called *conditionally independent* given Z if for each $a, b \in \mathbb{R}$ and for every $c \in \mathbb{R}$ such that $\mathbb{P}(Z \le c) > 0$ we have

$$\mathbb{P}(X \le a, Y \le b \mid Z \le c) = \mathbb{P}(X \le a \mid Z \le c)\mathbb{P}(Y \le b \mid Z \le c).$$

Further Reading

You can find a more extensive review of core concepts from probability theory in Chapter 6 of Norris's textbook.

Chapter 2

Markov chains

2.1 Stochastic processes

Definition 2.1.1 *Let* $(\Omega, \mathcal{A}, \mathbb{P})$ *be a probability space and let* T *be an index set (such as* $\mathbb{N}$ *or* $\mathbb{R}$*, often representing time), and let* $(\mathcal{E}, \xi)$ *be a measurable space. For every* $t \in T$*, let* $X_t : \Omega \to \mathcal{E}$ *be a random variable. The collection of random variables* $(X_t)_{t\in T}$ *(or* $\{X_t : t \in T\}$*) is then called a* stochastic process.

Instead of X_t we sometimes write $X(t)$, especially if T is an interval of the real line. Typical examples for the set T are

- $T = \mathbb{N}$; then, $(X_t)_{t\in T}$ is a random sequences with values in $\mathcal{E}$;
- $T = [0, \infty)$ or $T = \mathbb{R}$; a stochastic process with such an index set often represents the evolution of a system in time: at each moment t, the system is represented by the random variable X_t.

A stochastic process represents the evolution (generally in time) of a random variable (a "system"). If $\mathcal{E} = \mathbb{R}$ and if ξ is the Borel σ-algebra on $\mathbb{R}$, then the distribution of the stochastic process $(X_t)_{t\in T}$ is a probability measure on $(\mathbb{R}^T, \mathcal{B}(\mathbb{R}^T))$, where $\mathcal{B}(\mathbb{R}^T)$ denotes the Borel σ-algebra on $\mathbb{R}^T$. The distribution of $(X_t)_{t\in T}$ specifies all finite-dimensional distributions, i.e. the distributions of all random vectors $(X_{t_1}, \dots, X_{t_m})$ for any $m \in \mathbb{N}$ and any $t_1, \dots, t_m \in T$. Thus, it captures the dependence between the single random variables.

NOTE: In order to specify the distribution of a stochastic process $(X_t)_{t\in T}$, it is not enough to just specify the distribution of each individual random variable X_t as this doesn't take into account the dependence structure.

Example 2.1.2 *Consider the stochastic process* $\{X_n : n \in \mathbb{N}\}$*, where the random variables* X_n *are i.i.d. (independent and identically distributed), with c.d.f.* F*. For all* $n \in \mathbb{N}$ *and for all* $a_0, \dots, a_n \in \mathbb{R}$*, we have*

$$\mathbb{P}(X_0 \le a_0, \dots, X_n \le a_n) = \mathbb{P}(X_0 \le a_0) \cdot \ldots \cdot \mathbb{P}(X_n \le a_n) = F(a_0) \cdot \ldots \cdot F(a_n).$$

The hypothesis that the random variables X_n are i.i.d. is usually too simplistic. We will now define some processes which treat more general settings.

2.2 Markov chains

Definition 2.2.1 *Let $X = (X_n)_{n \in \mathbb{N}}$ be a stochastic process with values in a countable set E, i.e. X_n maps to E for every $n \in \mathbb{N}$. If E is finite, we may assume that $E = \{1, \dots, N\}$ for some positive integer N. If E is countably infinite, we assume $E = \mathbb{N} \setminus \{0\}$. We call X a* Markov chain *if for each $n \in \mathbb{N}$, X_{n+1} is conditionally independent of $X_0, X_1, \dots, X_{n-1}$, given X_n. This is called the* Markov property*. The Markov property implies in particular that for all $n \in \mathbb{N}$, for all $j \in E$, and for all $i_0, i_1, \dots, i_n \in E$ such that $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$,*

$$\mathbb{P}(X_{n+1} = j \mid X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = j \mid X_n = i_n).$$

Markov chains are a first step to relax the assumption of independence, which is too simplistic to represent reality. In short, the Markov property means that predicting the state of a system in the future, with the present known, is not made more precise through knowledge of supplementary information about the past, as this information becomes unnecessary.
In this lecture we will only consider time-homogeneous Markov chains.

Definition 2.2.2 *A Markov chain X is called (time-)*homogeneous *if for all $n, m \in \mathbb{N}$ and for all $i, j, \in E$ such that $\mathbb{P}(X_n = i), \mathbb{P}(X_m = i) > 0$ we have*

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_{m+1} = j \mid X_m = i) =: p_{ij}.$$

The number p_{ij} is undefined if $\mathbb{P}(X_n = i) = 0$ for all $n \in \mathbb{N}$. If p_{ij} is defined for all $i, j \in E$, we call the matrix $P = (p_{ij})$ of size $|E| \times |E|$ (possibly infinite) the transition matrix*. The* initial distribution $\boldsymbol{\alpha} = (\alpha_i)_{i \in E}$ *of X is defined by*

$$\alpha_i = \mathbb{P}(X_0 = i), \quad i \in E.$$

If X is a homogeneous Markov chain with initial distribution $\boldsymbol{\alpha}$ and transition matrix P, we say that X is $Markov(\boldsymbol{\alpha}, P)$ and write $X \sim Markov(\boldsymbol{\alpha}, P)$.

In Theorem 2.2.4, we shall see that the distribution of a homogeneous Markov chain is completely determined by the initial distribution $\boldsymbol{\alpha}$ and the transition matrix P. In an abuse of terminology, we will often identify a Markov chain X with its distribution, i.e. we will forget about the underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and only focus on the finite-dimensional distributions of X.

Remark 2.2.3 *p_{ij} is the probability to go from i to j in one time step. Clearly*

(i) Each entry of P is nonnegative because each p_{ij} is a probability;

(ii) If $\mathbf{1} = (1, 1, \dots, 1)^\top$ *for finite* E *and* $\mathbf{1} = (1, 1, \dots)^\top$ *for countably infinite* E, *we have* $P\mathbf{1} = \mathbf{1}$. *This is because for every* $i \in E$,

$$\begin{aligned}\sum_{j \in E} p_{ij} &= \sum_{j \in E} \mathbb{P}(X_{n+1} = j \mid X_n = i) \\ &= \mathbb{P}(X_{n+1} \in \bigcup_{j \in E} \{j\} \mid X_n = i) \\ &= \mathbb{P}(X_{n+1} \in E \mid X_n = i) \\ &= 1.\end{aligned}$$

A stochastic matrix $P = (p_{ij})$ *is a matrix which satisfies both* $p_{ij} \geq 0$ *for every* $i, j \in E$ *and* $P\mathbf{1} = \mathbf{1}$. *In particular, if* P *is the transition matrix of a Markov chain, then* P *is also a stochastic matrix.*

Theorem 2.2.4 *Let* X *be Markov*$(\boldsymbol{\alpha}, P)$. *Then, for all* $n \in \mathbb{N}$ *and for all* $i_0, \dots, i_n \in E$ *we have*

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}.$$

Proof. We show the statement by induction. In the base case $n = 0$, we have

$$\mathbb{P}(X_0 = i_0) = \alpha_{i_0}.$$

In the induction step, suppose there is $n \in \mathbb{N}$ for which the statement holds. Let $i_0, \dots, i_{n+1} \in E$. By induction hypothesis,

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}. \tag{2.1}$$

If $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = 0$, this implies $\alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} = 0$ and hence

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1}) = 0 = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n} p_{i_n i_{n+1}}.$$

If $\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) > 0$, we may condition on this event and obtain

$$\begin{aligned}&\mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1}) \\ &\quad = \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_0, \dots, X_n = i_n).\end{aligned}$$

The Markov property and homogeneity imply

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

Together with (2.1), this yields the desired result. □

So far, we have defined Markov chains, but we do not know yet whether they exist, i.e. whether there are stochastic processes that are homogeneous and satisfy the Markov property. Theorem 2.2.4 can be used to construct Markov processes explicitly and to thus show their existence: Fix any probability measure $\boldsymbol{\alpha}$ on E, and fix any stochastic matrix

P of dimension $|E| \times |E|$. One can show that there is a stochastic process $X = (X_n)_{n \in \mathbb{N}}$ on E with finite-dimensional distributions

$$\mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = \alpha_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

for $n \in \mathbb{N}$ and $i_0, \dots, i_n \in E$. Once this is established, it is not hard to see that X is homogeneous and satisfies the Markov property.

Let $X \sim Markov(\boldsymbol{\alpha}, P)$ and let $i \in E$ such that $\alpha_i > 0$. Then, we define the probability measure

$$\mathbb{P}_i(\bullet) = \mathbb{P}(\bullet \mid X_0 = i).$$

For $i, j \in E$ and $n \in \mathbb{N}$, we denote the entry in the ith row and jth column of the matrix $P^n = P \dots P$ (n times) by

$$p_{ij}^{(n)}.$$

Note that $p_{ij}^{(1)} = p_{ij}$.

Proposition 2.2.5 *For each $n \in \mathbb{N}$ and for each $j \in E$,*

$$\mathbb{P}_i(X_n = j) = p_{ij}^{(n)}.$$

Proof. We prove the statement by induction. The case $n = 0$ follows from the convention that P^0 is the identity matrix with 1s on the diagonal and 0s off the diagonal. In the induction step, suppose the formula holds for some $n \in \mathbb{N}$. We need to show that $\mathbb{P}_i(X_{n+1} = j) = p_{ij}^{(n+1)}$. The law of total probability applied to $\mathbb{P}_i$ yields

$$\mathbb{P}_i(X_{n+1} = j) = \sum_{\substack{h \in E: \\ \mathbb{P}_i(X_n = h) > 0}} \mathbb{P}_i(X_n = h) \mathbb{P}_i(X_{n+1} = j \mid X_n = h).$$

The Markov property and the fact that X is homogeneous let us write

$$\mathbb{P}_i(X_{n+1} = j \mid X_n = h) = \mathbb{P}(X_{n+1} = j \mid X_n = h) = p_{hj}.$$

And by induction hypothesis,

$$\mathbb{P}_i(X_n = h) = p_{ih}^{(n)}.$$

As a result,

$$\mathbb{P}_i(X_{n+1} = j) = \sum_{\substack{h \in E: \\ \mathbb{P}_i(X_n = h) > 0}} p_{ih}^{(n)} p_{hj}.$$

For any $h \in E$ such that $\mathbb{P}_i(X_n = h) = 0$, we also have $p_{ih}^{(n)} = 0$ by induction hypothesis. Thus, we may sum over the entire set E and obtain

$$\mathbb{P}_i(X_{n+1} = j) = \sum_{h \in E} p_{ih}^{(n)} p_{hj} = p_{ij}^{(n+1)},$$

as desired. □

In light of Proposition 2.2.5, we can interpret $p_{ij}^{(n)}$ as the probability of X going from state i to state j in n steps.

Remark 2.2.6 *Since X is homogeneous, we have the following slight generalization of Proposition 2.2.5:*
For any $i, j \in E$ and any $m, n \in \mathbb{N}$ with $\mathbb{P}(X_m = i) > 0$, we have

$$\mathbb{P}(X_{m+n} = j \mid X_m = i) = p_{ij}^{(n)}.$$

Corollary 2.2.7 *a) Let $X \sim Markov(\boldsymbol{\alpha}, P)$. For any $n \in \mathbb{N}$ and any $i \in E$, we have*

$$\mathbb{P}(X_n = i) = (\boldsymbol{\alpha} P^n)_i,$$

i.e. $\mathbb{P}(X_n = i)$ is the ith component of the row vector $\boldsymbol{\alpha} P^n$.

b) [Chapman–Kolmogorov equation] *For all m, $n \in \mathbb{N}$ and for all i, $j \in E$ with $\alpha_i > 0$, we have*

$$\mathbb{P}_i(X_{m+n} = j) = \sum_{h \in E} \mathbb{P}_i(X_m = h) \mathbb{P}(X_{m+n} = j \mid X_m = h).$$

Here we interpret $\mathbb{P}_i(X_m = h)\mathbb{P}(X_{m+n} = j \mid X_m = h)$ as 0 if $\mathbb{P}(X_m = h) = 0$.

Draw a picture to illustrate the Chapman–Kolmogorov equation.
Proof.

a) By the law of total probability,

$$\mathbb{P}(X_n = i) = \sum_{\substack{h \in E, \\ \alpha_h > 0}} \mathbb{P}(X_0 = h) \mathbb{P}_h(X_n = i).$$

By definition $\mathbb{P}(X_0 = h) = \alpha_h$, and Proposition 2.2.5 yields $\mathbb{P}_h(X_n = i) = p_{hi}^{(n)}$ for every $h \in E$ such that $\alpha_h > 0$. As $p_{hi}^{(n)}$ is also defined if $\alpha_h = 0$, we obtain

$$\mathbb{P}(X_n = i) = \sum_{h \in E} \alpha_h p_{hi}^{(n)} = (\boldsymbol{\alpha} P)_i.$$

b) The definition of the matrix product yields

$$p_{ij}^{(m+n)} = \sum_{h \in E} p_{ih}^{(m)} p_{hj}^{(n)}.$$

Then, we just apply Proposition 2.2.5 to the terms on both sides:

$$p_{ij}^{(m+n)} = \mathbb{P}_i(X_{m+n} = j), \quad p_{ih}^{(m)} = \mathbb{P}_i(X_m = h)$$

provided that $\alpha_i > 0$, and

$$p_{hj}^{(n)} = \mathbb{P}(X_{m+n} = j \mid X_m = h)$$

if $\mathbb{P}(X_m = h) > 0$.

□

Remark 2.2.8 *a) To any Markov chain we can associate a* transition graph *between states: this is a weighted, directed graph with an edge from i to j if and only if $p_{ij} > 0$. The weight assigned to the edge from i to j is precisely p_{ij}.*

b) We say that state i leads *to state j (and write $i \to j$) if*

$$p_{ij}^{(n)} > 0, \quad \text{for some } n \in \mathbb{N}.$$

In this case, there is a path from i to j in the transition graph. Notice however that if $j = i$, we always have $i \to j$ and $j \to i$, even if there is no loop at the vertex i. This is because $p_{ii}^{(0)} = 1 > 0$.

c) We say that i communicates *with j (and write $i \leftrightarrow j$) if both $i \to j$ and $j \to i$.*

d) It is easy to check that the relation $\leftrightarrow$ is an equivalence relation on the state space E, i.e. we have $i \leftrightarrow i$; $i \leftrightarrow j$ if and only if $j \leftrightarrow i$; and $i \leftrightarrow j$ and $j \leftrightarrow k$ together imply $i \leftrightarrow k$. Therefore, the relation $\leftrightarrow$ partitions E into so-called equivalence classes. These are subsets of E of the form

$$\{i \in E : i \leftrightarrow j\}$$

for $j \in E$. We call these equivalence classes communicating classes*.*

e) We say that a communicating class C is closed *if*

$$i \in C,\ i \to j \quad \text{implies} \quad j \in C.$$

The state i is called absorbing *if $\{i\}$ is a closed communicating class. If i is absorbing, then*

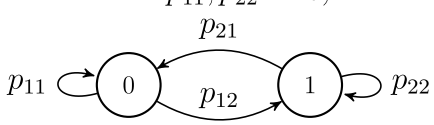
$$p_{ij} = \begin{cases} 0, & j \neq i, \\ 1, & j = i. \end{cases}$$

f) We call a Markov chain with transition matrix P (and the matrix P itself) irreducible *if the entire set E is a communicating class, i.e. if $i \leftrightarrow j$ for all $i, j \in E$.*

Example 2.2.9 *a) Consider a Markov chain on the state space $E = \{1, 2\}$ with transition matrix*

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}.$$

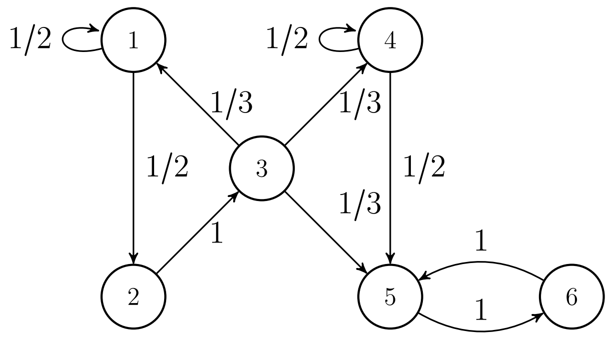
Suppose $p_{12}, p_{21} > 0$. Then $1 \leftrightarrow 2$, so the Markov chain is irreducible. If we have in addition $p_{11}, p_{22} > 0$, then the transition graph of the chain looks as follows.



b) Consider the transition matrix

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The corresponding transition graph is:



The communicating classes are $\{1,2,3\}$, $\{4\}$, *and* $\{5,6\}$. *Only* $\{5,6\}$ *is a closed communicating class.*

c) (One-dimensional random walk) A random walk on the set of integers $\mathbb{Z}$ *is a Markov chain on the state space* $E = \mathbb{Z}$ *of the form* $X_n = X_0 + \sum_{i=1}^n \epsilon_i$, $n \in \mathbb{N}$. *Here* X_0 *is an integer-valued random variable and* $(\epsilon_i)_{i\geq 1}$ *are integer-valued and i.i.d. If the distribution of the* ϵ_i*'s is given by* $\mathbb{P}(\epsilon_i = 1) = p$, $\mathbb{P}(\epsilon_i = -1) = 1-p$ *for some* $p \in (0,1)$, *we call* X *a simple random walk. A simple random walk starts at some randomly chosen integer given by* X_0. *Then, in each successive step, it jumps to its nearest neighbor on the right with probability* p *and to its nearest neighbor on the left with probability* $1-p$. *The* direction *of each jump (i.e. left or right) is independent of previous jumps. The* site *to which a jump leads does however depend on previous jumps as the walker can only jump to a nearest neighbor of its current location. The corresponding transition matrix is the infinite matrix*

$$P = \begin{pmatrix} \ddots & & & \vdots & & & \\ & 1-p & 0 & p & 0 & 0 & \\ \cdots & 0 & 1-p & 0 & p & 0 & \cdots \\ & 0 & 0 & 1-p & 0 & p & \\ & & & \vdots & & & \ddots \end{pmatrix}.$$

d) (Birth and death chain on $\mathbb{N}$*) For* $i \in \mathbb{N}$ *let* p_i, r_i, q_i *be real numbers in* $[0,1]$ *such that* $p_i + r_i + q_i = 1$. *Assume further that* $q_0 = 0$. *A birth and death chain on* $\mathbb{N}$ *is a*

Markov chain on the state space $E = \mathbb{N}$ with transition matrix $P = (p_{ij})$ satisfying

$$p_{ij} = \begin{cases} p_i, & j = i+1, \\ r_i, & j = i, \\ q_i, & j = i-1. \end{cases} \tag{2.2}$$

Here, X_n can be interpreted as the size of a population at time n. From one time step to the next, there is either exactly one birth or exactly one death or the population size stays constant. The probabilities of birth are given by the p_i's, the probabilities of death by the q_i's and the probabilities of nothing happening by the r_i's. Why do we need to impose $q_0 = 0$?

2.2.1 Strong Markov property

For a homogeneous Markov chain, we have the Markov property at each fixed time n: the random variable X_{n+1} is conditionally independent of $X_0, \ldots, X_{n-1}$ given X_n. What happens if we replace n with a random time T?
Let X be a Markov chain. For $n \in \mathbb{N}$, let $\sigma(X_0, \ldots, X_n)$ denote the σ-algebra generated by the random variables $X_0, \ldots, X_n$. By definition, this is the smallest σ-algebra on Ω with respect to which each X_i, $1 \leq i \leq n$, is measurable. We have $\sigma(X_0, \ldots, X_n) \subset \sigma(X_0, \ldots, X_n, X_{n+1})$, so as we increase the number of random variables, the σ-algebra they generate becomes larger. Such an increasing sequence of σ-algebras is called a *filtration*. One can think of $\sigma(X_0, \ldots, X_n)$ as measuring the information about the chain X that is known at time n. As n increases, more information becomes available.

Definition 2.2.10 *1. Let $(\Omega, \mathcal{A})$ be a measurable space and let $(\mathcal{A}_n)_{n\in\mathbb{N}}$ be a family of σ-algebras on Ω such that $\mathcal{A}_n \subset \mathcal{A}_{n+1} \subset \mathcal{A}$ for every $n \in \mathbb{N}$. Then we call $(\mathcal{A}_n)_{n\in\mathbb{N}}$ a* filtration *of $(\Omega, \mathcal{A})$.*

2. A random variable T with values in $\mathbb{N} \cup \{\infty\}$ is called a stopping time *with respect to a filtration $(\mathcal{A}_n)_{n\in\mathbb{N}}$ if the event $\{T \leq n\}$ is an element of $\mathcal{A}_n$ for every $n \in \mathbb{N}$. It is called a stopping time with respect to a Markov chain X if it is a stopping time with respect to the filtration $(\sigma(X_0, \ldots, X_n))_{n\in\mathbb{N}}$.*

Example 2.2.11 *Let $(X_n)_{n\in\mathbb{N}}$ be a random walk on $\mathbb{Z}$.*

- *For $j \in \mathbb{Z}$, set*

$$T_j = \inf\{n \geq 1 : X_n = j\}.$$

The random variable T_j indicates the first time the random walk visits the state j. It is called the first passage time *for state j. For any $n \in \mathbb{N}$, we have*

$$\{T_j \leq n\} = \bigcup_{k=1}^{n} \{X_k = j\}.$$

And for $1 \leq k \leq n$*, the event* $\{X_k = j\}$ *lies in* $\sigma(X_k) \subset \sigma(X_0, \dots, X_n)$*. As* $\sigma(X_0, \dots, X_n)$ *is closed under countable unions,* $\cup_{k=1}^n \{X_k = j\} \in \sigma(X_0, \dots, X_n)$*. This shows that* T_j *is a stopping time with respect to* X*.*

- *For* $j \in \mathbb{Z}$*, set*

$$L_j = \sup\{n \in \mathbb{N} : X_n = j\}.$$

The random variable L_j *indicates the last time* X *vists the state* j*. For any* $n \in \mathbb{N}$*,*

$$\{L_j \leq n\} = \bigcap_{k=n+1}^{\infty} \{X_k \neq j\}.$$

As this event depends on the random variables $X_{n+1}, X_{n+2}, \dots$*, it is typically not contained in* $\sigma(X_0, \dots, X_n)$*, i.e.* L_j *is typically not a stopping time with respect to* X*.*

Definition 2.2.12 *Let* T *be a stopping time with respect to a filtration* $(\mathcal{A}_n)_{n \in \mathbb{N}}$ *of a measurable space* $(\Omega, \mathcal{A})$*. Then, we set*

$$\mathcal{A}_T = \{A \in \mathcal{A} : A \cap \{T \leq n\} \in \mathcal{A}_n \ \forall n \in \mathbb{N}\}.$$

One can show that $\mathcal{A}_T$ *is a* σ*-algebra on* Ω*.*

Definition 2.2.13 *For* $i \in E$*, let* δ_i *denote the* Dirac measure *on* E *defined by*

$$\delta_i(\{j\}) = \begin{cases} 0, & j \neq i, \\ 1, & j = i. \end{cases}$$

Theorem 2.2.14 *Let* $X = (X_n)_{n \in \mathbb{N}}$ *be Markov*$(\boldsymbol{\alpha}, P)$*. Let* T *be a stopping time with respect to* X *and let* $i \in E$ *such that* $\mathbb{P}(T < \infty, X_T = i) > 0$*. Then,* $(X_{T+n})_{n \in \mathbb{N}}$ *is Markov*(δ_i, P) *with respect to the probability measure* $\mathbb{P}(\bullet \mid T < \infty, X_T = i)$*. Furthermore,* $(X_{T+n})_{n \in \mathbb{N}}$ *is independent of the* σ*-algebra* $\mathcal{A}_T$ *(where* $\mathcal{A}_n = \sigma(X_0, \dots, X_n)$*), meaning that for any* $n \in \mathbb{N}$*,* $i_0, \dots, i_n \in E$ *and* $B \in \mathcal{A}_T$ *we have*

$$\mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n, B \mid T < \infty, X_T = i) = \mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n \mid T < \infty, X_T = i) \tag{2.3}$$

$$\cdot \mathbb{P}(B \mid T < \infty, X_T = i). \tag{2.4}$$

Remark 2.2.15 *Any Markov process in discrete time – even if it has an uncountable state space – satisfies a version of Theorem 2.2.14. Most Markov processes in continuous time (e.g. Brownian motion) also satisfy a version of Theorem 2.2.14, but some don't. If a Markov process satisfies a version of Theorem 2.2.14, we say that it has the* strong Markov property*. Theorem 2.2.14 then tells us that any Markov chain on a discrete state space and in discrete time has the strong Markov property.*

Proof. Let $n \in \mathbb{N}$, $i_0, \dots, i_n \in E$ and $B \in \mathcal{A}_T$. If $i_0 \neq i$ we have 0 on both sides of the formula in (2.3) and (2.4), so we will assume from now on that $i_0 = i$. Then,

$$\begin{aligned}&\mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n, B \mid T < \infty, X_T = i)\\ &\qquad = \frac{\mathbb{P}(X_{T+1} = i_1, \dots, X_{T+n} = i_n, B, T < \infty, X_T = i)}{\mathbb{P}(T < \infty, X_T = i)}.\end{aligned}$$

As $\{T < \infty\} = \cup_{m=0}^{\infty} \{T = m\}$, we can write the term in the numerator as

$$\begin{aligned}&\sum_{m=0}^{\infty} \mathbb{P}(X_{T+1} = i_1, \dots, X_{T+n} = i_n, B, T = m, X_T = i) \qquad (2.5)\\ &= \sum_{m=0}^{\infty} \mathbb{P}(X_{m+1} = i_1, \dots, X_{m+n} = i_n, B, T = m, X_m = i)\\ &= \sum_{m=0}^{\infty} \mathbb{P}(X_{m+1} = i_1, \dots, X_{m+n} = i_n, B, T = m \mid X_m = i)\mathbb{P}(X_m = i).\end{aligned}$$

Fix $m \in \mathbb{N}$ such that $\mathbb{P}(X_m = i) > 0$. Since X satisfies the Markov property, the random variables $X_{m+1}, \dots, X_{m+n}$ are independent of $X_0, \dots, X_m$ under $\mathbb{P}(\bullet \mid X_m = i)$. (Technically, the Markov property only said that X_{m+1} is conditionally independent of $X_0, \dots, X_{m-1}$ given X_m, but one can derive this stronger statement without too much effort.) And since $B \in \mathcal{A}_T$, we have

$$B \cap \{T = m\} \in \sigma(X_0, \dots, X_m).$$

Hence,

$$\begin{aligned}&\mathbb{P}(X_{m+1} = i_1, \dots, X_{m+n} = i_n, B, T = m \mid X_m = i)\mathbb{P}(X_m = i)\\ =&\mathbb{P}(X_{m+1} = i_1, \dots, X_{m+n} = i_n \mid X_m = i)\mathbb{P}(B, T = m, X_m = i).\end{aligned}$$

Homogeneity implies

$$\mathbb{P}(X_{m+1} = i_1, \dots, X_{m+n} = i_n \mid X_m = i) = p_{ii_1} \dots p_{i_{n-1}i_n},$$

which doesn't depend on m. As a result, the term in (2.5) equals

$$p_{ii_1} \dots p_{i_{n-1}i_n} \sum_{m=0}^{\infty} \mathbb{P}(B, T = m, X_T = i) = p_{ii_1} \dots p_{i_{n-1}i_n} \mathbb{P}(B, T < \infty, X_T = i).$$

For the moment, let $B = \Omega$. Then, we have shown that

$$\mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n \mid T < \infty, X_T = i) = \delta_i(\{i_0\}) p_{ii_1} \dots p_{i_{n-1}i_n},$$

i.e. $(X_{T+n})_{n \in \mathbb{N}}$ is indeed $Markov(\delta_i, P)$ under $\mathbb{P}(\bullet \mid T < \infty, X_T = i)$. Hence, for arbitrary $B \in \mathcal{A}_T$,

$$\begin{aligned}&\mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n, B \mid T < \infty, X_T = i)\\ &= \mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n \mid T < \infty, X_T = i) \frac{\mathbb{P}(B, T < \infty, X_T = i)}{\mathbb{P}(T < \infty, X_T = i)}\\ &= \mathbb{P}(X_T = i_0, \dots, X_{T+n} = i_n \mid T < \infty, X_T = i)\mathbb{P}(B \mid T < \infty, X_T = i).\end{aligned}$$

□

Remark 2.2.16 *While the previous proof may look somewhat technical, its main steps are simple and can serve as a blueprint for dealing with Markov chains and stopping times: First, condition on the possible values the stopping time T may assume; then, for each fixed value of T you have a regular Markov chain that you can manipulate as needed; finally, remove the conditioning to get back T.*

2.3 Recurrence and transience

Let P be a stochastic matrix. We say that a state $i \in E$ is *recurrent* (with respect to P or with respect to a Markov chain X with transition matrix P) if

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) = 1,$$

where $X \sim Markov(\delta_i, P)$. We call it *transient* if

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) = 0$$

for $X \sim Markov(\delta_i, P)$. Recurrent states are those to which you keep coming back and transient states are those which you eventually leave forever. Our first important result will be that every state is either recurrent or transient.
Recall the first passage time

$$T_j = \inf\{n \geq 1 : X_n = j\},$$

where we use the convention that $\inf \varnothing = \infty$. Define inductively

$$T_j^{(0)} = 0, \quad T_j^{(1)} = T_j$$

and

$$T_j^{(r+1)} = \inf\{n \geq T_j^{(r)} + 1 : X_n = j\}, \quad r \geq 1.$$

Then, $T_j^{(r)}$ is the time at which X visits the state j for the rth time. For $r \geq 1$, the length of the rth excursion from j is defined as

$$S_j^{(r)} = \begin{cases} T_j^{(r)} - T_j^{(r-1)} & \text{if } T_j^{(r)} < \infty \\ +\infty & \text{otherwise.} \end{cases}$$

Draw a picture to illustrate this.

Lemma 2.3.1 *Let $r = 2, 3, \ldots,$ and suppose that $\mathbb{P}(T_j^{(r-1)} < \infty) > 0$. Conditional on $T_j^{(r-1)} < \infty$, $S_j^{(r)}$ is independent of $\mathcal{A}_{T_j^{(r-1)}}$ and*

$$\mathbb{P}(S_j^{(r)} = n \mid T_j^{(r-1)} < \infty) = \mathbb{P}_j(T_j = n).$$

Proof. We apply the strong Markov property for the stopping time $T = T_j^{(r-1)}$. By definition of T, $X_T = j$, which spares us conditioning on the event $\{X_T = j\}$. Under the probability measure $\mathbb{P}(\bullet \mid T < \infty)$, $(X_{T+n})_{n\in\mathbb{N}}$ is independent from $\mathcal{A}_T$ and is $Markov(\delta_j, P)$. Conditional on $T < \infty$, we have

$$S_j^{(r)} = \inf\{n \geq 1 : X_{T+n} = j\}.$$

Therefore, under $\mathbb{P}(\bullet \mid T < \infty)$, $S_j^{(r)}$ is the first passage time to j of the Markov chain $(X_{T+n})_{n\in\mathbb{N}}$ that has the same distribution as the original chain X under $\mathbb{P}_j$. It follows that $S_j^{(r)}$ is independent of $\mathcal{A}_T$ and that

$$\mathbb{P}(S_j^{(r)} = n \mid T_j^{(r-1)} < \infty) = \mathbb{P}_j(T_j = n).$$

□

Define the *number of visits* V_i to state i as

$$V_i = \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=i\}},$$

and note that

$$\mathbb{E}_i[V_i] = \mathbb{E}_i \sum_{n=0}^{\infty} \mathbb{1}_{\{X_n=i\}} = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Here, $\mathbb{E}_i$ stands for expected value with respect to the probability measure $\mathbb{P}_i$. With this notation at hand, we can now say that a state i is recurrent if $\mathbb{P}_i(V_i = \infty) = 1$. It is transient if $\mathbb{P}_i(V_i = \infty) = 0$. We can compute the distribution of V_i under $\mathbb{P}_i$ in terms of the *return probability*

$$f_i = \mathbb{P}_i(T_i < \infty).$$

Lemma 2.3.2 *For any $r \in \mathbb{N}$,*

$$\mathbb{P}_i(V_i > r) = f_i^r.$$

Proof. We prove the statement by induction. To deal with the base case $r = 0$, observe that $\mathbb{P}_i(V_i > 0) = 1$. This is because V_i already counts the 0th visit to state i (i.e. the start of the chain at i) as one visit. By convention, $f_i^0 = 1$, so the base case has been verified. In the induction step, we will use the general observation that

$$\{V_i > r\} = \{T_i^{(r)} < \infty\}, \quad r \in \mathbb{N}. \tag{2.6}$$

Suppose the statement holds for some $r \in \mathbb{N}$. Then,

$$\begin{aligned}
\mathbb{P}_i(V_i > r+1) &= \mathbb{P}_i(T_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(T_i^{(r)} < \infty, S_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(S_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty)\mathbb{P}_i(T_i^{(r)} < \infty).
\end{aligned}$$

With (2.6) and the induction hypothesis, we obtain

$$\mathbb{P}_i(T_i^{(r)} < \infty) = \mathbb{P}_i(V_i > r) = f_i^r.$$

Thus, if $\mathbb{P}_i(T_i^{(r)} < \infty) = 0$, we also have $f_i = 0$ and the desired formula holds. If $\mathbb{P}_i(T_i^{(r)} < \infty) > 0$, Lemma 2.3.1 yields

$$\mathbb{P}_i(S_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty) = \sum_{n=1}^{\infty} \mathbb{P}_i(S_i^{(r+1)} = n \mid T_i^{(r)} < \infty) = \sum_{n=1}^{\infty} \mathbb{P}_i(T_i = n) = \mathbb{P}_i(T_i < \infty) = f_i.$$

To summarize:

$$\mathbb{P}_i(S_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty)\mathbb{P}_i(T_i^{(r)} < \infty) = f_i f_i^r = f_i^{r+1}.$$

□

Recall the basic formula for the expectation of a non-negative random variable V with values in $\mathbb{N}$:

$$\mathbb{E}[V] = \sum_{r=0}^{\infty} \mathbb{P}(V > r).$$

(To derive it, write $\mathbb{E}[V]$ as $\sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mathbb{P}(V = n)$ and change the order of summation.) The next theorem gives two useful criteria to establish recurrence or transience of a given state.

Theorem 2.3.3 *(i) If $f_i = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$;*

(ii) If $f_i < 1$, then i is transient and $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.
In particular, every state is either transient or recurrent.

Proof. If $f_i = 1$, then

$$\mathbb{P}_i(V_i = \infty) = \lim_{r\to\infty} \mathbb{P}_i(V_i > r) = \lim_{r\to\infty} f_i^r = 1.$$

The first equality follows from $\{V_i = \infty\} = \cap_{r=0}^{\infty}\{V_i > r\}$ and the fact that probability measures are continuous from above, meaning that if $\mathbb{P}$ is a probability measure and if $(A_n)_{n\in\mathbb{N}}$ is a family of events such that $A_1 \supset A_2 \supset A_3 \supset \dots$, then $\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}(\cap_{n\in\mathbb{N}} A_n)$. The second equality follows from Lemma 2.3.2. Thus, i is recurrent and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i[V_i] = \infty.$$

If $f_i < 1$, we have

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i[V_i] = \sum_{r=0}^{\infty} \mathbb{P}_i(V_i > r) = \sum_{r=0}^{\infty} f_i^r = \frac{1}{1 - f_i} < \infty.$$

The key point here is the convergence of the geometric series. If the event $\{V_i = \infty\}$ had a positive $\mathbb{P}_i$-probability, then $\mathbb{E}_i[V_i]$ would be infinite. Hence, $\mathbb{P}_i(V_i = \infty) = 0$ and i is transient. □

This result allows us to provide simple criteria for recurrence/ transience. The criteria we are about to develop hold for an arbitrary countable state space E, but they are particularly helpful if the Markov chain evolves on a *finite* state space. First, we show that recurrence/ transience is a property shared by all members of a communicating class.

Theorem 2.3.4 *Let C be a communicating class. Then either all states in C are transient or recurrent.*

Proof. Take $i, j \in C$ and suppose that i is transient. In order to prove the theorem, it is enough to show that j is transient because every state is either transient or recurrent. Since $i \leftrightarrow j$, there are $n, m \in \mathbb{N}$ such that $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$. As a corollary of the Chapman–Kolmogorov equation, we have the estimate

$$p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}, \quad r \in \mathbb{N}.$$

This should also be clear on a more probabilistic level: On the left, we have the probability of going from i to i in $(n + r + m)$ steps. On the right, we have the probability of going from i to i in $(n + r + m)$ steps AND of being at j after n steps AND of being again at j after another r steps. The inequality above yields the estimate

$$\sum_{r=0}^{\infty} p_{jj}^{(r)} \leq \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(n+r+m)}.$$

Since state i is transient, part (ii) of Theorem 2.3.3 implies that the series on the right converges. Hence, the series on the left converges as well and j is also transient, again by Theorem 2.3.3. □

We can therefore speak of a recurrent or transient communicating class (or Markov chain/ transition matrix if it is irreducible).

Theorem 2.3.5 *Every recurrent communicating class is closed.*

Proof. We show the contraposition: Every communicating class that isn't closed is transient. Let C be a communicating class which is not closed. Then, there exist $i \in C$ and $j \notin C$ such that $i \to j$. In particular, there is $m \geq 1$ such that $p_{ij}^{(m)} > 0$. Since j is not in the same communicating class as i and since $i \to j$, we necessarily have $j \nrightarrow i$. This implies

$$\mathbb{P}_i(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}) = 0,$$

which, in light of $p_{ij}^{(m)} > 0$, can only hold if

$$\mathbb{P}_i(\{X_n = i \text{ for infinitely many } n\}) < 1.$$

As a result, i is transient. □

Theorem 2.3.5 states that any recurrent communicating class is closed. How about the converse statement? Is it always true that any closed communicating class is recurrent? When we discuss recurrence and transience for the simple random walk on $\mathbb{Z}$ in the following section, we shall see that this is not the case. However, we can turn Theorem 2.3.5 into an "if and only if"-statement if we restrict ourselves to finite communicating classes.

Theorem 2.3.6 *Every finite closed communicating class is recurrent. In particular, a finite communicating class is recurrent if and only if it is closed.*

Proof. Let C be a closed and finite communicating class of a Markov chain X whose initial distribution $\boldsymbol{\alpha}$ satisfies $\alpha_i = 0$ if $i \notin C$. Let us first show that there exists a state $i \in C$ such that

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) > 0.$$

If this wasn't the case, we would have

$$0 = \sum_{i \in C} \mathbb{P}(X_n = i \text{ for infinitely many } n) = \mathbb{P}\Big(\bigcup_{i \in C} \{X_n = i \text{ for infinitely many } n\}\Big).$$

But since X starts in C and since C is closed, each random variable X_n takes on values only in C. And since C is finite,

$$\mathbb{P}\Big(\bigcup_{i \in C} \{X_n = i \text{ for infinitely many } n\}\Big) = 1,$$

a contradiction. Let $i \in C$ such that

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) > 0.$$

Since $\{X_n = i \text{ for infinitely many } n\} = \{X_n = i \text{ for infinitely many } n > T_i\}$, we can write the probability on the left as

$$\mathbb{P}(X_{T_i+n} = i \text{ for infinitely many } n \mid T_i < \infty)\mathbb{P}(T_i < \infty).$$

By the strong Markov property,

$$\mathbb{P}(X_{T_i+n} = i \text{ for infinitely many } n \mid T_i < \infty) = \mathbb{P}_i(X_n = i \text{ for infinitely many } n).$$

Thus, the probability on the right is positive and i is recurrent. □

It is relatively easy to spot closed communicating classes, so determining whether a finite communicating class is transient or recurrent is usually not hard.

Corollary 2.3.7 *If E is the finite state space of a Markov chain, it can be uniquely partitioned as*

$$E = I \cup E_1 \cup E_2 \cup \cdots \cup E_m,$$

where I is the set of all transient states and $E_1, E_2, \ldots, E_m$ are the closed communicating classes.

The following theorem is often useful.

Theorem 2.3.8 *Suppose P is irreducible and recurrent. Then, for all $j \in E$, we have $\mathbb{P}(T_j < \infty) = 1$.*

Notice that we are not making any assumptions on the initial distribution.
Proof. By the law of total probability,

$$\mathbb{P}(T_j < \infty) = \sum_{i \in E} \alpha_i \mathbb{P}_i(T_j < \infty).$$

Thus, we only need to show $\mathbb{P}_i(T_j < \infty) = 1$ for all $i \in E$ with $\alpha_i > 0$. Fix such $i \in E$. Since P is irreducible, there is $m \in \mathbb{N}$ such that $p_{ji}^{(m)} > 0$. As j is recurrent,

$$\begin{aligned} 1 &= \mathbb{P}_j(X_n = j \text{ for infinitely many } n) \\ &= \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1) \\ &= \sum_{k \in E} \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1 \mid X_m = k)\mathbb{P}_j(X_m = k). \end{aligned}$$

Let $k \in E$ such that $\mathbb{P}_j(X_m = k) > 0$. By the Markov property and homogeneity, $(X_{n+m})_{n \in \mathbb{N}}$ is $Markov(\delta_k, P)$ under $\mathbb{P}_j(\bullet \mid X_m = k)$. Hence,

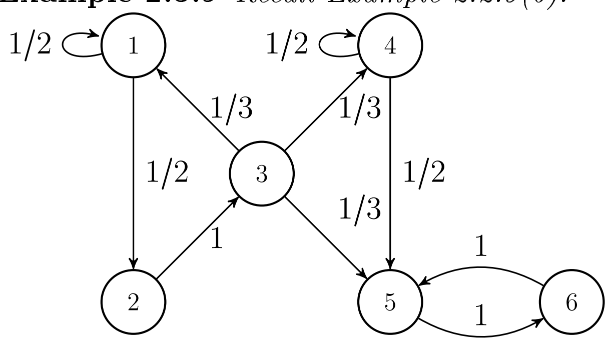
$$\mathbb{P}_j(X_n = j \text{ for some } n \geq m+1 \mid X_m = k) = \mathbb{P}_k(T_j < \infty).$$

This gives

$$1 = \sum_{k \in E} \mathbb{P}_k(T_j < \infty) p_{jk}^{(m)}.$$

As $\sum_{k \in E} p_{jk}^{(m)} = 1$, we have $\mathbb{P}_k(T_j < \infty) = 1$ for all $k \in E$ with $p_{jk}^{(m)} > 0$. In particular, $\mathbb{P}_i(T_j < \infty) = 1$. □

Example 2.3.9 *Recall Example 2.2.9(b):*



with communicating classes $\{1,2,3\}$, $\{4\}$, and $\{5,6\}$. Since only $\{5,6\}$ is closed, this is the only recurrent communicating class. The communicating classes $\{1,2,3\}$ and $\{4\}$ are transient.

The only interesting case left is then the case of communicating classes that are infinite and closed. A priori, such communicating classes could be recurrent or transient, and we will see in the next section that both of these situations occur.

2.3.1 Recurrence and transience for classical chains

Simple random walk on $\mathbb{Z}$

Recall that a simple random walk on $\mathbb{Z}$ is a Markov chain with transition probabilities

$$\mathbb{P}(X_{n+1} = y \mid X_n = x) = \begin{cases} p & y = x+1 \\ 1-p & y = x-1 \\ 0 & \text{otherwise} \end{cases},$$

where $p \in (0,1)$. As we assume $p \in (0,1)$, this Markov chain is irreducible. Suppose we start at 0, then $p_{00}^{(2n+1)} = 0$ for all n. The fact that simple random walk starting at 0 never returns to 0 at odd times is often a source of minor and at times of major technical difficulties when dealing with it. Any given sequence of $2n$ steps from 0 to 0 has probability $p^n(1-p)^n$, provided of course that we always step to a nearest neighbor. The total number of such sequences is the number of ways to choose n steps (say the ones taken to the right) out of $2n$. Thus,

$$p_{00}^{(2n)} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{(n!)^2} (p(1-p))^n.$$

We will use Stirling's approximation to $n!$, which states that

$$\lim_{n\to\infty} \frac{n! e^n}{\sqrt{2\pi n}\; n^n} = 1.$$

With this we obtain

$$\lim_{n\to\infty} \frac{p_{00}^{(2n)} \sqrt{\pi n}}{(4p(1-p))^n} = 1.$$

In particular, there are constants $0 < c < C$ such that

$$c \frac{(4p(1-p))^n}{\sqrt{n}} \le p_{00}^{(2n)} \le C \frac{(4p(1-p))^n}{\sqrt{n}}, \quad n \in \mathbb{N}.$$

In the symmetric case $p = 1/2$, $4p(1-p) = 1$, so

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} \ge c \sum_{n=0}^{\infty} \frac{1}{\sqrt{n}} = \infty,$$

showing with Theorem 2.3.3 that the one-dimensional simple symmetric random walk is recurrent.
If $p \ne \frac{1}{2}$, then $4p(1-p) = r < 1$ and thus

$$\sum_{n=0}^{\infty} p_{00}^{(2n)} \le C \sum_{n=0}^{\infty} \frac{r^n}{\sqrt{n}} < \infty.$$

Thus the asymmetric simple random walk is transient.

Birth and death chain on $\mathbb{N}$

On $\mathbb{N}$, consider a simplified version of the birth and death chain with transition probabilities

$$p_{ij} = \begin{cases} p, & i \geq 1,\ j = i+1, \\ 1-p, & i \geq 1,\ j = i-1, \\ 1, & i = 0,\ j = 1, \end{cases} \tag{2.7}$$

where $p \in (0,1)$. We can think of this Markov chain as a simple random walk on $\mathbb{N}$ that is reflected at 0. Just like the simple random walk on $\mathbb{Z}$, it is irreducible. For $k \in \mathbb{N}$, let $u(k) = \mathbb{P}_k(X_n \neq 0 \quad \forall n \in \mathbb{N})$. The chain is recurrent if and only if $u(k) = 0$ for all $k \in \mathbb{N}$: If the chain is recurrent, we have

$$\mathbb{P}_k(T_0 < \infty) = 1, \quad k \in \mathbb{N},$$

by Theorem 2.3.8. Therefore,

$$u(k) = \mathbb{P}_k(T_0 = \infty, X_0 \neq 0) \leq \mathbb{P}_k(T_0 = \infty) = 1 - \mathbb{P}_k(T_0 < \infty) = 0.$$

Conversely, if $u(k) = 0$ for all $k \in \mathbb{N}$, we have in particular $u(1) = 0$, so

$$0 = \mathbb{P}_1(T_0 = \infty, X_0 \neq 0) = \mathbb{P}_1(T_0 = \infty) = 1 - \mathbb{P}_1(T_0 < \infty).$$

With $p_{01} = 1$, we obtain

$$\mathbb{P}_0(T_0 < \infty) = \mathbb{P}_1(T_0 < \infty) = 1$$

and recurrence follows with Theorem 2.3.3. Clearly $u(0) = 0$, and moreover

$$\begin{aligned} u(k) =& p_{kk-1}\mathbb{P}_k(X_n \neq 0 \quad \forall n \in \mathbb{N} \mid X_1 = k-1) + p_{kk+1}\mathbb{P}_k(X_n \neq 0 \quad \forall n \in \mathbb{N} \mid X_1 = k+1) \\ =& (1-p)u(k-1) + pu(k+1), \quad k \geq 1. \end{aligned}$$

After rearranging terms, this gives

$$u(k+1) - u(k) = \frac{1-p}{p}(u(k) - u(k-1)) = \left(\frac{1-p}{p}\right)^k (u(1) - u(0)) = \left(\frac{1-p}{p}\right)^k u(1).$$

Using telescopic summation,

$$u(k+1) = (u(k+1) - u(k)) + (u(k) - u(k-1)) + \cdots + (u(1) - u(0)) = u(1)\sum_{j=0}^{k}\left(\frac{1-p}{p}\right)^j. \tag{2.8}$$

Suppose now that $p \leq \frac{1}{2}$. In this case, the series

$$\sum_{j=0}^{\infty}\left(\frac{1-p}{p}\right)^j$$

diverges. Thus, if $u(1)$ was strictly positive, the sequence $u(1)\sum_{j=0}^{k}(\frac{1-p}{p})^j$ on the right side of (2.8) would diverge to ∞ as $k \to \infty$. However, the sequence $u(k+1)$ on the left side of (2.8) is bounded by 1 because every $u(k)$ is the probability of some event. This implies $u(1) = 0$, and on account of (2.8) then even $u(k) = 0$ for all $k \in \mathbb{N}$. As a result, the chain is recurrent if $p \leq \frac{1}{2}$.

In tutorial, you were also asked to show that the chain is transient if $p > \frac{1}{2}$. One way of doing this is by comparison to the simple random walk on $\mathbb{Z}$.

Simple symmetric random walk on $\mathbb{Z}^d$

For $d \geq 1$, we consider the simple symmetric random walk (SSRW) on $\mathbb{Z}^d$, the Markov chain X on $\mathbb{Z}^d$ with transition probabilities

$$\mathbb{P}(X_{n+1} = y | X_n = x) = \begin{cases} \frac{1}{2d}, & \|y - x\|_1 = 1 \\ 0, & \text{otherwise} \end{cases}.$$

Here, $\|\cdot\|_1$ is the 1-norm on $\mathbb{R}^d$ defined by $\|v\|_1 = |v_1| + \ldots + |v_d|$. In words, if X is at a point $x \in \mathbb{Z}^d$ at time n, it jumps to one of the $2d$ nearest neighbors of x on the lattice $\mathbb{Z}^d$ at time $(n+1)$. The Markov chain X is irreducible, and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \sum_{n=0}^{\infty} p_{00}^{(2n)}$ for every $i \in \mathbb{Z}^d$.

We have already shown that X is recurrent if $d = 1$. Next, we consider the case $d = 2$. As in the one-dimensional case, it is impossible for the SSRW in two dimensions (and in fact in any dimension) to move from 0 to 0 in an odd number of steps. In order for it to move from 0 to 0 in exactly $2n$ steps, the number of steps to the left must equal the number of steps to the right, and the number of steps up must equal the number of steps down. For $0 \leq i \leq n$, suppose there were i steps taken to the left and right, and $n - i$ steps up and down. The number of admissible paths of length $2n$ from 0 to 0 is thus given by the sum of multinomial coefficients

$$\sum_{i=0}^{n} \binom{2n}{i, i, n-i, n-i} = \sum_{i=0}^{n} \frac{(2n)!}{i!i!(n-i)!(n-i)!}.$$

As each path has the same likelihood 4^{-2n} (where 4 comes from $4 = 2d$), we have

$$p_{00}^{(2n)} = 4^{-2n} \sum_{i=0}^{n} \frac{(2n)!}{i!i!(n-i)!(n-i)!} = 4^{-2n} \binom{2n}{n} \sum_{i=0}^{n} \binom{n}{i}^2 = 4^{-2n} \binom{2n}{n}^2 \geq \frac{c}{n} \tag{2.9}$$

for some $c > 0$. In the last step, we used Stirling's formula. The second to last step is easiest to understand if we interpret the binomial coefficients combinatorially: The term

$$\binom{2n}{n}$$

gives the number of ways of choosing n balls from an urn with n red and n blue balls. For every choice we make, there is some $i \in \{0, 1, \ldots, n\}$ such that i of the chosen balls are

red and $n-i$ of the chosen balls are blue. The estimate in (2.9) implies $\sum_{n=0}^{\infty} p_{ii}^{(2n)} = \infty$, so SSRW in dimension 2 is recurrent as well.
For $d = 3$, we have

$$p_{00}^{(2n)} = 6^{-2n} \sum_{i,j,k \in \mathbb{N}, i+j+k=n} \frac{(2n)!}{(i!j!k!)^2} = 2^{-2n} \binom{2n}{n} 3^{-2n} \sum_{i,j,k \in \mathbb{N}, i+j+k=n} \binom{n}{i,j,k}^2$$
$$= 2^{-2n} \binom{2n}{n} 3^{-n} \max_{i,j,k \in \mathbb{N}, i+j+k=n} \binom{n}{i,j,k} 3^{-n} \sum_{i,j,k \in \mathbb{N}, i+j+k=n} \binom{n}{i,j,k}.$$

Next, observe that

$$\sum_{i,j,k \in \mathbb{N}, i+j+k=n} \binom{n}{i,j,k} = 3^n$$

because both terms give the number of ways of placing n balls in three boxes. For the case where $n = 3m$, we have

$$\binom{n}{i,j,k} = \frac{n!}{i!j!k!} \leq \binom{n}{m,m,m},$$

so

$$p_{00}^{(6m)} \leq 2^{-2n} \binom{2n}{n} 3^{-n} \binom{n}{m,m,m} \leq \frac{C}{n^{\frac{3}{2}}}$$

by Stirling's formula. Hence $\sum_{m=0}^{\infty} p_{00}^{(6m)} < \infty$, by the comparison test. On the other hand, $p_{00}^{(6m-2)} \leq 6^2 p_{00}^{(6m)}$ and $p_{00}^{(6m-4)} \leq 6^4 p_{00}^{(6m)}$, so we must have

$$\sum_{n=0}^{\infty} p_{00}^{(n)} < \infty.$$

This shows that SSRW in three dimensions is transient.

Remark 2.3.10 *For the d-dimensional SSRW, the probability $p_{00}^{(2n)}$ is of order $n^{-\frac{d}{2}}$. As the series $\sum_{n=1}^{\infty} n^{-\frac{d}{2}}$ converges for all $d \geq 3$, SSRW is transient in any dimension $d \geq 3$.*

2.4 Stationarity

Definition 2.4.1 *Let π be a probability measure on E. Since E is countable, such a probability measure can be identified with a vector of $|E|$ components that are nonnegative and sum up to 1. We call π* invariant measure *or* stationary distribution *of a Markov chain with transition matrix P if*

$$\pi_i = \sum_{j \in E} \pi_j p_{ji}, \qquad i \in E.$$

This means that $\pi = \pi P$, i.e. π, viewed as a row vector, is a left eigenvector of the matrix P to the eigenvalue 1.

Remark 2.4.2 *If π is an invariant measure, the following holds.*

1. $\pi = \pi P^n$ *for any* n,

2. *If* X_0 *has distribution* π, *then* X_n *has distribution* π *for any* $n \in \mathbb{N}$. *This is the reason we call* π *invariant. To see this fact, observe that*

$$\mathbb{P}_\pi(X_1 = i) = \sum_{j \in E} \pi_j \mathbb{P}(X_1 = i \mid X_0 = j) = \sum_{j \in E} \pi_j p_{ji} = \pi_i.$$

   Here, $\mathbb{P}_\pi$ *is the law of a Markov chain in* $Markov(\pi, P)$.

2.4.1 Long-run proportion of number of visits

Recall the total number of visits to state $j \in E$, $V_j = \sum_{k=0}^{\infty} \mathbb{1}_{\{X_k = j\}}$, and define the number of visits up to time $n-1$ $(n \geq 1)$ by

$$V_j^{(n)} = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = j\}}.$$

The next theorem gives the long-run proportion of time spent by a Markov chain in each state.

Theorem 2.4.3 *Let* X *be an irreducible Markov chain.* $\mathbb{P}$*–almost surely, we have*

$$\lim_{n \to \infty} \frac{V_j^{(n)}}{n} = \frac{1}{\mathbb{E}_j T_j}, \quad j \in E.$$

Remark 2.4.4 *In the theorem above,* T_j *denotes the first passage time of state* j. *(Recall that* $T_j = \inf\{n \geq 1 : X_n = j\}$*). The term* $1/\mathbb{E}_j T_j$ *should be interpreted as* 0 *if* $\mathbb{E}_j T_j = \infty$.

Proof. If the Markov chain is transient, then V_j is $\mathbb{P}$–almost surely finite and thus

$$\frac{V_j^{(n)}}{n} \leq \frac{V_j}{n} \to 0 = \frac{1}{\mathbb{E}_j T_j}, \quad \mathbb{P} - a.s.$$

Let us now consider the recurrent case and define the rth passage time $T_j^{(r)}$ and the length of the rth excursion $S_j^{(r)}$ as before. Since the Markov chain is irreducible and recurrent, a stronger version of Theorem 2.3.8 states that $\mathbb{P}$–almost surely, $T_j^{(r)} < \infty$ for every $r \in \mathbb{N}$. By Lemma 2.3.1, the random variables $(S_j^{(r)})_{r \geq 1}$ are thus independent and have the same distribution under $\mathbb{P}$ as T_j under $\mathbb{P}_j$. For $k \in \mathbb{N}$, we have $T_j^{(k)} = \sum_{r=1}^{k} S_j^{(r)}$. By the strong law of large numbers, $\mathbb{P}$–almost surely,

$$\lim_{k \to \infty} \frac{T_j^{(k)}}{k} = \mathbb{E}_j T_j \in (0, \infty]. \tag{2.10}$$

Fix $\omega \in \Omega$ such that $\lim_{k\to\infty} \frac{T_j^{(k)}(\omega)}{k} = \mathbb{E}_j T_j$. To finish the proof of Theorem 2.4.3, it is enough to show that

$$\limsup_{n\to\infty} \frac{V_j^{(n)}(\omega)}{n} \le \frac{1}{\mathbb{E}_j T_j} \tag{2.11}$$

and

$$\liminf_{n\to\infty} \frac{V_j^{(n)}(\omega)}{n} \ge \frac{1}{\mathbb{E}_j T_j}. \tag{2.12}$$

Let $M \in (0, \mathbb{E}_j T_j)$. Then, there is $n(1) \in \mathbb{N}$ such that $\frac{T_j^{(n)}(\omega)}{n} \ge M$ for all $n \ge n(1)$. This implies $T_j^{(n)}(\omega) \ge \lceil Mn \rceil$, $n \ge n(1)$. Here, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. Notice that $V_j^{(n)} \le k$ if and only if $T_j^{(k)} \ge n$. Therefore, $V_j^{\lceil Mn \rceil}(\omega) \le n$ for all $n \ge n(1)$. Hence,

$$\frac{V_j^{\lceil Mn \rceil}(\omega)}{\lceil Mn \rceil} \le \frac{n}{\lceil Mn \rceil} \le \frac{1}{M}, \quad n \ge n(1).$$

Now, let $k \ge \lceil Mn(1) \rceil$. Then, there are $n \ge n(1)$ and $l \in \{0, \dots, \lceil M \rceil\}$ such that $k = \lceil Mn \rceil + l$. With this representation,

$$\frac{V_j^{(k)}(\omega)}{k} \le \frac{V_j^{\lceil Mn \rceil}(\omega) + l}{\lceil Mn \rceil + l} \le \frac{V_j^{\lceil Mn \rceil}(\omega)}{\lceil Mn \rceil} + \frac{\lceil M \rceil}{\lceil Mn \rceil}.$$

As $\lim_{n\to\infty} \frac{\lceil M \rceil}{\lceil Mn \rceil} = 0$, we have

$$\limsup_{n\to\infty} \frac{V_j^{(n)}(\omega)}{n} \le \frac{1}{M}.$$

As $\frac{1}{M} > \frac{1}{\mathbb{E}_j T_j}$ can be chosen arbitrarily close to $\frac{1}{\mathbb{E}_j T_j}$, we deduce (2.11). The proof of (2.12) is similar. We leave it to you as an exercise. □

Corollary 2.4.5 *If X is an irreducible Markov chain, we have*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} = \frac{1}{\mathbb{E}_j T_j}, \quad j \in E.$$

Proof. Since $|V_j^{(n)}/n| \le 1$ and $\mathbb{E}_i \mathbb{1}_{\{X_k = j\}} = p_{ij}^{(k)}$, this follows from the dominated convergence theorem. □

Definition 2.4.6 *We call a state $j \in E$* positive recurrent *if (i) it is recurrent and (ii) it has finite* expected return time, *i.e.*

$$\mathbb{E}_j T_j < \infty.$$

We call j null recurrent *if it is recurrent, but not positive recurrent.*

Example 2.4.7 *On the state space $E = \mathbb{N}$, consider the Markov chain $(X_n)_{n\in\mathbb{N}}$ with transition probabilities*

$$p_{ij} = \begin{cases} 1, & i > 0,\ j = i-1, \\ p_j, & i = 0,\ j > 0. \end{cases}$$

Here, $(p_j)_{j\geq 1}$ are given real numbers in $[0,1]$. We would like to identify conditions on $(p_j)_{j\geq 1}$ under which the Markov chain is (i) irreducible, (ii) recurrent and (iii) positive recurrent.

First, observe that the Markov chain is irreducible if and only if the set $\{j \geq 1 : p_j > 0\}$ is unbounded. If X is irreducible, then it is recurrent because

$$\mathbb{P}_0(T_0 < \infty) = \sum_{j=1}^{\infty} p_j \mathbb{P}_j(T_0 < \infty) = \sum_{j=1}^{\infty} p_j = 1.$$

If X is not irreducible, the chain has transient states as well. As X cannot stay at 0, we have $\mathbb{P}_0(T_0 = 1) = 0$ and thus

$$\mathbb{E}_0 T_0 = \sum_{j=2}^{\infty} j \mathbb{P}_0(T_0 = j) = \sum_{j=2}^{\infty} j p_{j-1} = \sum_{j=1}^{\infty} (j+1) p_j = 1 + \sum_{j=1}^{\infty} j p_j.$$

If $\sum_{j=1}^{\infty} j p_j < \infty$ (e.g. $p_j = j^{-3}$), 0 (and thus X, as we'll see later) is positive recurrent, and if $\sum_{j=1}^{\infty} j p_j = \infty$ (e.g. $p_j = j^{-2}$), 0 is null recurrent.

Our immediate goal is to relate the concepts of positive recurrence and null recurrence to the question whether a given Markov chain has a unique invariant measure.

Theorem 2.4.8 *Let X be irreducible and let π be an invariant measure of X. Then, X is recurrent and*

$$\pi_j = \frac{1}{\mathbb{E}_j T_j}, \quad j \in E.$$

In particular, there exists a state that is positive recurrent.

Proof. For any $n \geq 1$,

$$\pi_j = \frac{1}{n} \sum_{k=0}^{n-1} (\pi P^k)_j = \sum_{i\in E} \pi_i \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)}$$

by invariance of π. For $i \in E$, Corollary 2.4.5 yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} p_{ij}^{(k)} = \frac{1}{\mathbb{E}_j T_j}.$$

Thus,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} (\pi P^k)_j = \sum_{i\in E} \frac{\pi_i}{\mathbb{E}_j T_j} = \frac{1}{\mathbb{E}_j T_j}.$$

Since π_j does not depend on n, $\pi_j = \frac{1}{\mathbb{E}_j T_j}$. Any state $j \in E$ such that $\pi_j > 0$ is therefore positive recurrent and in particular recurrent. Irreducibility of X then implies recurrence of the entire Markov chain. □

Remark 2.4.9 *We have just shown that if an irreducible Markov chain has an invariant measure, then the invariant measure is unique, i.e. there is not more than one invariant measure. It could happen, though, that an irreducible Markov chain has no invariant measure at all.*

Lemma 2.4.10 *If an irreducible Markov chain X is recurrent and if $i \in E$ is positive recurrent, then, $\mathbb{P}$–almost surely,*

$$\lim_{n\to\infty} \frac{V_j^{(n)}}{n} = \frac{\mathbb{E}_i[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}]}{\mathbb{E}_i T_i}, \quad j \in E.$$

Proof. Since X is irreducible and recurrent, we have $\lim_{n\to\infty} V_i^{(n)} = \infty$, $\mathbb{P}$–almost surely. For any realization of the Markov chain with $\lim_{n\to\infty} V_i^{(n)} = \infty$, there is $N \in \mathbb{N}$ such that $V_i^{(n)} \geq 1$ for all $n \geq N$. For such n, we have

$$\frac{V_j^{(n)}}{n} = \frac{1}{n} \sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \tag{2.13}$$

$$+ \frac{V_i^{(n)}}{n} \frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}-1} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} \tag{2.14}$$

$$+ \frac{1}{n} \sum_{k=T_i^{(V_i^{(n)})}}^{n-1} \mathbb{1}_{\{X_k=j\}} \quad , \tag{2.15}$$

where we also used that $T_i^{(V_i^{(n)})} \leq n-1$ by definition. Since $\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}$ does not depend on n, the term on the right side of (2.13) tends to 0 as $n \to \infty$. By Theorem 2.4.3 and our assumption that X is irreducible, we have $\mathbb{P}$–almost surely

$$\lim_{n\to\infty} \frac{V_i^{(n)}}{n} = \frac{1}{\mathbb{E}_i T_i}. \tag{2.16}$$

Recurrence and irreducibility of X also imply that the random variables

$$\sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}}, l \geq 1,$$

are independent and identically distributed by virtue of Theorem 2.3.8 and the strong Markov property. By the strong law of large numbers,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{n-1} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} = \mathbb{E}_i \left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \right], \quad \mathbb{P} - a.s.$$

As $\lim_{n\to\infty} V_i^{(n)} = \infty$, $\mathbb{P}$–almost surely, we also have

$$\lim_{n\to\infty} \frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}-1} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} = \mathbb{E}_i \left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \right], \quad \mathbb{P}-a.s.$$

So far, we haven't used the assumption that i is positive recurrent. This is needed now: As i is positive recurrent, $\frac{1}{\mathbb{E}_i T_i} > 0$ and

$$\mathbb{E}_i \left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \right] \leq \mathbb{E}_i T_i < \infty.$$

Hence, as $n \to \infty$, the term in (2.14) converges $\mathbb{P}$–almost surely to

$$\frac{\mathbb{E}_i [\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}]}{\mathbb{E}_i T_i}.$$

By definition,

$$\begin{aligned} \frac{1}{n} \sum_{k=T_i^{(V_i^{(n)})}}^{n-1} \mathbb{1}_{\{X_k=j\}} \leq & \frac{V_i^{(n)}}{n} \frac{1}{V_i^{(n)}} \sum_{k=T_i^{(V_i^{(n)})}}^{T_i^{(V_i^{(n)}+1)}-1} \mathbb{1}_{\{X_k=j\}} \\ = & \frac{V_i^{(n)}}{n} \left(\frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} - \frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}-1} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} \right). \end{aligned}$$

Thus, the strong law of large numbers implies

$$\begin{aligned} \lim_{n\to\infty} & \left(\frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} - \frac{1}{V_i^{(n)}} \sum_{l=1}^{V_i^{(n)}-1} \sum_{k=T_i^{(l)}}^{T_i^{(l+1)}-1} \mathbb{1}_{\{X_k=j\}} \right) \\ & = \mathbb{E}_i \left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \right] - \mathbb{E}_i \left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}} \right] = 0, \quad \mathbb{P}-a.s. \end{aligned}$$

Together with (2.16), this implies that the term in (2.15) converges to 0 $\mathbb{P}$–almost surely as $n \to \infty$. This finishes the proof. □

Theorem 2.4.11 *If a Markov chain is irreducible and has a positive recurrent state, then every state is positive recurrent and*

$$\pi_j = \frac{1}{\mathbb{E}_j(T_j)}, \quad j \in E \tag{2.17}$$

is the unique invariant measure of the chain.

Proof. Since the Markov chain is irreducible, Theorem 2.4.8 implies that if there is an invariant measure, it is unique and has the form of (2.17). We still need to show that positive recurrence of one state of an irreducible Markov chain implies positive recurrence of all states, and that (2.17) defines in fact an invariant measure. Let i be positive recurrent and $j \in E \setminus \{i\}$. Irreducibility gives $s, r \in \mathbb{N}$ such that $p_{ji}^{(r)} p_{ij}^{(s)} > 0$. As we saw in the proof of Theorem 2.3.4,

$$\frac{1}{n} \sum_{k=0}^{n-1} p_{jj}^{(r+k+s)} \geq p_{ji}^{(r)} p_{ij}^{(s)} \frac{1}{n} \sum_{k=0}^{n-1} p_{ii}^{(k)}.$$

Letting $n \to \infty$, we have

$$\frac{1}{\mathbb{E}_j T_j} \geq \frac{p_{ji}^{(r)} p_{ij}^{(s)}}{\mathbb{E}_i T_i} > 0$$

by virtue of Theorem 2.4.3. (For the term on the left, we used

$$\frac{1}{n} \sum_{k=0}^{n-1} p_{jj}^{(r+k+s)} = -\frac{1}{n} \sum_{k=0}^{r+s-1} p_{jj}^{(k)} + \frac{1}{n} \sum_{k=0}^{n-1} p_{jj}^{(k)} + \frac{1}{n} \sum_{k=n}^{r+s+n-1} p_{jj}^{(k)},$$

where the first and third term on the right tend to 0 as $n \to \infty$.) Since $\mathbb{E}_j T_j < \infty$, the state j is positive recurrent.

The only points left to check are that π as defined in (2.17) is a probability measure (i.e. that its components sum up to 1) and that π is invariant under the transition matrix P. Fix any state $i \in E$. Since the Markov chain is positive recurrent, Theorem 2.4.3, Lemma 2.4.10 and uniqueness of the limit imply

$$\frac{1}{\mathbb{E}_j T_j} = \frac{\mathbb{E}_i[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}]}{\mathbb{E}_i T_i}, \quad j \in E.$$

Hence,

$$\sum_{j \in E} \pi_j = \frac{\sum_{j \in E} \mathbb{E}_i[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}]}{\mathbb{E}_i T_i} = \frac{\mathbb{E}_i[\sum_{k=0}^{T_i-1} \sum_{j \in E} \mathbb{1}_{\{X_k=j\}}]}{\mathbb{E}_i T_i} = 1.$$

To prove invariance of π, fix $j \in E$ and choose any $i \in E \setminus \{j\}$. Then,

$$\begin{aligned} \mathbb{E}_i[T_i]\pi_j =& \mathbb{E}_i\left[\sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}\right] = \mathbb{E}_i\left[\sum_{n=1}^{\infty} \mathbb{1}_{\{T_i=n\}} \sum_{k=0}^{T_i-1} \mathbb{1}_{\{X_k=j\}}\right] \\ =& \mathbb{E}_i\left[\sum_{n=1}^{\infty} \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=j, T_i=n\}}\right] = \mathbb{E}_i\left[\sum_{k=0}^{\infty} \sum_{n=k+1}^{\infty} \mathbb{1}_{\{X_k=j, T_i=n\}}\right] = \sum_{k=0}^{\infty} \mathbb{P}_i(X_k = j, k < T_i). \end{aligned} \tag{2.18}$$

If $k = 0$ we have $\mathbb{P}_i(X_k = j, k < T_i) = 0$, and if $k = 1$ we have $\mathbb{P}_i(X_k = j, k < T_i) = p_{ij}$. Now, suppose $k \geq 2$. As $i \neq j$, we have $\mathbb{P}_i(X_k = j, T_i = k) = 0$. Thus,

$$\mathbb{P}_i(X_k = j, k < T_i) = \mathbb{P}_i(X_k = j, k-1 < T_i) = \sum_{l \in E \setminus \{i\}} \mathbb{P}_i(X_k = j, X_{k-1} = l, k-1 < T_i).$$

All this shows that the term at the very right of (2.18) equals

$$\begin{aligned} & p_{ij} + \sum_{k=2}^{\infty} \sum_{l \in E \setminus \{i\}} \mathbb{P}_i(X_k = j, X_{k-1} = l, k-1 < T_i) \\ =& p_{ij} + \sum_{k=2}^{\infty} \sum_{l \in E \setminus \{i\}} \mathbb{P}_i(X_{k-1} = l, k-1 < T_i) \mathbb{P}_i(X_k = j \mid X_{k-1} = l, k-1 < T_i). \end{aligned} \tag{2.19}$$

For $2 \le k < \infty$ and $l \in E \setminus \{i\}$, we have

$$\mathbb{P}_i(X_k = j \mid X_{k-1} = l, k-1 < T_i) = \mathbb{P}_i(X_k = j \mid X_{k-1} = l, X_{k-2} \neq i, \dots, X_1 \neq i) = p_{lj}$$

by the Markov property. As a result, the term on the right side of (2.19) can be written as

$$\begin{aligned} & p_{ij} + \sum_{l \in E \setminus \{i\}} \sum_{k=2}^{\infty} \mathbb{P}_i(X_{k-1} = l, k-1 < T_i) p_{lj} \\ =& \mathbb{E}_i[T_i] \pi_i p_{ij} + \sum_{l \in E \setminus \{i\}} \sum_{k=0}^{\infty} \mathbb{P}_i(X_k = l, k < T_i) p_{lj} \\ =& \mathbb{E}_i[T_i] \pi_i p_{ij} + \sum_{l \in E \setminus \{i\}} \mathbb{E}_i[T_i] \pi_l p_{lj} \\ =& \mathbb{E}_i[T_i] \sum_{l \in E} \pi_l p_{lj}. \end{aligned}$$

Dividing by $\mathbb{E}_i T_i$ shows that π is invariant. □

We now summarize several previous results for irreducible chains.
(1) There exists an invariant measure if and only if there exists a positive recurrent state.
(2) In this case, every state is positive recurrent, and the invariant measure is $\pi_i = (\mathbb{E}_i T_i)^{-1}$, $i \in E$.

Remark 2.4.12 *Let X be an irreducible Markov chain on a finite state space. As an exercise, show that X is positive recurrent. Hence, any irreducible Markov chain on a finite state space has a unique invariant measure.*

Remark 2.4.13 *To find invariant measures π of a transition matrix P, one can:*

1. *Solve the eigenvector equation $\pi P = \pi$ for a vector π with nonnegative entries that satisfies $\sum_{i \in E} \pi_i = 1$.*

2. *Try "detailed balance": see if there exists $\pi = (\pi_i)_{i \in E}$ with nonnegative entries and $\sum_{i \in E} \pi_i = 1$ such that $\pi_i p_{ij} = \pi_j p_{ji}$ for any $i, j \in E$. This is often easier to solve than $\pi_j = \sum_{i \in E} \pi_i p_{ij}$, $j \in E$, and it is a sufficient condition for π being an invariant measure. However, not every invariant measure satisfies detailed balance (see Ex. 2 on Série 5).*

3. Intuition.

Example 2.4.14 (Birth and death chain on $\mathbb{N}$) *Recall the transition probabilities of a general birth and death chain on $\mathbb{N}$:*

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \begin{cases} p_i, & j = i+1 \\ r_i, & j = i \\ q_i, & j = i-1 \end{cases}, \qquad q_0 = 0, \qquad p_i + q_i + r_i = 1. \tag{2.20}$$

To simplify the ensuing analysis, we will also assume that $q_n > 0$ for all $n \geq 1$ and $p_n > 0$ for all $n \in \mathbb{N}$. This guarantees that the chain is irreducible. If π is an invariant measure, then it must satisfy

$$\pi_n = \pi_{n+1} q_{n+1} + \pi_{n-1} p_{n-1} + \pi_n (1 - p_n - q_n), \quad n \geq 1,$$

so

$$\pi_{n+1} q_{n+1} - \pi_n p_n = \pi_n q_n - \pi_{n-1} p_{n-1} = \cdots = \pi_1 q_1 - \pi_0 p_0.$$

But $\pi_0 = \pi_0 r_0 + \pi_1 q_1$, i.e. $\pi_0(1 - r_0) = \pi_1 q_1$ and $\pi_0 p_0 = \pi_1 q_1$. Combining this, we obtain

$$\pi_n q_n - \pi_{n-1} p_{n-1} = 0 \quad \forall n \geq 1.$$

Hence,

$$\pi_n = \pi_{n-1} \frac{p_{n-1}}{q_n} = \ldots = \pi_0 \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n}, \quad n \geq 1. \tag{2.21}$$

According to what we have discovered before, the chain is positive recurrent if and only if it admits an invariant measure. In light of (2.21), this is in turn equivalent to the existence of $\pi_0 > 0$ such that

$$\pi_0 + \pi_0 \sum_{n=1}^{\infty} \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n} = 1.$$

Consider the special case $p_0 = 1$ and $p_i = p$, $q_i = 1 - p$ for $i \geq 1$. Then,

$$\pi_0 + \pi_0 \sum_{n=1}^{\infty} \frac{p_0 \cdots p_{n-1}}{q_1 \cdots q_n} = \pi_0 + \frac{\pi_0}{1-p} \sum_{n=0}^{\infty} \left(\frac{p}{1-p} \right)^n.$$

The series on the right converges if and only if $p < \frac{1}{2}$. In this case, the entire term on the right equals

$$\pi_0 \frac{2-2p}{1-2p},$$

which is 1 for $\pi_0 = \frac{1-2p}{2-2p}$. The invariant measure of the birth and death chain in case $p < \frac{1}{2}$ is thus

$$\pi_n = \begin{cases} \frac{1-2p}{2-2p}, & n = 0, \\ \frac{1-2p}{2(1-p)^2} \left(\frac{p}{1-p} \right)^{n-1}, & n \geq 1. \end{cases}$$

If $p \geq \frac{1}{2}$, there is no invariant measure. By our existence result for the invariant measure of an irreducible Markov chain, the birth and death chain is positive recurrent if and only if $p < \frac{1}{2}$. If $p = \frac{1}{2}$, the chain is null recurrent.

2.4.2 Periodicity

In order to discuss under which conditions a Markov chain converges in law to its invariant measure, we need the notion of periodicity that we introduce in this subsection.

Definition 2.4.15 *The* period d_i *of a state* $i \in E$ *is the greatest common divisor of* $\{r > 0 : p_{ii}^{(r)} > 0\}$. *If* $d_i = 1$, *we say that the state* i *is* aperiodic.

Example 2.4.16 *Consider the two-state Markov chain with transition matrix*

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For any $n \in \mathbb{N}$, P^{2n} *is the identity matrix, and* $P^{2n+1} = P$. *Thus, for* $i \in \{1,2\}$, $p_{ii}^{(r)} > 0$ *if and only if* r *is even. It follows that* $\{r > 0 : p_{ii}^{(r)} > 0\}$ *is the set of even positive integers and that* $d_i = 2$. *In particular, neither of the two states is aperiodic. Also notice that the sequence of matrices* $(P^n)_{n\in\mathbb{N}}$ *does not converge as* $n \to \infty$. *If a Markov chain is not aperiodic, the initial distribution still has a noticeable effect on the distribution of* X_n *for large* n. *In other words, we never forget where we started from.*

Theorem 2.4.17 *If* $i \leftrightarrow j$, *then* $d_i = d_j$. *In particular, if the Markov chain is irreducible, then there exists* $d \geq 1$ *such that* $d_i = d$ *for all* $i \in E$. *The integer* d *is called the* period *of the Markov chain. If* $d = 1$, *then the Markov chain is called* aperiodic.

Proof. If $i = j$, the statement clearly holds, so we may assume that $i \neq j$. Since $i \to j$, $j \to i$ and $i \neq j$, there exist $m, n \geq 1$ such that $p_{ij}^{(m)} > 0$ and $p_{ji}^{(n)} > 0$. By Chapman–Kolmogorov,

$$p_{ii}^{(m+n)} = \sum_{k\in E} p_{ik}^{(m)} p_{ki}^{(n)} \geq p_{ij}^{(m)} p_{ji}^{(n)} > 0.$$

Since d_i is a divisor for every $r > 0$ such that $p_{ii}^{(r)} > 0$, it follows that $d_i|(m+n)$. Now, let $r > 0$ such that $p_{jj}^{(r)} > 0$. Then,

$$p_{ii}^{(m+r+n)} \geq p_{ij}^{(m)} p_{jj}^{(r)} p_{ji}^{(n)} > 0.$$

This implies $d_i|(m+r+n)$, and hence $d_i|r$. We have thus shown that d_i is a common divisor of $\{r > 0 : p_{jj}^{(r)} > 0\}$. As d_j is the greatest common divisor of this set, we have $d_i \leq d_j$. Reversing the roles of i and j gives the desired equality. □

Lemma 2.4.18 *Let* t *be a positive integer and let* $n \in \mathbb{N}$. *If* $p_{ii}^{(nt)} > 0$ *and* $p_{ii}^{(n+1)t} > 0$ *then* $p_{ii}^{(vt)} > 0$ *for every integer* $v \geq n(n-1)$.

Proof. Let us first discuss the case $n = 0$. In this case, $p_{ii}^{(t)} > 0$, so

$$p_{ii}^{(vt)} \geq (p_{ii}^{(t)})^v > 0.$$

Now, suppose $n > 0$. Then, any $v \geq n(n-1)$ can be written as $kn+w$, where $0 \leq w \leq n-1 \leq k$. Adding and subtracting wn, we obtain

$$v = (k-w)n + w(n+1).$$

Thus,

$$p_{ii}^{(vt)} \geq p_{ii}^{((k-w)nt)} p_{ii}^{w(n+1)t} \geq (p_{ii}^{(nt)})^{k-w} (p_{ii}^{(n+1)t})^w > 0.$$

□

Remark 2.4.19 *Lemma 2.4.18 implies in particular that if there is $t \geq 1$ such that $p_{ii}^{(t)} > 0$, then $p_{ii}^{(vt)} > 0$ for every $v \in \mathbb{N}$.*

Theorem 2.4.20 *The state i is aperiodic if and only if there is $r_0 > 0$ such that $p_{ii}^{(r)} > 0$ for all $r \geq r_0$.*

Proof. A set of integers that contains all but finitely many integers only has 1 as common divisor, so it is clear that a state i with $p_{ii}^{(r)} > 0$ for every $r \geq r_0$ is aperiodic. For the converse implication, we will show that there is $n > 0$ such that $p_{ii}^{(n)}$ and $p_{ii}^{(n+1)} > 0$. This fact, together with Lemma 2.4.18 with $t = 1$, yields the desired result. Let us suppose there is no n such that $p_{ii}^{(n)}, p_{ii}^{(n+1)} > 0$, and derive a contradiction. Under this assumption,

$$c := \inf\{r - s : r > s, p_{ii}^{(r)} > 0, p_{ii}^{(s)} > 0\} > 1.$$

Then, we can show that there is some large n_0 such that $p_{ii}^{(nc)} > 0$ for all $n \geq n_0$: By definition of c, there is $n \in \mathbb{N}$ such that $p_{ii}^{(n)}, p_{ii}^{(n+c)} > 0$. Thus, $p_{ii}^{(nc)} \geq (p_{ii}^{(n)})^c > 0$ and $p_{ii}^{(c(n+1))} = p_{ii}^{((c-1)n+n+c))} \geq (p_{ii}^{(n)})^{c-1} p_{ii}^{(n+c)} > 0$. This lets us apply Lemma 2.4.18 with $t = c$. Next, we show that there are $m \geq n_0$ and $v \in \{1, \ldots, c-1\}$ such that $p_{ii}^{(mc+v)} > 0$. Since $c > 1$ and since $d_i = 1$, there is $r > 0$ such that $p_{ii}^{(r)} > 0$ and such that c does not divide r. (Otherwise c would be a common divisor of $\{r > 0 : p_{ii}^{(r)} > 0\}$ that is greater than the greatest common divisor.) Let $N \in \mathbb{N}$ be so large that $Nr > n_0 c$, and assume further that c does not divide Nr. (As c does not divide r, such N always exists, for otherwise we would have for large N both $c \mid Nr$ and $c \mid (N+1)r$, which gives $c \mid (N+1)r - Nr = r$.) Then, we can represent Nr as $mc + v$ for $m \geq n_0$ and $v \in \{1, \ldots, c-1\}$. And by Remark 2.4.19, $p_{ii}^{(Nr)} > 0$. As a result, $p_{ii}^{(mc)}, p_{ii}^{(mc+v)} > 0$ and $(mc+v) - mc = v < c$, which is in contradiction with the definition of c. □

A careful inspection of the proof shows that we have actually verified the following number-theoretic result: If $S \subset \mathbb{N} \setminus \{0\}$ has greatest common divisor 1 and is closed under both addition and multiplication by numbers $c \in \mathbb{N} \setminus \{0\}$, then S contains all but finitely many elements of $\mathbb{N} \setminus \{0\}$.

Remark 2.4.21 *If $i \in E$ such that $p_{ii} > 0$, then the state i is aperiodic. If an irreducible Markov chain has a state $i_0 \in E$ such that $p_{i_0 i_0} > 0$, then the entire Markov chain is aperiodic. However, the converse is wrong: there are aperiodic irreducible Markov chains that satisfy $p_{ii} = 0$ for all i, for example the chain with transition matrix*

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

2.4.3 Convergence to the invariant measure

Theorem 2.4.22 *Let $(X_n)_{n\in\mathbb{N}}$ be a Markov chain that is irreducible, positive recurrent and aperiodic, with some initial distribution α. Then, for any $i \in E$, we have*

$$\lim_{n\to\infty} \mathbb{P}(X_n = i) = \pi_i > 0,$$

where π is the unique invariant measure of X.

Remark 2.4.23 *If P is irreducible, aperiodic and not positive recurrent (i.e., transient or null recurrent), then*

$$\lim_{n\to\infty} \mathbb{P}(X_n = i) = 0, \quad i \in E.$$

In the transient case, this is an easy exercise. For the null recurrent case, see Theorem 1.8.5 in Norris's textbook. If P fails to be aperiodic, the sequence $\mathbb{P}(X_n = i)$ may not converge as $n \to \infty$ (see Example 2.4.16).

The proof of Theorem 2.4.22 relies on a technique called coupling. It was invented by the French mathematician Vincent Doeblin, son of the famous German writer Alfred Döblin. His Jewish family had escaped from Nazi Gerrmany to France in the 1930's. To evade capture by German troops in World War II, Doeblin committed suicide at the age of 25.

Remark 2.4.24 *(See also Ex. 6 on Série 3)*
Let α be a probability measure and let P be a stochastic matrix with respect to a countable state space E. We can assume without loss of generality that $E = \mathbb{N}$. (This should be clear if E is infinite. If E is finite, define a new probability measure $\tilde{\alpha}$ on $\mathbb{N}$ by assigning probability 0 to every state that is not already in E; similarly, define an infinite-dimensional stochastic matrix $\tilde{P}$ as the block-diagonal matrix with block P in the upper left corner and an infinite-dimensional identity matrix in the lower right corner.) We can construct a Markov chain $X \sim Markov(\alpha, P)$ on E as follows:

1. *Choose a random variable X_0 with law α;*
2. *Let $(U_n)_{n\geq 1}$ be an i.i.d. sequence of random variables that are uniformly distributed on the interval $[0,1]$ and live on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$ as X_0. We then construct the random variables $(X_n)_{n\geq 1}$ inductively: Suppose the random variable*

X_n is given, and we want to construct X_{n+1}. Fix $\omega \in \Omega$, and set $x = X_n(\omega)$. We set $X_{n+1}(\omega) = 1$ if $U_{n+1}(\omega) \in [0, p_{x,1})$, $X_{n+1}(\omega) = 2$ if $U_{n+1}(\omega) \in [p_{x,1}, p_{x,1} + p_{x,2})$, and so on.

Proof. (of Theorem 2.4.22) The main idea of coupling is to construct two copies of the Markov chain that have a very particular joint distribution. Let $(U_n)_{n\geq 1}$ and $(U_n')_{n\geq 1}$ be two i.i.d. sequences of random variables that are uniformly distributed on $[0,1]$ and live on the same probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Assume in addition that the sequences $(U_n)_{n\geq 1}$ and $(U_n')_{n\geq 1}$, viewed as stochastic processes, are independent of each other. Let X_0 be a random variable with law α, and let $X = (X_n)_{n\in\mathbb{N}}$ be $Markov(\alpha, P)$, generated from the random variables $(U_n)_{n\geq 1}$ as outlined in the previous remark. Then, let Y_0 be a random variable with law π (the unique invariant measure of P), independent of X_0, and let $Y = (Y_n)_{n\in\mathbb{N}} \sim Markov(\pi, P)$ be generated from $(U_n')_{n\geq 1}$.
First, observe that X and Y are independent because X_0 and Y_0 as well as $(U_n)_{n\geq 1}$ and $(U_n')_{n\geq 1}$ are independent. Next, note that Y_n has law π for every $n \in \mathbb{N}$ because π is invariant. In Série 6, we show that $(X,Y) = (X_n, Y_n)_{n\in\mathbb{N}}$ is an irreducible and aperiodic Markov chain on $E \times E$. It is also positive recurrent since $\tilde{\pi}_{(i,j)} = \pi_i \pi_j$ is an invariant measure. (Here, we use independence of X and Y.) Let x be any state in E, and let $T_{(x,x)}$ denote the first passage time of state (x,x) with respect to the Markov chain (X,Y). As (X,Y) is recurrent, $\mathbb{P}(T_{(x,x)} < \infty) = 1$, so in particular

$$T = \inf\{n \in \mathbb{N} : X_n = Y_n\} \leq T_{(x,x)} < \infty \quad \mathbb{P} - a.s.$$

In words, the two Markov chains X and Y meet in finite time with probability 1. Now, we define a new Markov chain $X' = (X_n')_{n\in\mathbb{N}}$ that corresponds to X up to the random time T of the first meeting of X and Y, and corresponds to Y after time T. In a sense, we force X and Y to stay together once they meet. Set $X_0' = X_0$. For fixed $\omega \in \Omega$, we set $X_n'(\omega) = X_n(\omega)$ if $T(\omega) \geq n$, and $X_n'(\omega) = Y_n(\omega)$ if $T(\omega) < n$. It is then not hard to see that X' is also $Markov(\alpha, P)$, just like X. The Markov chains X and X' are the two copies referred to in the first sentence of this proof. Because of how we defined X and X', we have already fixed their joint distribution. For any $n \in \mathbb{N}$,

$$\begin{aligned}
\mathbb{P}(X_n = i) = \mathbb{P}(X_n' = i) &= \mathbb{P}(X_n' = i, n > T) + \mathbb{P}(X_n' = i, n \leq T) \\
&= \mathbb{P}(Y_n = i, n > T) + \mathbb{P}(X_n = i, n \leq T) \\
&= \mathbb{P}(Y_n = i, n > T) + \mathbb{P}(Y_n = i, n \leq T) + \mathbb{P}(X_n = i, n \leq T) - \mathbb{P}(Y_n = i, n \leq T) \\
&= \mathbb{P}(Y_n = i) + \mathbb{P}(X_n = i, n \leq T) - \mathbb{P}(Y_n = i, n \leq T) \\
&= \pi_i + \mathbb{P}(X_n = i, n \leq T) - \mathbb{P}(Y_n = i, n \leq T).
\end{aligned}$$

Since $\lim_{n\to\infty} \mathbb{P}(n \leq T) = 0$, it follows that $\lim_{n\to\infty} \mathbb{P}(X_n = i) = \pi_i$. □

Remark 2.4.25 *What happens if P is not irreducible? If j is not positive recurrent, then $\lim_{n\to\infty} \mathbb{P}_i(X_n = j) = 0$, $i \in E$. If j is positive recurrent and aperiodic, then*

$$\lim_{n\to\infty} \mathbb{P}_i(X_n = j) = \pi_j \mathbb{P}_i(T_j < \infty),$$

where T_j is the first passage time for j and where π is the unique invariant measure supported on the communicating class of j.

2.5 Applications in statistics

2.5.1 Metropolis–Hastings algorithm

Markov chains are the basis for many important methods in statistics. We will first consider a way of generating random samples from a distribution.
Given a countable state space E and some probability measure p on E with $p_i > 0$ for all $i \in E$, the main idea for sampling from p is to construct an irreducible, aperiodic Markov chain $(X_n)_{n \geq 0}$ with invariant measure p. Once this has been accomplished, we can simply start the chain at some arbitrary state $j \in E$, and simulate transitions of the chain until the distribution of X_n is very close to p. More precisely, we use the result from Theorem 2.4.22 that

$$\lim_{n \to \infty} \mathbb{P}(X_n = i) = p_i.$$

Therefore, for n large enough, the distribution of X_n should be close to p.
How to construct a Markov chain with invariant measure p?

Lemma 2.5.1 (Metropolis–Hastings) *For any $i \in E$, let $q^i(\cdot)$ be a probability measure on E (called the proposal distribution) with $q^i(j) > 0$ if and only if $q^j(i) > 0$, $i, j \in E$. The Metropolis–Hastings algorithm constructs a Markov chain $(X_n)_{n \in \mathbb{N}}$ as follows:*

1. *Let $X_0 = i_0 \in E$ be an arbitrary deterministic starting value.*

2. *Suppose that $X_n = i$. Generate j from the proposal distribution q^i, and let U be uniformly distributed on $[0, 1]$. Then we define X_{n+1} by*

$$X_{n+1} = \begin{cases} j, & \text{if } U \leq \min\left\{\frac{p_j q^j(i)}{p_i q^i(j)}, 1\right\}, \\ i, & \text{otherwise.} \end{cases} \tag{2.22}$$

If the Markov chain $(X_n)_{n \in \mathbb{N}}$ is irreducible (e.g., if $q^i(j) > 0$ for all $i, j \in E$), then it has invariant measure p.

You will be asked to prove this lemma in tutorial. The proposal distribution q^i should

- be easy to simulate from;

- allow the Markov chain to efficiently cover the state space.

Remark 2.5.2 • *It is important to note that in the Metropolis–Hastings algorithm, the Markov chain* $(X_n)_{n\in\mathbb{N}}$ *only depends on the ratio* p_j/p_i*, that is, the probability measure* p *only has to be known* ***up to a constant****. Often,* p_i *has the form* h_i/Z*, where* $(h_i)_{i\in E}$ *are known and* $Z = \sum_{i\in E} h_i$ *is a normalizing constant. If* E *is large, it may be difficult to calculate* Z *explicitly even though we know the* h_i*'s. Fortunately, calculating* Z *is not needed because the chain in* (2.22) *only depends on* h_j/h_i*.*

- *The time period* $\{0, 1, \ldots, n_b\}$ *until* $\mathbb{P}(X_n \in \cdot)$ *is sufficiently close to* p *is called* burn-in time*. The corresponding values* $X_0, \ldots, X_{n_b}$ *have to be discarded since their distributions differ too much from* p*. A suitable number* n_b *is often determined by visual inspection of the chain.*

- *After the burn-in time, each* X_n*,* $n > n_b$*, has distribution close to* p*. Note that the samples are however correlated, i.e. not independent. If independent samples of* p *are required, then only every* m*th value of the chain should be taken, where* m *again depends on the specific Markov chain (e.g., on the the proposal distribution in the Metropolis–Hastings algorithm).*

- *In this course we only consider Markov chains and Markov processes on countable state spaces* E*. The theory of Markov chains and stationarity can be extended to uncountable state spaces (which are usually assumed to be complete, separable metric spaces), and the Metropolis–Hastings algorithm works in this setting as well (*p *and the* q^i*'s will then be densities).*

2.5.2 Bayesian statistics

Let $(p_\theta)_{\theta\in\Theta}$ be a parametric family of probability density functions (in the continuous case) or probability mass functions (in the discrete case) on $\mathbb{R}^d$, with countable parameter space Θ. We further suppose that we have independent observations $z_1, \ldots, z_n$ from the distribution associated with p_θ. In classical frequentist statistics, it is assumed that there is a true, deterministic value θ_0 that generated the data, and there are numerous methods to estimate this parameter, e.g. maximum likelihood.

In Bayesian statistics, one models the uncertainty about the underlying parameter as a random variable on Θ. In order to do so, one defines a *prior distribution* π on Θ, which incorporates the prior belief of the modeler about the parameter, without taking into account any data (the prior can include also beliefs about the shape/structure of the parameter). If now independent data $z_1, \ldots, z_n$ from the parametric model $\{p_\theta : \theta \in \Theta\}$ becomes available, one updates this prior belief with the new information. The result is the so-called *posterior distribution* on the parameter space Θ, which is given as $p(\theta \mid z_1, \ldots, z_n)$. By Bayes' theorem,

$$p(\theta \mid z_1, \ldots, z_n) = \frac{\pi(\theta) p(z_1, \ldots, z_n \mid \theta)}{p(z_1, \ldots, z_n)} = \frac{\pi(\theta) \prod_{i=1}^n p_\theta(z_i)}{p(z_1, \ldots, z_n)},$$

where $p(z_1, \dots, z_n) = \sum_{\theta \in \Theta} p(z_1, \dots, z_n \mid \theta)\pi(\theta)$ does not depend on θ.
Generating samples from the posterior distribution is generally not possible in a direct way. Instead, one simulates a Markov chain on the state space $E = \Theta$ with invariant measure $p(\theta \mid z_1, \dots, z_n)$. Since the posterior is usually only known up to a constant, the Metropolis–Hastings algorithm is one of the most popular methods to this end. Another widely used method is the so-called Gibbs sampler. If one is interested in a point estimator of the parameter, one can take for instance the mean of the posterior distribution.

2.6 Ergodic theorem

The ergodic theorem for Markov chains relates the time average of the Markov chain to the space average with respect to the invariant measure.

Theorem 2.6.1 (Ergodic theorem) *Let X be an irreducible, positive recurrent Markov chain with invariant measure π, and let $f : E \to \mathbb{R}$ be a bounded function. Then, $\mathbb{P}$–almost surely,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \mathbb{E}_\pi f = \sum_{i\in E} f(i)\pi_i.$$

Remark 2.6.2 *For a Markov chain with i.i.d. random variables $(X_n)_{n\in\mathbb{N}}$, the ergodic theorem (slightly generalized to unbounded functions) recovers the strong law of large numbers: In the i.i.d. case, the invariant measure of the Markov chain clearly equals the distribution of X_1.*

Proof. As f is bounded, there is $c > 0$ such that $|f(i)| \le c$ for all $i \in E$. Recall that

$$V_i^{(n)} = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k = i\}}, \quad i \in E,$$

and observe that

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) = \frac{1}{n} \sum_{k=0}^{n-1} \sum_{i\in E} \mathbb{1}_{\{X_k=i\}} f(X_k) = \sum_{i\in E} f(i) \frac{V_i^{(n)}}{n}.$$

Hence,

$$\left| \frac{1}{n} \sum_{k=0}^{n-1} f(X_k) - \mathbb{E}_\pi f \right| = \left| \sum_{i\in E} \left(\frac{V_i^{(n)}}{n} - \pi_i \right) f(i) \right| \le c \sum_{i\in E} \left| \frac{V_i^{(n)}}{n} - \pi_i \right|.$$

For pedagogical reasons, let us first consider the simpler case that E is finite. By Theorem 2.4.3 and Theorem 2.4.11, the set of $\omega \in \Omega$ such that

$$\lim_{n\to\infty} \frac{V_i^{(n)}(\omega)}{n} = \pi_i, \quad i \in E,$$

has probability measure 1 under $\mathbb{P}$. As E is finite, this implies for any such ω

$$\lim_{n\to\infty} \sum_{i\in E} \left| \frac{V_i^{(n)}(\omega)}{n} - \pi_i \right| = 0,$$

and the statement is proved. Now, suppose E is infinite. Since π is a probability measure, the infinite series $\sum_{i\in E} \pi_i$ converges to 1. Thus, for any $\epsilon > 0$ there is a finite subset J of E such that

$$\sum_{i\in J} \pi_i > 1 - \epsilon.$$

We have

$$\sum_{i\in E} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| = \sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| + \sum_{i\notin J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| \leq \sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| + \sum_{i\notin J} \left(\frac{V_i^{(n)}}{n} + \pi_i \right). \quad (2.23)$$

Since

$$\sum_{i\notin J} \frac{V_i^{(n)}}{n} = \sum_{i\in E} \pi_i - \sum_{i\in J} \frac{V_i^{(n)}}{n} \leq \sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| + \sum_{i\notin J} \pi_i,$$

the term on the right side of (2.23) is less than or equal to

$$2\sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| + 2\sum_{i\notin J} \pi_i \leq 2\sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| + 2\epsilon.$$

As J is finite,

$$\lim_{n\to\infty} \sum_{i\in J} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| = 0, \quad \mathbb{P} - a.s.$$

Thus,

$$\limsup_{n\to\infty} \sum_{i\in E} \left| \frac{V_i^{(n)}}{n} - \pi_i \right| \leq 2\epsilon, \quad \mathbb{P} - a.s.$$

Since $\mathbb{P}$ is continuous from above, letting $\epsilon \downarrow 0$ yields the desired convergence. □

Chapter 3

Markov processes

3.1 Continuous time Markov chains

Definition 3.1.1 *A* Markov process *(or* continuous time Markov chain*) is a collection* $(X(t))_{t\in\mathbb{R}_+}$ *of random variables with values in a countable set* E*. The index* t *often represents time. Unlike Markov chains, here we have continuous time processes.*
As for Markov chains, we will impose two conditions:

1. Markov property:

$$\mathbb{P}(X(t+s)=j \mid X(u), 0\le u\le t) = \mathbb{P}(X(t+s)=j \mid X(t)), \quad s,t\ge 0, i,j\in E.$$

2. Homogeneity:

$$\mathbb{P}(X(t+s)=j \mid X(t)=i) = \mathbb{P}(X(s)=j \mid X(0)=i) = P_{ij}(s), \quad s,t\ge 0, i,j\in E,$$

where the $P_{ij}(s)$ *are called* transition functions. *This means that the finite dimensional distributions of* X *are given by: for all* $n\in\mathbb{N}$, $0\le t_1\le\cdots\le t_n$ *and all states* $i_0,i_1,\dots,i_n\in E$*:*

$$\mathbb{P}(X(t_n)=i_n,\dots,X(t_1)=i_1|X(0)=i_0) = P_{i_0i_1}(t_1)P_{i_1i_2}(t_2-t_1)\dots P_{i_{n-1}i_n}(t_n-t_{n-1}).$$

Remark 3.1.2 *The set of stochastic matrices* $(P(t): t\ge 0)$ *with* $P(t)=(P_{ij}(t))_{i,j\in E}$ *is a* semigroup*, that is*

$$P(s+t) = P(s)P(t), \text{ for all } s,t\ge 0.$$

This is the Chapman-Kolmogorov equation for the continuous-time case.
Proof.

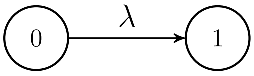
$$\begin{aligned} P_{ij}(s+t) &= \mathbb{P}(X(s+t)=j \mid X(0)=i) \\ &= \sum_{k\in E} P(X(s)=k \mid X(0)=i)P(X(s+t)=j \mid X(s)=k) \\ &= \sum_{k\in E} P_{i,k}(s)P_{k,j}(t) = (P(s)P(t))_{i,j}. \end{aligned}$$

□

Remark 3.1.3 *Let $E_i \sim Exp(\lambda_i)$, $i = 1, \dots, n$ be independent exponential random variables with rates $\lambda_i \in (0, \infty)$, i.e., $\mathbb{P}(E_i \le x) = 1 - e^{-\lambda_i x}$. Then*

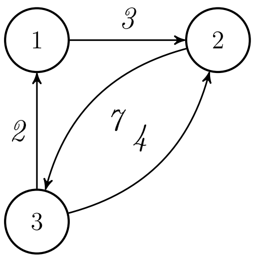
1. *$(E_i - s)|(E_i > s)$ is again $Exp(\lambda_i)$ distributed (memoryless property).*
2. *$\mathbb{E}(E_i) = 1/\lambda_i$;*
3. *$\min(E_1, \dots, E_n) \sim Exp(\lambda_1 + \dots + \lambda_n)$;*
4. *$\mathbb{P}(E_k = \min(E_1, \dots, E_n)) = \lambda_k/(\lambda_1 + \dots + \lambda_n)$.*

Example 3.1.4 *The simplest example is on the state space $E = \{0, 1\}$. When in state 0 we wait for a random exponential time $E_1 \sim Exp(\lambda)$ with parameter $\lambda \in (0, \infty)$ and then jump to 1.*



(Q-matrix)

Example 3.1.5 *Take $E = \{1, 2, 3\}$. In state 3 take two independent exponential times $E_1 \sim Exp(2)$ and $E_2 \sim Exp(4)$, if E_1 is the smaller then go to 1 after time E_1, and if E_2 is the smaller go to 2 after time E_2. Rules for states 1 and 3 are similar. The time spent in state 3 is $min(E_1, E_2) \sim Exp(2+4)$ (see Ex. 3, Série 1), and the probability of jumping from 3 to 1 is $2/(2+4) = 1/3$ (Ex. 2, Série 1).*



(Q-matrix)

Another way of thinking about the evolution of the Markov process X is in terms of its Q matrix which is known as the generator of the process.

Definition 3.1.6 (The Q-matrix) *A matrix $Q = (q_{ij})_{i,j \in E}$ is a Q-matrix if it satisfies*

1. *$q_{ii} \le 0$ for all $i \in E$;*
2. *$q_{ij} \ge 0$ for all $i \ne j$;*
3. *$\sum_{j \in E} q_{ij} = 0$ for all $i \in E$.*

The numbers $q_{ij}, j \ne i$ can be intrepeted as follows: Being in state i we sample for each other state $j \ne i$ an exponential random variable E_j with rate q_{ij} and then jump to state $k \in E$ after time E_k if $E_k = \min_{j \ne i} E_j$, then the process starts afresh. Equivalently, we can think as follows: Being in state i, the number $\delta_i = -q_{ii} = \sum_{j \ne i} q_{ij} \ge 0$ is the exponential rate with which the process leaves state i, and then jumps to state j with probability $\widehat{P}_{ij} = q_{ij}/\delta_i$. The matrix $\widehat{P}$ is called the jump matrix of the process X.

Example 3.1.7 (Poisson process) *Let us define a stochastic process $(N(t))_{t\geq 0}$ in the following way. For $i \in \mathbb{N}$, let E_i be independent copies of an $Exp(\lambda)$ distribution. Define $T_n = E_1 + \cdots + E_n$ and*

$$N(t) = \sum_{i=1}^{\infty} 1_{\{T_n \leq t\}}, \quad t \geq 0.$$

This counting process is called a homogeneous Poisson process with intensity λ. It can be used as model for the number of earthquakes, for instance. [picture]

- *The Q-matrix of the Poisson process is given by $q_{ii} = -\lambda$ and $q_{i,i+1} = \lambda$.*
- *For each $t \geq 0$, $N(t)$ has a Poisson distribution with parameter λt (Ex.), that is for all $j \in \mathbb{N}$*

$$P_{0j}(t) = e^{-\lambda t}\frac{(\lambda t)^j}{j!}.$$

- *$(N(t))_{t\geq 0}$ is a Markov process, i.e., for any $s \geq 0$, conditional on $N(s)$, $(N(s+t))_{t\geq 0}$, is again a Poisson process with rate λ, started in state $N(s)$, independent of $(N(r) : r \geq s)$. We even have that the Poisson process is homogeneous in space, that is $(\tilde{N}(t))_{t\geq 0} = (N(s+t) - N(s))_{t\geq 0}$ is a Poisson process with rate λ (started at 0), independent of $(N(r) : r \geq s)$. In fact, this even holds if the time s is replaced by a stopping time T (strong Markov property).*

Proof.

It suffices to show the claim conditional on the event $N(s) = i$ for all $i \geq 0$. We have

$$\{N(s) = i\} = \{T_i \leq s < T_{i+1}\} = \{T_i \leq s\} \cap \{E_{i+1} > s - T_i\}.$$

On this event

$$N(r) = \sum_{j=1}^{i} 1_{\{T_j \leq t\}}, \quad r \leq s,$$

and the holding times $\tilde{E}_1, \tilde{E}_2, \ldots$ are given by

$$\tilde{E}_1 = E_{i+1} - (s - T_i), \quad \tilde{E}_n = E_{i+n}, \quad n \geq 2.$$

Condition on $E_1, \ldots, E_i$ and $\{N(s) = i\}$, then by the memoryless property of E_{i+1} and independence, $\tilde{E}_1, \tilde{E}_2, \ldots$ are themselves independent $Exp(\lambda)$. Since $E_1, \ldots E_i$ are independent of $E_{i+1}, \ldots,$ it suffices to condition on $\{N(s) = i\}$, such that $\tilde{E}_1, \tilde{E}_2, \ldots$ are independent $Exp(\lambda)$, independent of $E_1, \ldots, E_i$. Hence, conditional on $\{N(s) = i\}$, $(\tilde{N}(t))_{t\geq 0}$ is a Poisson process with rate λ and independent of $(N(r) : r \leq s)$. □

Theorem 3.1.8 *Let $(N(t))_{t\geq 0}$ be a Poisson process. Then, conditional on the event $\{N(t) = n\}$, the jump times $T_1, T_2, \ldots, T_n$ have the same distribution as an ordered sample of size n from the uniform distribution on $[0, t]$.*

We refer to Theorem 2.4.6 in Norris for the proof.

Proposition 3.1.9 *Let $(N_1(t))_{t\geq 0}$ and $(N_2(t))_{t\geq 0}$ be two independent Poisson processes with parameters λ_1 and λ_2. The process $M(t) = N_1(t) + N_2(t), t \geq 0$ is again a Poisson process with parameters $\lambda_1 + \lambda_2$.*

Proof. We use induction to show that the waiting time E_i between the $(i-1)^{th}$ and i^{th} jumps of M follows an exponential law of parameters $(\lambda_1 + \lambda_2)$ independent of the time E_j for all $1 \leq j < i$, we prove this for all $i \in \mathbb{N}$.
For $i = 1$, we have that $T_1 = E_1 = \min\{S_1, S_2\}$ where S_i is the first arrival time of the process $(N_i(t))_{t\geq 0}$, $i = 1, 2$. As $S_i \sim Exp(\lambda_i)$ for $i = 1, 2$ and as S_1 and S_2 are independent, we have that $T_1 = E_1 \sim Exp(\lambda_1 + \lambda_2)$. Let us assume that $E_1, \cdots, E_i$ are independent and have the same law $Exp(\lambda_1 + \lambda_2)$. Using that $T_i = E_1 + \cdots + E_i$ is a stopping time for N_1 and N_2, we can use the strong Markov property

$$\tilde{M}(t) := M(t+T_i) - M(T_i) = (N_1(t+T_i) - N_1(T_i)) + (N_2(t+T_i) - N_2(T_i)) \stackrel{d}{=} N_1(t) + N_2(t) = M(t).$$

The process $(\tilde{M}(t))_{t\geq 0}$ is then independent of $E_1, \cdots, E_i$. Implying $E_{i+1} := \inf\{t > 0 \mid M(t + T_i) - M(T_i) > 0\} \stackrel{d}{=} \inf\{t > 0 \mid M(t) > 0\}$. We finally get that E_{i+1} is an exponential variable with parameter $(\lambda_1 + \lambda_2)$ and is independent of $E_1, \cdots, E_i$. □

Example 3.1.10 *In a population of size N, a rumour is begun by a single individual who tells it to everyone he meets; they in turn pass the rumour to everyone they meet. Assume that each individual meets another randomly with exponential rate $1/N$. How long does it take until everyone knows the rumour? If i people know the rumour, then $N-i$ do not, and the rate at which the rumour is passed on is (Q-matrix)*

$$q_{i,i+1} = i(N-i)/N, \quad i \in \{1,\dots,N\}.$$

The expected time until everyone knows the rumour is then

$$\mathbb{E}\sum_{i=1}^{N-1} E_i = \sum_{i=1}^{N-1} q_i^{-1} = \sum_{i=1}^{N-1} \frac{i(N-i)}{N} = \sum_{i=1}^{N-1} \frac{1}{i} + \frac{1}{N-i} = 2\sum_{i=1}^{N-1} \frac{1}{i} \sim 2\log N,$$

as $N \to \infty$.

Note that the *Strong Markov property* extends to the continuous-time setting: for any stopping time T for X and any state $j \in E$, we have

$$\mathbb{P}(X(T+s) = j \mid X(u), 0 \le u \le T) = \mathbb{P}(X(s) = j \mid X(T)).$$

We are now going to study the length of time that X spends in each state before the next transition, and formally prove that it is exponentially distributed, using the Markov property and the homogeneity property.

Definition 3.1.11 *We define by W_t the length of time the Markov process X remains in the state being occupied at time t, that is,*

$$W_t = \inf\{s > 0 \mid X(t+s) \neq X(t)\}.$$

We then have the important

Theorem 3.1.12 *Take a Markov process X. Then for all $i \in E$, there exists $\delta_i \in [0,\infty]$ such that for all $t, x \ge 0$,*

$$\mathbb{P}(W_t > x \mid X(t) = i) = e^{-\delta_i x},$$

and therefore $W_t | \{X(t) = i\} \sim Exp(\delta_i)$.

Proof. By the homogeneity property:

$$\mathbb{P}(W_t > x \mid X(t) = i) = \mathbb{P}(W_0 > x \mid X(0) = i) := f_i(x).$$

Since the event $\{W_0 > x + y\}$ is equivalent to the event $\{W_0 > x, W_x > y\}$, we have

$$\begin{aligned}
f_i(x+y) &= \mathbb{P}(W_0 > x + y \mid X(0) = i), \\
&= \mathbb{P}(W_0 > x, W_x > y \mid X(0) = i), \\
&= \mathbb{P}(W_0 > x \mid X(0) = i)\mathbb{P}(W_x > y \mid X(0) = i, W_0 > x), \\
&= f_i(x)\mathbb{P}(W_x > y \mid X(u) = i, 0 \le u \le x), \\
&= f_i(x)\mathbb{P}(W_x > y \mid X(x) = i), \quad \text{(Markov property)} \\
&= f_i(x)\mathbb{P}(W_0 > y \mid X(0) = i), \quad \text{(homogeneity)} \\
&= f_i(x)f_i(y).
\end{aligned}$$

The function $f_i(\cdot)$ is bounded by 0 and 1 and satisfies $f_i(x+y) = f_i(x)f_i(y)$ for all $x, y \ge 0$. Therefore it must be of the form $f_i(x) = e^{-\delta_i x}$ for some $\delta_i \in [0, \infty]$ (see also Serie 1, exercise 4(2)). □

Note that we accept $\delta_i = +\infty$ to cover the case where $W_t = 0$ with probability one. We then classify the states as follows, depending on the value of δ_i.

- If $0 < \delta_i < +\infty$:
  $$\mathbb{P}(W_t \le x \mid X(t) = i) = 1 - e^{-\delta_i x}.$$
  State i is a *stable* state.

- If $\delta_i = 0$:
  $$\mathbb{P}(W_t \le x \mid X(t) = i) = 0 \text{ for all } x \ge 0.$$
  Therefore $W_t = \infty$ with a probability 1, and i is an *absorbing* state.

- If $\delta_i = +\infty$:
  $$\mathbb{P}(W_t \le x \mid X(t) = i) = 1 \text{ for all } x \ge 0.$$
  State i is an *instantaneous* state (the process jumps of an instantaneous state as soon as it enters it, but also returns to it infinitely often within arbitrarily short times).

We will restrict ourselves to Markov processes with no instantaneous states.

Definition 3.1.13 *A Markov process is* conservative *if all its states are stable or absorbing. This is equivalent to each of the following:*

(i) the function $t \to X(t)$ is right-continuous almost surely;

(ii)

$$\lim_{s\to 0^+} P_{ij}(s) = \delta_{ij}.$$

Indeed, this follows since in this case

$$\begin{aligned} 1 = \mathbb{P}(W_t > 0 \mid X(t) = i) &= \mathbb{P}(\lim_{s\to 0^+} X(t+s) = X(t) \mid X(t) = i) \\ &= \mathbb{P}(\lim_{s\to 0^+} X(s) = i \mid X(t) = i) = \lim_{s\to 0^+} P_{ii}(s) \end{aligned}$$

A conservative Markov process stays a certain time in each state visited and does not leave it immediately. We suppose from now on that the Markov process is conservative. In the next section, we will study together the time spent in a state, the jumps between states, and the times between jumps.

3.2 Jump chain

For a conservative Markov process with no absorbing state the *jump times* are defined in terms of the holding times W_t as

$$T_0 = 0, \qquad T_{n+1} = T_n + W_{T_n} \quad \text{for all } n \geq 0,$$

and we define the values of X at the jump times as

$$\widehat{X}_0 = X(0), \qquad \widehat{X}_{n+1} = X(T_{n+1}).$$

$\widehat{X}$ is the chain of the successive states visited by the Markov process X. It is called the *jump chain* of the process X. This chain does not contain all the Markov process' information: it does not record the length of time spent in each state.

In this section, we examine the underlying structure of X defined by the jump chain $\widehat{X}$ and the jump times T_n. We first prove that $\widehat{X}$ is a Markov chain, and $T_{n+1} - T_n$ is exponentially distributed with rate depending on $\widehat{X}_n$.

Theorem 3.2.1 *For any $n \geq 0$, $j \in E$, and $u \in \mathbb{R}^+$,*

$$\begin{aligned} &\mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_0, \ldots, \widehat{X}_n, T_0, \ldots, T_n) \\ &\qquad = \mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_n). \qquad (\star) \end{aligned}$$

Furthermore, if $\widehat{X}_n = i$,

$$\mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_n = i) = \widehat{P}_{ij}\, e^{-\delta_i u}, \qquad (\star\star)$$

where $\widehat{P}$ is a stochastic matrix such that $\widehat{P}_{ii} = 0$ if i is stable, and $\widehat{P}_{ii} = 1$ if i is absorbing.

Proof. Recall that $\mathbb{P}(W_0 > u \mid \widehat{X}_0 = i) = e^{-\delta_i u}$. For any $n \geq 0$,

$$
\begin{aligned}
\mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_0, \dots, \widehat{X}_n, T_0, \dots, T_n) \\
&= \mathbb{P}(X(T_{n+1}) = j, T_{n+1} - T_n > u \mid X(t), 0 \leq t \leq T_n) \\
&= \mathbb{P}(X(T_{n+1}) = j, T_{n+1} - T_n > u \mid X(T_n)) \quad \text{(strong Markov property)} \\
&= \mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_n)
\end{aligned}
$$

which shows $(\star)$. Now, if $X(T_n) = i$,

$$
\begin{aligned}
\mathbb{P}(\widehat{X}_{n+1} = j, T_{n+1} - T_n > u \mid \widehat{X}_n = i) \\
&= \mathbb{P}(X(W_0) = j, T_1 > u \mid X(0) = i) \quad \text{(homogeneity)} \\
&= \mathbb{P}(W_0 > u \mid X(0) = i)\mathbb{P}(X(W_0) = j \mid X(0) = i, W_0 > u) \\
&= e^{-\delta_i u}\mathbb{P}(X(W_0) = j \mid X(t) = i \quad \text{for all } t : 0 \leq t \leq u) \\
&= e^{-\delta_i u}\mathbb{P}(X(u + W_u) = j \mid X(t) = i, 0 \leq t \leq u) \\
&= e^{-\delta_i u}\mathbb{P}(X(u + W_u) = j \mid X(u) = i), \quad \text{(Markov property)} \\
&= e^{-\delta_i u}\mathbb{P}(X(W_0) = j \mid X(0) = i) \quad \text{(homogeneity)} \\
&= e^{-\delta_i u}\mathbb{P}(\widehat{X}_1 = j \mid \widehat{X}_0 = i) \\
&= e^{-\delta_i u}\widehat{P}_{ij},
\end{aligned}
$$

and we have shown $(\star\star)$. □

Corollary 3.2.2 *We have*

(i) The jump chain $\widehat{X}$ is a Markov chain with transition matrix $\widehat{P}$ (putting $u = 0$ in the previous theorem).

(ii) Given $\widehat{X}_0, \widehat{X}_1, \dots$, the intervals $(T_1 - T_0), (T_2 - T_1), \dots$ are independent.

In other words, the times between transitions are conditionally independent of each other given the successive states being visited, and each such sojourn time has an exponential distribution with the parameter dependent on the state being visited. This and the fact that the successive states visited form a Markov chain clarify the structure of a Markov process.

If there are absorbing states, then T_n may potentially be infinite, in which case we define $W_\infty = +\infty$, and the definition of the jump chain becomes

$$
\widehat{X}_0 = X(0), \qquad \widehat{X}_{n+1} = \begin{cases} X(T_{n+1}) & \text{if } T_{n+1} < \infty \\ \widehat{X}_n & \text{if } T_{n+1} = \infty \end{cases} .
$$

If i is an absorbing state for X, then it is an absorbing state for $\widehat{X}$ as well and in that case $\widehat{P}_{ii} = 1$.

Definition 3.2.3 *A conservative Markov process is* regular *if*

$$\lim_{n\to\infty} T_n = +\infty \ \ a.s. \quad \textit{(non explosiveness)}$$

Can we define X in terms of $\widehat{X}$ and $\{T_n\}_{n\geq 0}$? It follows from the definitions that

$$X(t) = \widehat{X}_n \qquad \text{for } t \in [T_n, T_{n+1}),$$

hence X can be defined in terms of $\widehat{X}$ and $\{T_n\}_{n\geq 0}$ provided that for any real t there exists some nonnegative integer n such that $t \in [T_n, T_{n+1})$, or in other words, provided that the Markovian process is regular. From now on, we will assume that this is the case.

3.3 Kolmogorov's equations

There are two equivalent ways to describe how a Markov process X evolves. The first is in terms of the jump chain $\widehat{X}$ and the holding times $\{W_{T_n}\}$. The second is in terms of the semigroup $P(t)$. The objective here is to go from $P_{ij}(t)$ to $\widehat{P}_{ij}$ and δ_i $(i, j \in E)$, and vice versa in a regular Markov process.

3.3.1 $\{\widehat{P}_{ij}, \delta_i\} \longrightarrow P_{ij}(t)$

Theorem 3.3.1

$$P_{ij}(t) = e^{-\delta_i t}\delta_{ij} + \int_0^t \sum_{k\in E, k\neq i} e^{-\delta_i u} \quad \delta_i \widehat{P}_{ik} P_{kj}(t-u)\, du,$$

where δ_{ij} is Kronecker's delta.

Remark 3.3.2 *We can write this theorem in a matrix form, by using the definition of the exponential of a matrix A:*

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!},$$

and setting $\Lambda = diag(\delta_i, i \in E)$*:*

$$P(t) = e^{-\Lambda t} + \int_0^t e^{-\Lambda u} \quad \Lambda \widehat{P} P(t-u) du.$$

Proof. First recall that

$$\begin{aligned} & \mathbb{P}(X(T_1) = k, T_1 \le u \mid X(0) = i) = \widehat{P}_{ik}(1 - e^{-\delta_i u}) \\ \Rightarrow & \frac{d}{du}\mathbb{P}(X(T_1) = k, T_1 \le u \mid X(0) = i) = \widehat{P}_{ik}\, \delta_i\, e^{-\delta_i u}. \end{aligned}$$

So

$$\mathbb{P}(X(T_1) = k, u \le T_1 \le u + du \mid X(0) = i) = \widehat{P}_{ik}\, \delta_i\, e^{-\delta_i u} du.$$

Next, we will use the theorem of total probabilities twice. We first condition on the value of T_1, and then on the value of the next state visited after state i:

$$\begin{aligned} P_{ij}(t) &= \mathbb{P}(X(t) = j \mid X(0) = i) \\ &= \int_0^\infty \mathbb{P}(X(t) = j \mid X(0) = i, T_1 = u) \qquad \underbrace{f_{T_1}(u)}_{= \delta_i e^{-\delta_i u} \text{ because } T_1 \sim exp(\delta_i)} \qquad du, \\ &= \int_0^\infty \delta_i e^{-\delta_i u} \sum_{k \in E} \mathbb{P}(X(t) = j \mid X(0) = i, T_1 = u, X(u) = k) \\ & \quad \cdot \mathbb{P}(X(u) = k \mid X(0) = i, T_1 = u)\, du, \\ &= \int_0^\infty \sum_{k \in E} \delta_i e^{-\delta_i u} \widehat{P}_{ik} \underbrace{\mathbb{P}(X(t) = j \mid X(0) = i, T_1 = u, X(u) = k)}_{(*)} du. \end{aligned}$$

We now separate the integral into two parts, depending on the value of u:

- If $u < t$:

$$\begin{aligned} (*) &= \mathbb{P}(X(t) = j \mid X(x) : 0 \le x \le u,\, X(u) = k), \\ &= \mathbb{P}(X(t) = j \mid X(u) = k), \quad \text{(Markov property)} \\ &= \mathbb{P}(X(t-u) = j \mid X(0) = k), \quad \text{(homogeneity)} \\ &= P_{kj}(t-u). \end{aligned}$$

- If $u \ge t$: $(*) = \delta_{ij}$ (no state change before $u \ge t$.)

Therefore:

$$
\begin{aligned}
P_{ij}(t) &= \int_0^t \sum_{k\neq i} \delta_i e^{-\delta_i u} \widehat{P}_{ik} P_{kj}(t-u)\, du \\
&+ \int_t^\infty \delta_i e^{-\delta_i u} \delta_{ij} \underbrace{\sum_{k\in E} \widehat{P}_{ik}}_{=1 \text{ because } \widehat{P} \text{ is stochastic.}} du,
\end{aligned}
$$

and observing that $\int_t^\infty \delta_i e^{-\delta_i u} du = e^{-\delta_i t}$ completes the proof. □

3.3.2 $P_{ij}(t) \to \widehat{P}_{ij}, \delta_i$

Definition 3.3.3 *The* generator *of the Markov process is the Q-matrix (of size $|E|\times|E|$) such that:*

$$
\begin{cases}
Q_{ii} = -\delta_i, \\
Q_{ij} = \delta_i \widehat{P}_{ij} \quad (i \neq j).
\end{cases}
$$

Theorem 3.3.4

$$
Q = \frac{d}{dt} P(t)\Big|_{t=0^+},
$$

and:

$$
P'(t) = QP(t) = P(t)Q.
$$

Proof.

$$
\begin{aligned}
P_{ij}(t) &= e^{-\delta_i t}\delta_{ij} + \int_0^t \sum_{k\neq i} e^{-\delta_i u} \delta_i \widehat{P}_{ik} P_{kj}(t-u) du, \\
&= e^{-\delta_i t}\left(\delta_{ij} + \int_0^t \sum_{k\neq i} e^{\delta_i s} \delta_i \widehat{P}_{ik} P_{kj}(s) ds\right) \quad (s = t-u), \\
\Rightarrow \frac{d}{dt} P_{ij}(t) &= \underbrace{-\delta_i}_{=Q_{ii}} \underbrace{e^{-\delta_i t}(\dots)}_{=P_{ij}(t)} + e^{-\delta_i t} \sum_{k\neq i} \underbrace{\delta_i \widehat{P}_{ik}}_{=Q_{ik}} P_{kj}(t) e^{\delta_i t}, \\
&= Q_{ii} P_{ij}(t) + \sum_{k\neq i} Q_{ik} P_{kj}(t), \\
&= \sum_{k\in E} Q_{ik} P_{kj}(t) = (Q\, P(t))_{ij},
\end{aligned}
$$

which implies $P'(t) = Q\, P(t)$.

Next,

$$\lim_{t\to 0^+} P'(t) = \lim_{t\to 0^+} Q\,P(t) = Q,$$

where $\lim_{t\to 0^+} P(t) = I$ because X is a conservative Markov process.
Finally, by the Chapman-Kolmogorov Theorem (Chapter 2),

$$\frac{d}{ds}P(t+s) = P(t)\frac{d}{ds}P(s),$$

therefore by taking $\lim_{s\to 0^+}$ we find that $P'(t) = P(t)\,Q$. □

The differential equations $P'(t) = QP(t)$ and $P'(t) = P(t)Q$ are called, respectively, *Kolmogorov's backward and forward equations.*

Corollary 3.3.5 *For any $t \geq 0$, we have $P(t) = e^{Qt}$.*

Remark 3.3.6 *We have seen that $(\boldsymbol{\alpha}, Q)$ entirely determines the Markov process X (where $\boldsymbol{\alpha}$ = the vector of initial probabilities.) We have also shown that a relationship between every pair $P(t) \longleftrightarrow \{\hat{P}, (\delta_i, i \in E)\}$, $P(t) \longleftrightarrow Q$, $Q \longleftrightarrow \{\hat{P}, (\delta_i, i \in E)\}$.*

3.3.3 Interpretation of Q:

If we expand $P(t) = e^{Qt}$ using the definition of the matrix exponential, we find that for t small,

$$\begin{aligned} P(t) &= I + Qt + \frac{(Qt)^2}{2} + \dots \\ &= I + Qt + o(t), \end{aligned}$$

$$\Rightarrow \begin{cases} i \neq j: & P_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i) = Q_{ij}t + o(t), \\ i = j: & P_{ii}(t) = 1 + Q_{ii}t + o(t) = 1 - \delta_i t + o(t). \end{cases}$$

Also note that $e^{-\delta_i t} = 1 - \delta_i t + o(t)$ for small t.

The fact that $P(t)$ is defined by its first derivative at $t = 0$ makes these results interesting. For this reason, Q is called the *generator* of the process X.

The class structure of a continuous-time Markov chain X is simply the discrete-time class structure of its corresponding jump chain $\widehat{X}$.

Theorem 3.3.7 *The following three affirmations are equivalent:*

(i) $\exists t > 0 : P_{ij}(t) > 0$,

(ii) $i \rightsquigarrow j$ *in the graph of* $\widehat{P}$,

(iii) $P_{ij}(t) > 0$ *for all* $t > 0$.

Proof.
$(iii) \Rightarrow (i)$: trivial.

$(i) \Rightarrow (ii)$: trivial as well because we can go from one state to another on the condition that there exists n such that $(\widehat{P}^n)_{ij} > 0$.

$(ii) \Rightarrow (iii)$:
Case 1: we suppose that $\widehat{P}_{ij} > 0$.
The event $\{T_1 \le t, \widehat{X}_1 = j, T_2 - T_1 > t\}$ implies that $\{X(t) = j\}$. We then have

$$\begin{aligned} \mathbb{P}(X(t) = j \mid X(0) = i) &\geq \mathbb{P}(T_1 \le t, \widehat{X}_1 = j, T_2 - T_1 > t \mid X(0) = i), \\ &= (1 - e^{-\delta_i t})\widehat{P}_{ij} e^{-\delta_j t}, \\ &> 0 \quad \text{for all } t \ge 0. \end{aligned}$$

Case 2: If Case 1 does not apply, that is, if $\widehat{P}_{ij} = 0$, the chain goes from i to j through intermediate states, and (ii) can be reformulated:

$$\exists i_1, \ldots, i_{n-1} \text{ such that } \widehat{P}_{i,i_1}, \widehat{P}_{i_1,i_2}, \widehat{P}_{i_2,i_3}, \ldots, \widehat{P}_{i_{n-1},j} > 0.$$

We then have

$$\begin{aligned} \mathbb{P}(X(t) = j \mid X(0) = i) &\geq \mathbb{P}(X(t/n) = i_1, X(2t/n) = i_2, \ldots, X(t) = j \mid X(0) = i), \\ &= \underbrace{P_{i,i_1}(t/n)}_{>0 \text{ by Case 1}} \underbrace{P_{i_1,i_2}(t/n)}_{>0} \ldots \underbrace{P_{i_{n-1},j}(t/n)}_{>0}, \\ &> 0 \quad \text{for all } t > 0. \end{aligned}$$

□

Condition (iii) shows that the situation is simpler than in discrete-time, where it may be possible to reach a state, but only after a certain length of time, and then only periodically.

3.4 States

Let θ_i be the total cumulative time spent in the state i by the Markov process X (this is the equivalent of V_i in discrete time Markov chains). We have

$$\theta_i = \int_0^\infty 1_{\{X(u)=i\}} du.$$

We can decompose θ_i into the following:

$$\theta_i = \sum_{1\leq n\leq N_i} Y_n^{(i)} \quad \text{where} \begin{cases} N_i = \text{ total number of visits of state } i \text{ in } \widehat{X}, \\ Y_n^{(i)} = \text{ time spent in } i \text{ during the } n^{th} \text{ visit}, \\ Y_n^{(i)} \text{ i.i.d., } Y_n^{(i)} \sim exp(\delta_i). \end{cases}$$

We will therefore classify the states into two groups, namely transient states and recurrent states. They have the following properties:

Definition 3.4.1
State i is transient *if (equivalent conditions):*

- $\theta_i < \infty$ *with probability* 1,
- $N_i < \infty$ *with probability* 1,
- *i is transient in $\widehat{X}$,*
- $\mathbb{E}[N_i] < \infty$.

State i is recurrent *if*

- $\theta_i = \infty$ *with probability* 1,
- $N_i = \infty$ *with probability* 1,
- *i is recurrent in $\widehat{X}$,*
- $\mathbb{E}[N_i \mid X(0) = i] = \infty$.

State i is positive recurrent *if it is recurrent and if the expected return time to i is finite; otherwise a recurrent state i is called* null recurrent.

We have

$$\begin{aligned}
\mathbb{E}[\theta_i \mid X(0) = i] &= \mathbb{E}\left[\int_0^\infty 1_{\{X(u)=i|X(0)=i\}} du\right], \\
&= \int_0^\infty \mathbb{E}[1_{\{X(u)=i|X(0)=i\}}]\, du, \qquad \text{(Fubini)} \\
&= \int_0^\infty \mathbb{P}(X(u) = i \mid X(0) = i)\, du \\
&= \int_0^\infty P_{ii}(u)\, du.
\end{aligned}$$

Therefore, i is transient if and only if

$$\begin{aligned}
\mathbb{E}[\theta_i] &= \mathbb{E}\left[\sum_{1 \le n \le N_i} Y_n^{(i)}\right], \\
&= \mathbb{E}[N_i]\mathbb{E}[Y^{(i)}], \qquad \text{(Wald's identity)} \\
&< \infty,
\end{aligned}$$

and we also have:

- i transient $\Leftrightarrow \int_0^\infty P_{ii}(u)\, du < \infty$,
- i recurrent $\Leftrightarrow \int_0^\infty P_{ii}(u)\, du = \infty$.

3.5 Limit behaviour of $P(t)$

We are now interested in calculating

$$\lim_{t\to\infty} P_{ij}(t) = ?$$

Problem: $\lim_{t\to\infty} P(X(t) = j \mid X(0) = i)$ is not necessarily equal to $\lim_{t\to\infty} \mathbb{P}(\widehat{X}_n = j \mid \widehat{X}_0 = i)$.
Counter-example: consider a Markov process X with corresponding jump chain

$$\widehat{P} = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \\ 1/2 & 1/2 & 0 \end{bmatrix},$$

and such that $1/\delta_1 = 1$, $1/\delta_2 = 1$, $1/\delta_3 = 10^{27}$. We the have $\lim_{t\to\infty} \mathbb{P}(X(t) = 3) >> \lim_{t\to\infty} \mathbb{P}(X(t) = i)$, but the stationary distribution of $\widehat{X}$ is $(1/3, 1/3, 1/3)$.

We define a new Markov chain: For all $h > 0$, the discrete skeleton $Z_n := X(nh)$, $n \geq 0$, is a Markov chain with probability transition matrix $P(h)$:

$$\mathbb{P}(Z_{n+1} = j \mid Z_n = i) = \mathbb{P}(X((n+1)h) = j \mid X(nh) = i) = P(h).$$

$\{Z_n\}$ is called the *h-skeleton* of $X(t)$ (of the Markov process). The next theorem shows that recurrence and transience of a state are determined by any discrete-time sampling of X.

Proposition 3.5.1 *State i is transient for X* if and only if *i is transient for the h-skeleton Z, for any $h > 0$.*

Proof. It suffices to show that

$$\int_0^\infty P_{ii}(t)dt < \infty \Leftrightarrow \sum_{n\geq 0} (P(h)^n)_{ii} < \infty.$$

We have

$$\begin{aligned} (P(h)^n)_{ii} &= \mathbb{P}(Z_n = i \mid Z_0 = i) = \mathbb{P}(X(nh) = i \mid X(0) = i) \\ &= P_{ii}(nh). \end{aligned}$$

Therefore, all we need to show is

$$\int_0^\infty P_{ii}(t)dt < \infty \Leftrightarrow \sum_{n\geq 0} P_{ii}(nh) < \infty.$$

Let $t \in [nh, (n+1)h]$ for some $n \in \mathbb{N}$. We have:

$$\begin{aligned} \{X(t) = i, W_t > h\} &\subseteq \{X((n+1)h) = i\}, \quad (\star) \\ \{X(nh) = i, W_{nh} > h\} &\subseteq \{X(t) = i\} \quad (\star\star) \end{aligned}$$

Therefore, using $(\star)$,

$$\begin{aligned} \mathbb{P}(X((n+1)h) = i \mid X(0) = i) &\geq \mathbb{P}(X(t) = i, W_t > h \mid X(0) = i), \\ &= \mathbb{P}(X(t) = i \mid X(0) = i)\mathbb{P}(W_t > h \mid X(t) = i), \\ \Rightarrow P_{ii}((n+1)h) &\geq P_{ii}(t)e^{-\delta_i h}. \end{aligned}$$

Similarly, using $(\star\star)$, we obtain a second inequality, and

$$P_{ii}((n+1)h)e^{\delta_i h} \geq P_{ii}(t) \geq P_{ii}(nh)e^{-\delta_i h}. \quad (\#)$$

Therefore,

$$\begin{aligned}
\int_0^\infty P_{ii}(t)dt &= \sum_{n\geq 0} \int_{nh}^{(n+1)h} P_{ii}(t)dt, \\
&\geq \sum_{n\geq 0} \int_{nh}^{(n+1)h} e^{-\delta_i h} P_{ii}(nh)dt, \qquad \text{(by (\#))} \\
&= \sum_{n\geq 0} \underbrace{he^{-\delta_i h}}_{>0} P_{ii}(nh), \\
&= C_1 \sum_{n\geq 0} P_{ii}(nh)
\end{aligned}$$

for some strictly positive constant C_1. Therefore, we have

$$\int_0^\infty P_{ii}(t)dt < \infty \quad \Rightarrow \quad \sum_{n\geq 0} P_{ii}(nh) < \infty.$$

Similarly, using the other inequality in (#),

$$\begin{aligned}
\int_0^\infty P_{ii}(t)dt &\leq he^{\delta_i h} \sum_{n\geq 1} P_{ii}(nh) \\
&= C_2 \sum_{n\geq 1} P_{ii}(nh),
\end{aligned}$$

for some strictly positive constant C_2, and

$$\sum_{n\geq 0} P_{ii}(nh) < \infty \quad \Rightarrow \quad \int_0^\infty P_{ii}(t)dt < \infty.$$

□

The limiting behaviour of the transition function $P(t)$ as $t \to \infty$ is just as in the case of discrete-time Markov chains, except that it is made simpler by the disappearance of periodicity.

Theorem 3.5.2 (Convergence to equilibrium) *Let X be an irreducible conservative Markov process. Then, for any $i, j \in E$,*

$$\lim_{t\to\infty} \mathbb{P}(X(t) = j \mid X(0) = i) = \begin{cases} 0 & \text{if } j \text{ is transient or null recurrent,} \\ \pi_j & \text{if } j \text{ is positive recurrent} \end{cases}$$

Proof.
Case 1: $t \to \infty$ with $t = nh$.

$$\begin{aligned}
\lim_{t\to\infty,\, t=nh} \mathbb{P}(X(t) = j \mid X(0) = i) &= \lim_{n\to\infty} \mathbb{P}(Z_n = j \mid Z_0 = i) \\
&= \begin{cases} 0 & \text{if } j \text{ is transient or null recurrent,} \\ \alpha_j(h) & \text{if } j \text{ is positive recurrent,} \end{cases}
\end{aligned}$$

by the results on discrete-time Markov chains.
We now want to show that the limit $\alpha_j(h)$ actually does not depend on the choice of h.

Case 2: $t \to \infty$ in an arbitrary way.
Let $\varepsilon > 0$, and let us fix an arbitrary $h > 0$ such that $e^{-\delta_i h} > 1 - \varepsilon/2$. Therefore,

$$e^{-\delta_i s} > 1 - \varepsilon/2 \quad \text{for all } s \in]0, h[.$$

By Case 1, there exists N such that for all $n \geq N$,

$$| P_{ij}(nh) - \alpha_j(h) | \leq \varepsilon/2.$$

Let $T = Nh$, and fix $t > T$. We need to show that $| P_{ij}(t) - \alpha_j(h) | < \varepsilon$. There exists $v \geq N$ such that $vh \leq t \leq (v+1)h$. Then,

$$| P_{ij}(t) - \alpha_j(h) | \leq \underbrace{| P_{ij}(t) - P_{ij}(vh) |}_{\leq \varepsilon/2??} + \underbrace{| P_{ij}(vh) - \alpha_j(h) |}_{\leq \varepsilon/2}.$$

Take $t = s + vh$. We have

$$\begin{aligned} P_{ij}(t) &= P_{ij}(s + vh), \\ &= (P(s)P(vh))_{ij}, \quad \text{(Chapman-Kolmogorov)} \\ &= \sum_{k \in E} P_{ik}(s) P_{kj}(vh), \end{aligned}$$

which implies

$$\begin{aligned} | P_{ij}(t) - P_{ij}(vh) | &= \left| \sum_{\substack{k \in E \\ k \neq i}} P_{ik}(s) P_{kj}(vh) - (1 - P_{ii}(s)) P_{ij}(vh) \right|, \\ &\leq \max \left\{ \sum_{\substack{k \in E \\ k \neq i}} P_{ik}(s) P_{kj}(vh), (1 - P_{ii}(s)) P_{ij}(vh) \right\}, \\ &\leq \max \left\{ \sum_{\substack{k \in \mathbb{E} \\ k \neq i}} P_{ik}(s), 1 - P_{ii}(s) \right\}, \\ &= 1 - P_{ii}(s), \quad (P(s) \text{ is a stochastic matrix}) \\ &< \varepsilon/2, \end{aligned}$$

because $P_{ii}(s) > e^{-\delta_i s} > 1 - \varepsilon/2$.
We have shown that, for the fixed value of h, $P_{ij}(t) \to \alpha_j(h)$ as $t \to \infty$. This implies that for any sequence $\{t_k\}$ with $t_k \to \infty$ as $k \to \infty$, $P_{ij}(t_k) \to \alpha_j(h)$. By taking $t_k = kh_2$ for any arbitrary h_2, and using Case 1, we have that $\alpha_j(h) = \alpha_j(h_2)$. This proves that the limit $\alpha_j(h)$ actually does not depend on h, and we write $\alpha_j(h) = \pi_j$. □

Remark 3.5.3 *If X is reducible and j is positive recurrent, then*

$$\lim_{t\to\infty} \mathbb{P}(X(t) = j \mid X(0) = i) = f_{ij}\,\pi_j,$$

where f_{ij} is the probability that, starting from i, state j is visited after a finite time, that is,

$$f_{ij} = \mathbb{P}(\tau_j < \infty \mid X(0) = i),$$

where $\tau_j = \inf\{t \geq 0 : X(t) = j\}$, and $\pi_j = \lim_{t\to\infty} \mathbb{P}(X(t) = j \mid X(0) = j)$.

The next theorem characterizes the limiting distribution $\boldsymbol{\pi}$ in the positive recurrent case as the solution of some systems of linear equations.

Theorem 3.5.4 *The following three assertions are equivalent:*

(i) The states of a conservative irreducible Markov process are positive recurrent

(ii) There exists $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}P(s) = \boldsymbol{\pi}$ for all s,

(iii) There exists $\boldsymbol{\pi}$ such that $\boldsymbol{\pi}Q = \mathbf{0}$,

In (ii) and (iii), $\boldsymbol{\pi}$ is such that $\pi_j = \lim_{t\to\infty} \mathbb{P}(X(t) = j \mid X(0) = j)$, with $\boldsymbol{\pi} \geq \mathbf{0}$, and $\boldsymbol{\pi}\mathbf{1} = 1$.

Proof.
(i) $\Leftrightarrow$ (ii). We have

$$\lim_{t\to\infty} \mathbb{P}(X(t) = j \mid X(0) = i) = \pi_j \quad \text{if} \quad \lim_{n\to\infty} \mathbb{P}(X(nh) = j \mid X(0) = i) = \pi_j \quad \text{for all } h > 0.$$

For the h-skeleton, the states are positive recurrent if and only if there exists $\boldsymbol{x}(h)$ such that

$$\begin{cases} \boldsymbol{x}(h)P(h) = \boldsymbol{x}(h) \\ \boldsymbol{x}(h)\mathbf{1} = 1 \end{cases}$$

In this case, $x_j(h) = \lim_{n\to\infty} \mathbb{P}(X(nh) = j \mid X(0) = i) = \pi_j$, as we have shown in the previous theorem that the limit does not depend on h.

(ii) $\Leftrightarrow$ (iii). If the number of states is finite,

$$\begin{array}{rrcll} & \boldsymbol{\pi}P(s) & = & \boldsymbol{\pi} & \text{for all } s, \\ \Leftrightarrow & \boldsymbol{\pi}P'(s) & = & \mathbf{0} & \text{(since the sum is finite)}\ , \\ \Leftrightarrow & \boldsymbol{\pi}P(s)Q & = & \mathbf{0}, & \\ \Leftrightarrow & \boldsymbol{\pi}Q & = & \mathbf{0}. & \end{array}$$

For an arbitrary number of states, first observe that $\boldsymbol{\pi}Q = \mathbf{0}$ implies that $\boldsymbol{\pi}P(s) = \boldsymbol{\pi}e^{Qs} = \boldsymbol{\pi}\sum_{n\geq 0}(Qs)^n/n! = \boldsymbol{\pi}$ for all s, so we have (iii) $\Rightarrow$ (ii). We remains to show (ii) $\Rightarrow$ (iii).

We look at the Markov jump chain, whose transition matrix is $\widehat{P}$. We know that its states are positive recurrent if and only if there exists $\widehat{\boldsymbol{\pi}}$ such that $\widehat{\boldsymbol{\pi}}\widehat{P} = \widehat{\boldsymbol{\pi}}$. However, recall that

$$\begin{cases} Q_{ij} = \delta_i \widehat{P}_{ij} & (i \neq j), \\ Q_{ii} = -\delta_i. \end{cases}$$

Letting Λ =diag(δ_i), we can then write

$$Q = \Lambda(\widehat{P} - I) \quad \Leftrightarrow \quad \widehat{P} = I + \Lambda^{-1}Q$$

and $\widehat{P}_{ii} = 0$. Therefore

$$\begin{array}{rrcl} & \widehat{\boldsymbol{\pi}}\widehat{P} & = & \widehat{\boldsymbol{\pi}}, \\ \Leftrightarrow & \widehat{\boldsymbol{\pi}} + \widehat{\boldsymbol{\pi}}\Lambda^{-1}Q & = & \widehat{\boldsymbol{\pi}}, \\ \Leftrightarrow & \widehat{\boldsymbol{\pi}}\Lambda^{-1}Q & = & \mathbf{0}. \end{array}$$

In order to have (iii), we need to show that $\widehat{\boldsymbol{\pi}}\Lambda^{-1} = \boldsymbol{\pi}$.

Let $t - v$ be the time of the last state change before time t. We have

$$P_{ij}(t) = e^{-\delta_i t}\delta_{ij} + \int_0^t \sum_k P_{ik}(t-v)\delta_k \widehat{P}_{kj} e^{-\delta_j v} dv,$$

or in matrix form,

$$P(t) = e^{-\Lambda t} + \int_0^t P(t-v)\Lambda\widehat{P}e^{-\Lambda v} dv.$$

But we know that $\boldsymbol{\pi}P(t) = \boldsymbol{\pi}$ for all t, hence

$$\begin{array}{rrcll} & \boldsymbol{\pi}e^{-\Lambda t} + \int_0^t \boldsymbol{\pi}P(t-v)\Lambda\widehat{P}e^{-\Lambda v}dv & = & \boldsymbol{\pi}, & \\ \Leftrightarrow & \boldsymbol{\pi}e^{-\Lambda t} + \int_0^t \boldsymbol{\pi}\Lambda\widehat{P}e^{-\Lambda v}dv & = & \boldsymbol{\pi} & (\boldsymbol{\pi}P(t-v) = \boldsymbol{\pi}), \\ \Leftrightarrow & \boldsymbol{\pi}e^{-\Lambda t} + \boldsymbol{\pi}\Lambda\widehat{P}\int_0^t e^{-\Lambda v}dv & = & \boldsymbol{\pi}, & \\ \Leftrightarrow & \boldsymbol{\pi}e^{-\Lambda t} - \boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1}[e^{-\Lambda v}]_0^t & = & \boldsymbol{\pi}, & \\ \Leftrightarrow & \boldsymbol{\pi}e^{-\Lambda t} - \boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1}e^{-\Lambda t} + \boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1} & = & \boldsymbol{\pi} & \text{for all } t. \end{array}$$

We take the limit of each term as $t \to \infty$ and observe that $\lim_{t\to\infty} e^{-\Lambda t} = 0$, $\lim_{t\to\infty} \boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1}e^{-\Lambda t} = 0$, and $\boldsymbol{\pi}, \boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1}$ do not dependent on t. We then obtain

$$\boldsymbol{\pi}\Lambda\widehat{P}\Lambda^{-1} = \boldsymbol{\pi} \quad \Leftrightarrow \boldsymbol{\pi}\Lambda\widehat{P} = \boldsymbol{\pi}\Lambda.$$

On the other hand, we know that $\widehat{\boldsymbol{\pi}}$ is the unique solution to $\widehat{\boldsymbol{\pi}}\widehat{P} = \widehat{\boldsymbol{\pi}}$ to one multiplying constant, therefore

$$\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}\Lambda$$

to one multiplying constant, and we have

$$\widehat{\boldsymbol{\pi}}\Lambda^{-1}Q = \mathbf{0} \quad \Leftrightarrow \quad \boldsymbol{\pi}\Lambda\Lambda^{-1}Q = \mathbf{0} \quad \Leftrightarrow \quad \boldsymbol{\pi}Q = \mathbf{0}.$$

This shows $(ii) \Rightarrow (iii)$.

□

Remark 3.5.5

- *In the last proof, we showed that if* $\widehat{\boldsymbol{\pi}}$ *is the stationary distribution of the jump chain* $\widehat{X}$ *(with positive recurrent states, irreducible), then*

$$\boldsymbol{\pi} = \widehat{\boldsymbol{\pi}}\Lambda^{-1}/(\widehat{\boldsymbol{\pi}}\Lambda^{-1}\mathbf{1}),$$

where $\Lambda = diag(\delta_i) = diag(-Q_{ii})$.

- *In the last theorem, we supposed that all the states were stable (if a state is absorbing in an irreducible process, that state will be visited and will never be left).*

The complete description of limiting behaviour for irreducible chains in continuous-time is provided by the following result.

Theorem 3.5.6 *Let* X *be an irreducible Markov process with arbitrary initial distribution and generator* Q*. Then*

$$\mathbb{P}(X(t) = j) \to 1/(\delta_j m_j) \quad \text{as } t \to \infty \text{ for all } j \in E,$$

where $m_j = \mathbb{E}(\tau_j \mid X(0) = j)$ *is the expected return time to state* j.

3.6 Example: The M/M/1 queue

The simplest queueing model has exponential interarrival times with mean $1/\lambda$, exponential service times with mean $1/\mu$ and a single server. Customers are served in order of arrival. Let $X(t)$ denote the number of customers in the system at time t (including the one being served, if there is one). $\{X(t)\}$ forms a Markovian process with generator

$$Q = \begin{bmatrix} -\lambda & \lambda & 0 & 0 & \cdots \\ \mu & -(\mu+\lambda) & \lambda & 0 & \cdots \\ 0 & \mu & -(\mu+\lambda) & \lambda & \cdots \\ 0 & 0 & \mu & -(\mu+\lambda) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

So in this example, the parameters of the exponential sojourn time distributions are $\delta_0 = \lambda$, $\delta_i = \lambda + \mu$ for $i \geq 1$. The Markov process is irreducible and there is no absorbing state. The probability transition matrix of the corresponding jump chain is

$$\widehat{P} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots \\ \mu/(\mu+\lambda) & 0 & \lambda/(\mu+\lambda) & 0 & \cdots \\ 0 & \mu/(\mu+\lambda) & 0 & \lambda/(\mu+\lambda) & \cdots \\ 0 & 0 & \mu/(\mu+\lambda) & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix},$$

which corresponds to a random walk on $\mathbb{Z}^+$.

What is the distribution of $X(t)$ given that the chain start in some initial distribution $\boldsymbol{\alpha}$ ($\alpha_i = \mathbb{P}(X(0) = i)$)? It is given by $\boldsymbol{p}(t) := \boldsymbol{\alpha} P(t)$ where $p_n(t) = \mathbb{P}(X(t) = n)$, $n \geq 0$. We write the forward Kolmogorov equation, $P'(t) = P(t)\,Q$ which, after pre-multiplication by $\boldsymbol{\alpha}$, gives $\boldsymbol{p}'(t) = \boldsymbol{p}(t)\,Q$. This is the matrix expression for the infinite system of differential equations

$$p_0'(t) = -\lambda p_0(t) + \mu p_1(t), \tag{3.1}$$
$$p_n'(t) = \lambda p_{n-1}(t) - (\lambda+\mu) p_n(t) + \mu p_{n+1}(t), \quad n = 1, 2, \ldots \tag{3.2}$$

It is difficult to solve these differential equations. An explicit solution for the probabilities $p_n(t)$ can be written but it involves an infinite sum of modified Bessel functions. So already one of the simplest interesting queueing models leads to a difficult expression for the time-dependent behaviour of its state probabilities. For more general systems we can only expect more complexity. However, after some transition period the system will become stable, and the limiting or equilibrium behaviour of this system is much easier to analyse. Of course the state will permanently change, but the probabilities of various numbers of customers in the system will be constant.

We have seen that a stationary probability vector $\boldsymbol{\pi}$ satisfies $\boldsymbol{\pi} Q = \mathbf{0}$ and $\boldsymbol{\pi}\mathbf{1} = 1$. This gives

$$-\lambda\pi_0 + \mu\pi_1 = 0 \tag{3.3}$$
$$\lambda\pi_{n-1} - (\lambda+\mu)\pi_n + \mu\pi_{n+1} = 0, \quad n = 1, 2, \ldots \tag{3.4}$$

Note that this is exactly the system we obtain if we let $t \to \infty$ in the forward Kolmogorov equations. From this system, we obtain that

$$\pi_n = (\lambda/\mu)^n\,\pi_0, \quad n = 0, 1, 2, \ldots$$

so a solution such that $\boldsymbol{\pi}\mathbf{1} = 1$ exists if and only if the states are positive recurrent, and it happens if and only if $\sum_{n\geq 0}(\lambda/\mu)^n < \infty$, if and only if $\lambda < \mu$. In that case, the stationary distribution is given by

$$\pi_n = (\lambda/\mu)^n (1 - \lambda/\mu), \quad n = 0, 1, 2 \ldots .$$

You will study the M/M/1 queue in more detail in Serie 9, including criteria for null recurrence and transience.

Chapter 4

Renewal processes

In the stochastic processes we have studied so far, an important property has been the existence of times, usually random, from which onward the future of the process is a probabilistic replica of the original process. In Markov chains and Markov processes, for example, if the initial state is i, then the times of successive entrances to that state i plays this role; and this fact in turn enables us to obtain many of the limiting results we listed before. This "regeneration" property may hold in much more general situations, and when it holds, surprisingly sharp results can be obtained by the methods we are going to develop in this chapter.

4.1 Definition

Let $F(\cdot)$ be the distribution function of a non-negative random variable (that is, $F(x) = 0$ if $x < 0$). Let $W_1, W_2, \ldots$ be i.i.d. random variables $\sim F(\cdot)$.

Definition 4.1.1 *The* renewal process *associated with $F(\cdot)$ is the process $\{S_n : n \in \mathbb{N}\}$ such that:*

- $S_0 = 0$ *almost surely,*
- $S_{n+1} = S_n + W_{n+1}$ *for all* $n = 0, 1, 2, \ldots$.

The S_n are called renewal times.

Example 4.1.2 *Consider an item installed at time $S_0 = 0$. When it fails it is replaced by an identical item; when that item fails, it in turn is replaced by a new item, and so on. Suppose the lifetime of the successive items are $U_1, U_2, \ldots$ and the replacements take $V_1, V_2, \ldots$ units of time. Hence, the successive items start working at times $S_0 = 0$, $S_1 = U_1 + V_1$, $S_2 = U_2 + V_2$, and so on. It is reasonable to assume the U_i to be i.i.d. and the same for the V_i, and the U_i and V_i to be independent. Then $W_i = U_i + V_i$ are i.i.d. and the S_n form a renewal process.*

Definition 4.1.3 *Let $F(\cdot)$ and $G(\cdot)$ be the distribution functions of two non-negative random variables. The* convolution *of $F(\cdot)$ and $G(\cdot)$ is defined as*

$$(F * G)(x) = \int_0^x G(x-u)dF(u).$$

We know that if X and Y are independent random variables with $X \sim F(\cdot)$, and $Y \sim G(\cdot)$, then $X + Y \sim (F * G)(\cdot)$ (exercise), that is,

$$\mathbb{P}(X + Y \leq x) = (F * G)(x).$$

Note that the convolution operation is commutative. Here, the renewal times are such that:

$$\begin{array}{lll} S_0 & = & 0 \\ S_1 & = & W_1 \qquad\qquad\quad \sim F(\cdot) \\ S_2 & = & W_1 + W_2 \qquad\quad \sim (F * F)(\cdot) \\ S_3 & = & W_1 + W_2 + W_3 \quad \sim (F * F * F)(\cdot) \end{array}$$

Definition 4.1.4

$$\begin{aligned} F^{(0)}(x) &= \mathbb{1}_{\{x \geq 0\}} \\ &= \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0, \end{cases} \\ F^{(n+1)}(x) &= (F * F^{(n)})(x) \quad \text{for all } n \geq 0. \end{aligned}$$

These $F^{(n)}$ (nth fold convolution of F with itself) are the distribution functions of the renewal times:

- $F^{(0)}(\cdot)$ = distribution function of the random variable = 0 a.s.,
- $F^{(1)}(\cdot) = F(\cdot)$ = distribution function of $W_1 = S_1$,
- $F^{(2)}(\cdot) = (F * F)(\cdot)$ = distribution function of $W_1 + W_2 = S_2$, and so on.

Therefore

$$\mathbb{P}(S_n \leq x) = F^{(n)}(x).$$

Definition 4.1.5 *The number of renewals $N(t)$ in the interval $[0,t]$ is defined as*

$$N(t) = \inf\{n : S_n > t\}.$$

Remark 4.1.6 $N(0) \geq 1$ *($N(0) = 1$, except if there are several simultaneous renewals in $t = 0$.)*

We have

$$\{N(t) \leq k\} = \{S_k > t\}.$$

Indeed, the event $\{N(t) \leq k\}$ is equivalent to the event that before time t, we have at most k renewals, that is, at most $S_0, S_1, \ldots, S_{k-1}$, which is $\{S_k > t\}$.
We therefore have

$$\mathbb{P}(N(t) \leq k) = \mathbb{P}(S_k > t) = 1 - F^{(k)}(t),$$

and by calculating the $F^{(n)}(\cdot)$, we know the distribution of S_n and $N(t)$.

Definition 4.1.7 *The* renewal function *of the renewal process is the expected number of renewals in the interval $[0,t]$:*

$$\begin{aligned} R(t) &= \mathbb{E}[N(t)], \\ &= \sum_{k\geq 0} F^{(k)}(t), \end{aligned}$$

The last equality is true because

$$\begin{aligned} \mathbb{E}[N(t)] &= \sum_{k\geq 0} \mathbb{P}(N(t) > k), \\ &= \sum_{k\geq 0} 1 - \underbrace{\mathbb{P}(N(t) \leq k)}_{=1-F^{(k)}(t)}, \\ &= \sum_{k\geq 0} F^{(k)}(t). \end{aligned}$$

The renewal function plays an important role in the study of renewal processes. Before showing this, we will first examine different possible behaviours of the renewal process depending on the behaviour of $F(\cdot)$.

Behaviour in function of $F(0)$.

- *If* $F(0) = 0$*:* then $\mathbb{P}(W = 0) = 0$, therefore the intervals of time between two renewals cannot be of zero length:

- *If* $0 < F(0) < 1$*:* then $0 < \mathbb{P}(W = 0) < 1$, therefore we can have times where several renewals take place:

- *If* $F(0) = 1$*:* then all the renewal intervals have length zero, therefore $\mathbb{P}(S_n = 0) = 1$ for all n. This case is not interesting, since an infinity number of renewals take place at time 0.

Assumption (A): From now on, we will always suppose that $F(0) < 1$, that is, we cannot have infinitely many renewals at one time.

Behaviour in function of $F(\infty)$.

$$F(\infty) = \lim_{x \to \infty} F(x).$$

- *If* $F(\infty) = 1$*:* then $\mathbb{P}(W_i < \infty) = 1$ for all i, therefore all the renewal intervals are of finite length.

- *If* $F(\infty) < 1$*:* then $\mathbb{P}(W_i = \infty) > 0$ for all i, therefore $W_k = +\infty$ for some k, so that the next renewal comes after an infinite time. Therefore, in this case, the renewal process is "dead", since no more renewals will ever occur.

Definition 4.1.8 *We say that a renewal process is*

- recurrent *if* $F(\infty) = 1$,
- transient *if* $F(\infty) < 1$.

Let us return to the renewal function $R(t)$. We will show that, under Assumption (A), the renewal function $R(t)$ is always finite for $0 \leq t < \infty$, that is, that we cannot have an infinite number of renewals in a finite time:

Proposition 4.1.9 *If* $F(0) < 1$, *then* $R(t) < \infty$ *for all finite* $t \geq 0$.

Proof. We will construct a (finite) bound for $R(t)$.
Since $F(0) < 1$, and F is right continuous, there is some $b > 0$ such that $F(b) < 1$ (that is, $\mathbb{P}(W > b) > 0$). We have

$$\begin{cases} S_0 = 0, \\ S_n = S_{n-1} + W_n \quad \text{for all } n \geq 1. \end{cases}$$

We construct a new renewal process by taking:

$$\widetilde{W}_n = \begin{cases} 0 & \text{if } W_n \leq b, \\ b & \text{if } W_n > b. \end{cases}$$

We then have a renewal process

$$\begin{cases} \widetilde{S}_0 = 0, \\ \widetilde{S}_n = \widetilde{S}_{n-1} + \widetilde{W}_n \quad \text{for all } n \geq 1, \end{cases}$$

such that for all n, t,

$$\begin{cases} S_n \geq \widetilde{S}_n, \\ N(t) \leq \widetilde{N}(t). \end{cases}$$

This implies that

$$R(t) = \mathbb{E}[N(t)] \leq \mathbb{E}[\widetilde{N}(t)],$$

and $\mathbb{E}[\widetilde{N}(t)] < \infty$ because

$$\mathbb{E}[\widetilde{N}(t)] \leq \left(\frac{t}{b} + 1\right) \mathbb{E}[Y]$$

where Y is the number of instantaneous renewals ($\widetilde{W}_i = 0$) between two non-instantaneous renewals ($\widetilde{W}_i = b$), so $Y \sim Geom(F(b))$ and $\mathbb{E}[Y] = 1/(1 - F(b)) < \infty$ because $F(b) < 1$. □

Example 4.1.10 *Suppose $F(t) = 1 - e^{-\lambda t}$. Then the renewal process is actually a Poisson process with an arrival at time 0. Hence, $R(t) = 1 + \lambda t$ for all $t \geq 0$.*

Proposition 4.1.11 *Under Assumption (A), $R(t)$ is right continuous, and non-decreasing. Furthermore, if $f : \mathbb{R}^+ \to \mathbb{R}^+$ is bounded and null outside a finite interval, we have*

$$\int_0^\infty f(u)dR(u) = \mathbb{E}\left[\sum_{n=0}^{\infty} f(S_n)\right] < \infty.$$

Proof. The function $N(t)$ is increasing and right-continuous (by definition), therefore $R(t)$ in particular is also increasing.
In order to show that $R(t)$ is right-continuous, let $(t_n) \searrow t$, $n \geq 0$. We have

$$\begin{aligned} N(t_n) &\searrow N(t) \quad (N(t) \text{ right-continuous}) \\ N(t) &\leq N(t_0) \quad \text{for all } n \geq 0 \quad (N(t) \text{ increasing}). \end{aligned}$$

Since $\mathbb{E}[N(t_0)] = R(t_0) < \infty$, we have, by the theorem of dominated convergence,

$$\begin{aligned} \lim_{n\to\infty} R(t_n) &= \lim_{n\to\infty} \mathbb{E}[N(t_n)], \\ &= \mathbb{E}[N(t)], \\ &= R(t). \end{aligned}$$

Therefore $R(t)$ is right-continuous.
Next, if f is bounded by C and vanishes outside $[0,t]$,then

$$\sum_{n\geq 0} f(S_n) \leq CN(t), \quad \text{and} \quad \mathbb{E}\left[\sum_{n=0}^{\infty} f(S_n)\right] \leq CR(t) < \infty.$$

To prove $\int_0^\infty f(u)dR(u) = \mathbb{E}\left[\sum_{n=0}^\infty f(S_n)\right]$, we observe the following:

- For $f = \mathbb{1}_{\{(s,t]\}}$, we have:

$$\begin{aligned} \mathbb{E}[\sum_{n\geq 0} f(S_n)] &= \mathbb{E}[N(t) - N(s)], \\ &= R(t) - R(s), \\ &= \int_0^\infty \mathbb{1}_{\{(s,t]\}}(u)dR(u) \quad \to \text{ ok.} \end{aligned}$$

- It is therefore also true for all step functions (or *simple* functions, that is, functions that are linear combinations of finitely many indicator functions) since the expectation of a sum is the sum of the expectations, and the integral of a sum is the sum of the integrals.

- It is therefore also true for an arbitrary $f \geq 0$, taking f as the limit of an increasing sequence (f_n) of step functions, then applying the monotone convergence theorem.

□

Remark 4.1.12 *From here, we have for all renewal processes:*

$$R(0) \geq 1,$$

therefore, when we calculate $\int_0^\infty f(u)dR(u)$*, we must not forget the term* $R(0)f(0)$ *which is to be added to the integral over* $(0,\infty)$*.*

Example 4.1.13 *Let X be a discrete-time Markov chain, and let j be a fixed state. Let $S_1, S_2, \ldots$ be the successive step numbers at which state j is visited. If $X_0 = j$, then the times S_1, $S_2 - S_1$, $S_3 - S_2$, $\ldots$, between returns in j are independent and identically distributed, hence $\{S_n\}$ forms a renewal process (this fact will be used later to prove an important result about Markov chain using the theory of renewal processes). Consider the renewal function $R(\cdot)$ for this process. Since the number of visits to j during $[0,t]$ is $\sum_{n\leq t} \mathbb{1}_{\{X_n=j\}}$, we have*

$$R(t) = \mathbb{E}_j\left[\sum_{n\leq t} \mathbb{1}_{\{X_n=j\}}\right] = \sum_{n\leq t}(P^n)_{jj},$$

where P^n is the n-step probability transition matrix. Note that in this case, all the S_n are integer-valued, so $R(\cdot)$ is a step function whose jumps are restricted to the times $0, 1, 2, \ldots$.

Remark 4.1.14 *We have previously defined the convolution $F * G$ between two distribution functions F and G. We now define the convolution $F * g$ where F is a distribution function, and g is any non-negative function defined on $\mathbb{R}^+$ which is bounded over any finite interval:*

$$(F * g)(t) = \int_0^t g(t-u)\, dF(u).$$

(Lebesgue-Stieltje integral). Note that if $F^{(0)}$ is the distribution of a degenerate random variable equal to 0 a.s., that is, if $F^{(0)}(x) = \mathbb{1}_{\{x\geq 0\}}$ is the Heaviside step function, then $(F^{(0)} * g)(t) = g(t)$ for all t. This comes from the fact that $dF^{(0)}(u) = \delta(u)du$ where $\delta(\cdot)$ is the Dirac delta function. It is also because, since $F^{(0)}(\cdot)$ has point mass at zero,

$$\int_0^t g(t-u)\, dF^{(0)}(u) = g(t)F^{(0)}(0) + \int_{(0,t]} g(t-u)\, dF^{(0)}(u) = g(t).$$

4.2 Renewal equations

A *renewal equation* is an equation of the form:

$$f(t) = g(t) + (F * f)(t), \quad (\star)$$

where

- $f(t)$ is the unknown function,
- $F(t)$ is the distribution function of a non-negative random variable,
- $g(t)$ is a given function, bounded on all finite intervals, and such that $g(t) = 0$ if $t < 0$.

This equation will arise when studying properties of *regenerative processes*. These are stochastic processes Z which are such that every time a certain phenomenon occurs, the future of Z after that time becomes a probabilistic replica of the future after time 0. Such times (usually random) are called *regeneration times* of Z. For example, if Z is a Markov chain and if j is a fixed state, then every time at which state j is entered is a regeneration time for Z starting at j.

Theorem 4.2.1 *The renewal equation* $(\star)$ *has the unique solution given by*

$$f(t) = (R * g)(t),$$

where $R(t)$ *is the renewal function of the renewal process associated with* $F(\cdot)$*,*

$$R(t) = \sum_{n\geq 0} F^{(n)}(t).$$

This theorem will be very useful in future: often, the computation of quantities related to renewal processes will reduce to the resolution of an equation of type $(\star)$. This theorem provides us with the unique solution.

Proof. We first show that $R * g$ is a solution:

$$\begin{aligned}
(R * g)(t) &= \sum_{n\geq 0}(F^{(n)} * g)(t), \\
&= g(t) + \sum_{n\geq 1}(F * F^{(n-1)} * g)(t) \quad (F^{(0)}(t) = \mathbb{1}_{\{t\geq 0\}}), \\
&= g(t) + (F * \sum_{n\geq 0}(F^{(n)} * g))(t), \\
&= g(t) + (F * (R * g))(t).
\end{aligned}$$

We now show the uniqueness of the solution: if $f_1(t)$ and $f_2(t)$ are both solutions of $(\star)$, then

$$f_1(t) = g(t) + (F * f_1)(t)$$
$$f_2(t) = g(t) + (F * f_2)(t).$$

Take $h(t) = f_1(t) - f_2(t)$. We then have:

$$\begin{aligned} h(t) &= (F * h)(t), \\ &= (F * (F * h))(t), \\ &= \dots \\ \Rightarrow h(t) &= (F^{(n)} * h)(t) \quad \text{for all } n, \\ \Rightarrow h(t) &= \lim_{n\to\infty} (F^{(n)} * h)(t). \end{aligned}$$

Since $R(t) < \infty$ for all t, the general term of the series $R(t) = \sum_{n\geq 0} F^{(n)}(t)$ goes to 0 for all t therefore

$$\begin{aligned} \lim_{n\to\infty} (F^{(n)} * h)(t) &= 0, \\ \Rightarrow h(t) &= 0 \quad \text{for all } t, \end{aligned}$$

which shows uniqueness. □

4.3 Transient renewal processes

In the transient case ($F(\infty) < 1$), $W_n = \infty$ for some $n \geq 0$ with probability one, which stops the arrival of renewals times.

Example 4.3.1 *The renewal process formed by the times of successive entrances to a fixed state j of a Markov process is transient if and only if j is transient.*

The total number of renewals in $[0, \infty)$ then follows a geometric distribution with success probability $1 - F(\infty)$, and in particular,

Theorem 4.3.2 *In the case of transient renewal processes,*

$$\lim_{t\to\infty} R(t) = \frac{1}{1 - F(\infty)}$$

Proof.

$$\begin{aligned} \lim_{t\to\infty} F^{(k)}(t) &= \lim_{t\to\infty} \mathbb{P}(S_k \leq t), \\ &= \mathbb{P}(S_k < \infty), \\ &= \mathbb{P}(W_1 + \cdots + W_k < \infty), \\ &= \mathbb{P}(W_1 < \infty, \dots, W_k < \infty), \\ &= \mathbb{P}(W_1 < \infty)\mathbb{P}(W_2 < \infty)\dots\mathbb{P}(W_k < \infty), \\ &= F(\infty)^k. \end{aligned}$$

Therefore,

$$\begin{aligned}
\lim_{t\to\infty} R(t) &= \lim_{t\to\infty} \sum_{k\geq 0} F^{(k)}(t), \\
&= \sum_{k\geq 0} \lim_{t\to\infty} F^{(k)}(t), \quad (\star) \\
&= \sum_{k\geq 0} F(\infty)^k, \\
&= \frac{1}{1-F(\infty)} \quad \text{(geometric series)}.
\end{aligned}$$

($\star$) By dominated convergence, since $F^{(k)}(t) \leq F(\infty)^k$ for all t and $\sum_{k\geq 0} F(\infty)^k < \infty$. □

Theorem 4.3.3 *If $F(\infty) < 1$, then*

$$\lim_{t\to\infty} (R * g)(t) = R(\infty)\, g(\infty)$$

provided that $g(\infty) = \lim_{t\to\infty} g(t)$ exists.

Definition 4.3.4 *The* lifetime *of a transient renewal process is the time of the last renewal:*

$$L = max\{S_n : S_n < \infty\}$$

In the next theorem, we will calculate the distribution of L. Combined with the previous theorem, the next theorem also shows that $\mathbb{P}(L < \infty) = 1$. The proof of the theorem provides a good example of the "renewal-theoretic reasoning" that will be used a couple of times.

Theorem 4.3.5

$$\mathbb{P}(L \leq x) = (1 - F(\infty))R(x).$$

Proof. Let us write $f(x) = \mathbb{P}(L > x)$. We will show that $f(\cdot)$ satisfies a renewal equation, and we will then obtain its expression using the theorem in the previous section.
In order to do this, as we often do in the context of renewal processes, we will condition on the time S_1 of the first renewal:

$$\begin{aligned} f(x) &= \mathbb{P}(L > x), \\ &= \int_0^\infty \mathbb{P}(L > x \mid S_1 = u)\, dF(u). \end{aligned}$$

We then distinguish three cases:

- Case 1: $W_1 = +\infty \quad (\to u = \infty)$.

  In this case,

$$\begin{aligned} L = S_0 &= 0, \\ \Rightarrow \mathbb{P}(L > x \mid S_1 = \infty) &= 0. \end{aligned}$$

- Case 2: $x < u < \infty$.

  We have $L \geq u$, and $x < u$, which implies that

$$\mathbb{P}(L > x \mid S_1 = u) = 1.$$

- Case 3: $0 \leq u \leq x$.

  We "rest the clock" at time $S_1 = u$, therefore we have a new renewal process $\tilde{S}$ of the same type as S (same function F), which starts at the new time 0.

$$\begin{aligned}\Rightarrow \mathbb{P}(L > x \mid S_1 = u) &= \mathbb{P}(\tilde{L} + u > x),\\ &= \mathbb{P}(L > x - u).\end{aligned}$$

Therefore,

$$\begin{aligned} f(x) &= \int_0^\infty \mathbb{P}(L > x \mid S_1 = u)\, dF(u),\\ &= \int_x^\infty \underbrace{\mathbb{P}(L > x \mid S_1 = u)}_{=1 \text{ (Case 2)}} dF(u) + \int_0^x \underbrace{\mathbb{P}(L > x \mid S_1 = u)}_{=f(x-u) \text{ (Case 3)}} dF(u),\\ \Rightarrow f(x) &= (F(\infty) - F(x)) + (F * f)(x).\end{aligned}$$

So, $f(\cdot)$ satisfies a renewal equation with $g(x) = F(\infty) - F(x) = P[x < W < \infty]$. According to the theorem in the previous section,

$$\begin{aligned} f(x) &= (R * g)(x),\\ &= \int_0^x (F(\infty) - F(x-u))\, dR(u),\\ &= F(\infty) \int_0^x dR(u) - \int_0^x F(x-u) dR(u),\\ &= F(\infty) R(x) - \underbrace{(R * F)(x)}_{=R(x)-1\ (\star)},\\ &= 1 - (1 - F(\infty)) R(x),\\ \Rightarrow \mathbb{P}(L \le x) &= (1 - F(\infty)) R(x).\end{aligned}$$

□

Remark 4.3.6 *1. Justification of* $(\star)$*:*

$$\begin{aligned} (R * F)(x) &= \sum_{n \ge 0} (F^{(n)} * F)(x) = \sum_{n \ge 0} F^{(n+1)}(x),\\ &= \sum_{n \ge 1} F^{(n)}(x) = \sum_{n \ge 0} F^{(n)}(x) - F^{(0)}(x),\\ &= R(x) - 1 \quad (F^{(0)}(x) = \mathbb{1}_{\{x \ge 0\}} = 1).\end{aligned}$$

2. *First moment of L:*

$$\begin{aligned}
\mathbb{E}[L] &= \mathbb{E}[L\,\mathbb{1}_{\{S_1<\infty\}}] \quad \textit{(since } S_1=\infty \Rightarrow L=0\textit{)},\\
&= \mathbb{E}[(L+S_1)\mathbb{1}_{S_1<\infty}], \quad (\star\star)\\
&= \mathbb{E}[L]F(\infty) + \int\limits_{[0,\infty)} u\,dF(u),\\
\Rightarrow \mathbb{E}[L] &= \frac{1}{1-F(\infty)}\int\limits_0^\infty (F(\infty)-F(u))du.
\end{aligned}$$

$(\star\star)$ *By doing the same as above, "resetting the clock at S_1."*

4.4 Recurrent renewal processes

4.4.1 Renewal theorems

In the case of recurrent renewal processes ($F(\infty)=1$), $R(t)$ tends towards infinity when $t\to\infty$. We will now examine how this convergence happens.

Theorem 4.4.1 (Fundamental Theorem) *In the case of recurrent renewal processes,*

(i) $\lim_{t\to\infty}\frac{N(t)}{t}=\frac{1}{\mathbb{E}[W]}$ *a.s.,*

(ii) $\lim_{t\to\infty}\frac{R(t)}{t}=\frac{1}{\mathbb{E}[W]}$,

where $\mathbb{E}[W]=\int\limits_0^\infty u\,dF(u)$ *is the expected time between two renewals.*

Note that almost sure convergence of a sequence $\{X_n\}$ of random variables does not necessarily imply convergence of the means (take for example $\mathbb{P}(X_n=0)=1-1/n^2, \mathbb{P}(X_n=2^n)=1/n^2$ where $X_n\to 0$ a.s. but $\mathbb{E}(X_n)$ does not converge to 0)! It does so only if the sequence is uniformly integrable .

In order to prove this theorem, we will need the following lemma:

Lemma 4.4.2

$$\mathbb{E}[S_{N(t)}]=R(t)\,\mathbb{E}[W]$$

Remark 4.4.3

$$S_{N(t)}=\sum_{i=1}^{N(t)}W_i,$$

is the time of the first renewal after time t. ($N(0) \geq 1$, because we count a renewal at time S_0.)

Proof. We write $f(t) = \mathbb{E}[S_{N(t)}]$. We will show that $f(\cdot)$ satisfies a renewal equation. In order to do this, we will as usual condition on S_1, the time of the first renewal after 0:

$$\begin{aligned} f(t) &= \mathbb{E}[S_{N(t)}], \\ &= \int_0^\infty \mathbb{E}[S_{N(t)} \mid S_1 = u] \, dF(u). \end{aligned}$$

- Case 1: $u > t$.

  We have

$$\begin{aligned} S_{N(t)} &= S_1, \\ \Rightarrow \mathbb{E}[S_{N(t)} \mid S_1 = u] &= u. \end{aligned}$$

- Case 2: $u \leq t$.

We "reset the clocks" to the renewal time u. We have a new renewal process which is the same as the old one, but starts at the new time "0", and

$$\begin{aligned} S_{N(t)} &= u + \tilde{S}_{N(t-u)}, \\ \Rightarrow \mathbb{E}[S_{N(t)} \mid S_1 = u] &= u + \underbrace{\mathbb{E}[\tilde{S}_{N(t-u)}]}_{=\mathbb{E}[S_{N(t-u)}]} \\ &\quad \text{because i.i.d. renewal times} \\ &= u + f(t-u). \end{aligned}$$

We therefore have:

$$\begin{aligned} f(t) &= \int_t^\infty u \, dF(u) + \int_0^t (u + f(t-u)) \, dF(u), \\ &= \int_0^\infty u \, dF(u) + \int_0^t f(t-u) \, dF(u), \\ &= \mathbb{E}[W] + (F * f)(t). \end{aligned}$$

Therefore, $f(t)$ is the solution of a renewal equation, with $g(t) = E[W]$. We then have

$$\begin{aligned} f(t) &= (R * g)(t), \\ &= \mathbb{E}[W] \int_0^t dR(u), \\ &= \mathbb{E}[W] R(t). \end{aligned}$$

□

Proof of the Fundamental Theorem. By the strong law of large numbers,

$$\frac{W_1 + W_2 + \cdots + W_n}{n} \xrightarrow{n \to \infty} \mathbb{E}[W] \quad a.s.$$

Since the renewal process is recurrent, we have $N(t) \to +\infty$ as $t \to \infty$ a.s., therefore

$$\frac{S_{N(t)}}{N(t)} = \frac{W_1 + W_2 + \cdots + W_{N(t)}}{N(t)} \xrightarrow{t \to \infty} \mathbb{E}[W] \quad a.s.$$

Furthermore, $S_{N(t)-1} \leq t \leq S_{N(t)}$ a.s., therefore, with probability one,

$$\begin{aligned} \frac{S_{N(t)-1}}{N(t)} &\leq \frac{t}{N(t)} \leq \frac{S_{N(t)}}{N(t)} \\ \underbrace{\frac{N(t)-1}{N(t)}}_{\to 1} \underbrace{\frac{S_{N(t)-1}}{N(t)-1}}_{\to \mathbb{E}[W]} &\leq \frac{t}{N(t)} \leq \underbrace{\frac{S_{N(t)}}{N(t)}}_{\to \mathbb{E}[W]} \end{aligned}$$

By the sandwich theorem, we then have $\frac{t}{N(t)} \to \mathbb{E}[W]$ a.s., and

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{\mathbb{E}[W]} \quad a.s. \quad (\star)$$

This proves (i).
We will now show (ii), that is,

$$\lim_{t\to\infty} \frac{R(t)}{t} = \lim_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] = \frac{1}{\mathbb{E}[W]}.$$

By Fatou's lemma,

$$\begin{aligned} \liminf_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \quad &\geq \quad \mathbb{E}\left[\liminf_{t\to\infty} \frac{N(t)}{t}\right], \\ &= \quad \frac{1}{\mathbb{E}[W]} \quad \text{(by } (\star)). \end{aligned}$$

It is then enough to show that

$$\limsup_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \leq \frac{1}{\mathbb{E}[W]} \quad (\star\star)$$

because we would then have

$$\begin{aligned} &\frac{1}{\mathbb{E}[W]} \leq \liminf_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \leq \limsup_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \leq \frac{1}{\mathbb{E}[W]}, \\ \Longrightarrow\ &\lim_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] = \frac{1}{\mathbb{E}[W]} \quad \text{(and therefore we have } (ii)). \end{aligned}$$

To show $(\star\star)$, we fixe $a > 0$, and we define a new renewal process:

$$Y_n = \begin{cases} W_n & \text{if } W_n \leq a, \\ 0 & \text{otherwise.} \end{cases}$$

Therefore we have the renewal process

$$\begin{cases} \tilde{S}_0 = 0, \\ \tilde{S}_{n+1} = \tilde{S}_n + Y_{n+1}. \end{cases}$$

It is clear that $S_n \geq \tilde{S}_n$ and $N(t) \leq \tilde{N}(t)$ a.s., therefore

$$\limsup_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \leq \limsup_{t\to\infty} \mathbb{E}\left[\frac{\tilde{N}(t)}{t}\right].$$

However, by the previous lemma,

$$\mathbb{E}[\tilde{N}(t)] = \frac{\mathbb{E}[\tilde{S}_{\tilde{N}(t)}]}{\mathbb{E}[Y]},$$

which implies that

$$\begin{aligned}\limsup_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] &\le \limsup_{t\to\infty} \frac{1}{t}\frac{1}{\mathbb{E}[Y]} \underbrace{\mathbb{E}[\tilde{S}_{\tilde{N}(t)}]}_{\le t+a}, \\ &\le \underbrace{\limsup_{t\to\infty} \frac{t+a}{t}}_{=1} \frac{1}{\mathbb{E}[Y]}, \\ &= \frac{1}{\mathbb{E}[Y]}.\end{aligned}$$

Therefore,

$$\limsup_{t\to\infty} \mathbb{E}\left[\frac{N(t)}{t}\right] \le \frac{1}{\mathbb{E}[Y(a)]} \quad \text{for all } a > 0.$$

Since $\mathbb{E}[Y(a)] \to_{a\to\infty} \mathbb{E}[W]$ (exercise), we have shown the inequality $(\star\star)$. □

Definition 4.4.4 *A function $g : \mathbb{R}^+ \to \mathbb{R}^+$ is* directly Riemann-integrable *($g \in \mathbb{D}$) if:*

(i) g is bounded on all finite intervals, and

(ii) if, for $n \in \mathbb{N}$ and fixed h, we define

$$\begin{aligned} m_n(h) &= \min\{g(x) : x \in [(n-1)h, nh]\}, \\ M_n(h) &= \max\{g(x) : x \in [(n-1)h, nh]\}, \end{aligned}$$

then we have:

- ♣ $\sum_{n=1}^{\infty} m_n(h)$ *converges absolutely for all h,*
- ♣ $\sum_{n=1}^{\infty} M_n(h)$ *converges absolutely for all h,*
- ♣ $\lim_{h\to 0} \sum_{n=1}^{\infty} (M_n(h) - m_n(h)) = 0$.

We will now give one of the most important renewal theorems:

Theorem 4.4.5 (Key Renewal Theorem) *For a recurrent renewal process, if the function g is directly Riemann-integrable,*

$$\lim_{t\to\infty} (R * g)(t) = \frac{1}{\mathbb{E}[W]} \int_0^{\infty} g(u)\, du.$$

(Proof: see Theorem 2.8 in Chapter 9 of Çinlar's book.)

Corollary 4.4.6 (Blackwell's Theorem)

$$\lim_{t\to\infty}(R(t+h)-R(t)) = \frac{h}{\mathbb{E}[W]}.$$

Proof. See exercise Serie 12. □

Remark 4.4.7 *In the chapter on Markov chains, we gave the following result without proving it:*
In an irreducible Markov chain with recurrent states:

$$\lim_{n\to\infty}\mathbb{P}(X_n = j \mid X_0 = j) = \begin{cases} 0 & \text{if the states are null recurrent,} \\ \pi_j > 0 & \text{if the states are positive recurrent.} \end{cases}$$

In fact, this is a consequence of Blackwell's Theorem: in the Markov chain $\{X_n\}$, if we start from state j, the intervals of time between two transitions through state j are i.i.d.:

Therefore, we have a recurrent renewal process $\{S_n : n \geq 0\}$, where S_n is the time of the n^{th} visit to state j,

$$\begin{cases} S_0 = 0, \\ S_{n+1} = S_n + W_{n+1}, \end{cases}$$

where the W_i are i.i.d. and have the same distribution as the first return time T_j given that the chain starts in j ($\mathbb{E}[W] = \mathbb{E}[T_j \mid X_0 = j]$, see chapter on Markov chains).
For $t \in \mathbb{N}$, we consider the renewal function of this renewal process:

$$\begin{aligned} R(t) - R(t-1) &= \mathbb{E}[N(t) - N(t-1)], \\ &= \mathbb{E}[\textit{number of passages in } j \textit{ between } t-1 \textit{ and } t], \\ &= \mathbb{E}[\mathbb{1}_{\{X(t)=j\}}] \quad \textit{(discrete time)}. \end{aligned}$$

Therefore, by Blackwell's Theorem,

$$\begin{aligned}\lim_{t\to\infty}\mathbb{P}(X_t = j \mid X_0 = j) &= \lim_{t\to\infty}(R(t) - R(t-1)), \\ &= \frac{1}{\mathbb{E}[W]} \\ &= \frac{1}{\mathbb{E}[T_j \mid X_0 = j]},\end{aligned}$$

and consequently,

$$\begin{cases}\mathbb{E}[W] = +\infty \Rightarrow \lim = 0 \Rightarrow \textit{null recurrent null states,} \\ \mathbb{E}[W] < \infty \Rightarrow \lim > 0 \Rightarrow \textit{ positive recurrent states.}\end{cases}$$

4.4.2 Survival of a renewal process

Definition 4.4.8 *The* survival *of a recurrent renewal process at time t is the random variable*

$$Z_t = S_{N(t)} - t.$$

(Recall that $S_{N(t)}$ is the time of the first renewal after time t.)

We will study the asymptotic distribution of Z_t as $t \to \infty$. Before doing so, we study the distribution of Z_t for any finite t:

Theorem 4.4.9

$$\mathbb{P}(Z_t > x) = \int_0^t \mathbb{P}(W > t - u + x)\, dR(u).$$

Proof. For any fixed t, we write $f(t,x) = \mathbb{P}(Z_t > x)$. By the usual renewal conditioning argument, we have

$$f(t,x) = \int_0^\infty \mathbb{P}(Z_t > x \mid S_1 = u)\, dF(u).$$

- If $u > t$:

  Then $S_{N(t)} = S_1 = u \Rightarrow Z_t = u - t$, and

  $$\mathbb{P}(Z_t > x \mid S_1 = u) = \mathbb{1}_{\{u-t>x\}}.$$

- If $u \leq t$: we reset the clock, and

  $$\mathbb{P}(Z_t > x \mid S_1 = u) = \mathbb{P}(Z_{t-u} > x) = f(t-u, x).$$

Therefore,

$$\begin{aligned} f(t,x) &= \int_t^\infty \mathbb{1}_{\{u \geq t+x\}} \, dF(u) + \int_0^t f(t-u,x) \, dF(u), \\ &= \mathbb{P}(W > t+x) + (F * f(\cdot, x))(t). \end{aligned}$$

We have a renewal equation with $g(t,x) = \mathbb{P}(W > t+x)$, and therefore

$$\begin{aligned} f(t,x) &= (R * g(\cdot, x))(t), \\ &= \int_0^t g(t-u,x) \, dR(u). \end{aligned}$$

□

Proposition 4.4.10 (Asymptotic survival of a recurrent renewal process) *If*

$$V(x) := \lim_{t\to\infty} \mathbb{P}(Z_t \leq x),$$

we have

$$V(x) = \frac{1}{\mathbb{E}[W]} \int_0^x (1 - F(u)) \, du.$$

Note that this means that the density of the asymptotic survival is given by $(1 - F(x))/\mathbb{E}[W]$.

Proof. With $g(t,x) = \mathbb{P}(W > t + x) = 1 - F(t + x)$, by the previous theorem,

$$\begin{aligned}
\lim_{t\to\infty} \mathbb{P}(Z_t > x) &= \lim_{t\to\infty} (R * g)(t), \\
&= \frac{1}{\mathbb{E}[W]} \int_0^\infty g(u,x)\, du, \quad \text{(Key Renewal Theorem)} \\
&= \frac{1}{\mathbb{E}[W]} \int_x^\infty (1 - F(u))\, du.
\end{aligned}$$

This implies

$$\begin{aligned}
V(x) &= 1 - \lim_{t\to\infty} \mathbb{P}(Z_t > x) \\
&= \frac{1}{\mathbb{E}[W]} \left(\mathbb{E}[W] - \int_x^\infty (1 - F(u))\, du \right) \\
&= \frac{1}{\mathbb{E}[W]} \left(\int_0^\infty (1 - F(u))\, du - \int_x^\infty (1 - F(u))\, du \right),
\end{aligned}$$

because $\mathbb{E}[W] = \int_0^\infty \mathbb{P}(W > u)\, du$. □

4.5 General renewal processes

Definition 4.5.1 *A* general renewal process *(also called* delayed renewal process*) is a process* $\{\widehat{S}_n : n \in \mathbb{N}\}$ *such that*

$$\widehat{S}_{n+1} = \widehat{S}_n + W_{n+1} \quad \text{for all } n,$$

where:

- $W_1, W_2, \ldots$ *are non-negative i.i.d. random variables with distribution function* $F(\cdot)$*,*
- $\widehat{S}_0$ *is a non-negative random variable with distribution function* $G(\cdot)$*, independent of the* W_i*.*

It is therefore a renewal process, but one for which the first renewal $\widehat{S}_0$ doesn't necessarily come at time 0, but after a random time given by the distribution function $G(\cdot)$. There are no new tools needed. In handling a general renewal process, first we condition the event in question on the time $\widehat{S}_0$ of first renewal, and then we use the fact that at time $\widehat{S}_0$ there starts an ordinary renewal process.

$\widehat{S}_n$ is still the time of the n^{th} renewal. If $\widehat{N}(t)$ is the number of renewals before t,

$$\begin{aligned} \mathbb{P}(\widehat{S}_n \le t) &= (G * F^{(n)})(t), \\ \mathbb{P}(\widehat{N}(t) > k) &= \mathbb{P}(\widehat{S}_k \le t). \end{aligned}$$

Indeed, for the first equality,

$$\begin{aligned} \mathbb{P}(\widehat{S}_n \le t) &= \mathbb{P}(\widehat{S}_0 + W_1 + \cdots + W_n \le t), \\ &= (G * F^{(n)})(t) \end{aligned}$$

(sum of independent random variables).

The renewal function takes the following form:

$$\begin{aligned} \widehat{R}(t) &= \mathbb{E}[\widehat{N}(t)], \\ &= \sum_{n\ge 0} (G * F^{(n)}(t) \\ &= (G * R)(t). \end{aligned}$$

4.6 Stationary processes

Definition 4.6.1 *Let $\{S_n : n \in \mathbb{N}\}$ be a renewal process whose inter-renewal times W_i have distribution function $F(\cdot)$. The* stationary process *$\{\widehat{S}_n : n \in \mathbb{N}\}$ associated to $\{S_n\}$ is the general renewal process given by the W_i, for which the distribution of $\widehat{S}_0$ is the asymptotic survival distribution $V(\cdot)$ ($G(x) = V(x)$).*

Intuitively, the stationary process corresponds to a renewal process that has been going on for a long time before time $t = 0$. We will see that, as expected, the stationary process verifies properties which are only asymptotic for the associated renewal process.

Proposition 4.6.2 *If $\{\widehat{S}_n\}$ is a stationary process, then*

(i) $\widehat{R}(t) = \dfrac{t}{\mathbb{E}[W]}$, *and*

(ii) $\mathbb{P}(\widehat{Z}_t \le x) = V(x)$ *for all* t.

Proof. <u>*(i)*</u>

$$\begin{aligned}
\widehat{R}(t) &= (V * R)(t) \quad \text{(see general renewal processes)},\\
&= \int_0^t V(t-u)\,dR(u),\\
&= \frac{1}{\mathbb{E}[W]}\int_0^t \int_0^{t-u} (1-F(v))\,dv\,dR(u),\\
&= \frac{1}{\mathbb{E}[W]}\int_0^t \int_u^{t} (1-F(s-u))\,ds\,dR(u),\\
&= \frac{1}{\mathbb{E}[W]}\int_0^t \int_0^{s} (1-F(s-u))\,dR(u)\,ds,
\end{aligned}$$

where the last equality is obtained by changing the way of integrating on the domain. However,

$$\begin{aligned}
\int_0^s (1-F(s-u))\,dR(u) &= R(s) - F * R(s),\\
&= 1 + F * R - F * R = 1.
\end{aligned}$$

This implies

$$\widehat{R}(t) = \frac{1}{\mathbb{E}[W]}\int_0^t ds = \frac{t}{\mathbb{E}[W]}.$$

<u>*(ii)*</u> Let $f_x(t) = \mathbb{P}(\widehat{Z}_t > x)$. We have

$$f_x(t) = \int_0^\infty \mathbb{P}(\widehat{Z}_t > x \mid \widehat{S}_0 = u)\,dV(u).$$

- If $u > t$

Then $\widehat{Z}_t = u - t$, so that

$$\mathbb{P}(\widehat{Z}_t > x \mid \widehat{S}_0 = u) = \mathbb{1}_{\{u>t+x\}}.$$

- If $u \le t$

We reset the clock to $\widehat{S}_0$. If we start from $\widehat{S}_0$, by independence, the stationary process behaves afterwards like a normal renewal process, and therefore

$$\mathbb{P}(\widehat{Z}_t > x \mid \widehat{S}_0 = u) = \mathbb{P}(Z_{t-u} > x).$$

So, we have

$$f_x(t) = \underbrace{\int_{t+x}^{\infty} dV(u)}_{=1-V(t+x)} + \underbrace{\int_0^t \mathbb{P}(Z_{t-u} > x)\, dV(u)}_{(\star)}$$

Here we do not have a renewal equation because $(\star)$ is not $(V * f_x)$, and $\mathbb{P}(Z_{t-u} > x) \neq \mathbb{P}(\widehat{Z}_{t-u} > x) \dots$. But we have already calculated $\mathbb{P}(Z_{t-u} > x)$ in a previous section, and

$$\mathbb{P}(Z_t > x) = (R * g_x)(t),$$

where $g_x(t) = 1 - F(x+t) = \mathbb{P}(W > x+t)$. We therefore have

$$\begin{aligned}
\mathbb{P}(\widehat{Z}_t > x) &= 1 - V(t+x) + \int_0^t (R * g_x)(t-u)\, dV(u), \\
&= 1 - V(t+x) + (V * R * g_x)(t), \\
&= 1 - V(t+x) + (\widehat{R} * g_x)(t), \\
&= 1 - V(t+x) + \frac{1}{\mathbb{E}[W]} \int_0^t g_x(t-u)\, du, \qquad \text{by (i)}, \\
&= 1 - V(t+x) + \frac{1}{\mathbb{E}[W]} \int_0^t (1 - F(x+t-u))\, du, \\
&= 1 - V(t+x) + \frac{1}{\mathbb{E}[W]} \int_x^{x+t} (1 - F(v))\, dv \quad (v = x+t-u), \\
&= 1 - V(t+x) + V(t+x) - V(x), \\
&= 1 - V(x),
\end{aligned}$$

which shows that $\mathbb{P}(\widehat{Z}_t \le x) = V(x)$. □

Example 4.6.3 *The simplest stationary renewal process is the Poisson process. In this case, the survival at time t, Z_t, has exactly the same (exponential) distribution as the inter-renewal times (exercise). This is actually the only renewal process satisfying this property.*