Markov chains and Martingales applied to the analysis of discrete random structures.

Cécile Mailler

Nablus, August 18–28, 2014

Contents

Introduction

Probability and theoretical Computer Science interact in many ways: from stochastic algorithms such as `ethernet` to analysis of algorithms on average. This course aims at presenting two very classical objects in probability theory: Markov chains and martingales through their applications in Computer Science. Our goal is not to give the complete theory, but only to give definitions, basic results and numerous examples. Not all proofs will be developed.

Let us start with a story. John gets out of a bar in Manhattan and wants to go to his hotel. He his so drunk though, that at each crossing, he does not remember where he comes from and choose one road out of the four at random. The next crossing he visits thus only depends on where he is now and what will be his decision, but it does not depend on the past. This is the heuristic of a Markov chain: *the future only depends on the present and not on the past.* Random walks are the classical example of Markov chains, and we will prove in this course that, John will almost surely reach his hotel in finite time – whereas a drunken fish in a 3D undersea Manhattan would almost surely never find his hotel.

A martingale models a fair game: let us say you play heads-or-tails against you banker. Each time you toss a coin, if its heads, you win one peso, if its tail, you loose one peso. If the coin is fair, your expected wealth after the next toss is equal to your actual wealth. This is the heuristic definition of a martingale.

The course is divided into 4 sections: the two first ones concern discrete time Markov chains and martingales, while the two last ones detail continuous time versions of both objects. The discrete time objects being less intricate, we will study them in full detail. Instead of studying continuous time Markov chains in full generality, we will focus on queuing processes, very useful in Computer Science and which study is more basic. In all sections, our aim will be to state convergence results for the considered stochastic processes.

Prerequisites for this course are elementary probability: in particular conditional expectation, convergence of sequences of random variables. It could also be useful to know about σ-algebras, even if a heuristic description should be enough.

This course does not aim to be exhaustive. Many references are available to go further: one can for example cite the following

[Norris] J. R. Norris: **Markov Chains**. *Cambridge University Press*, 1998.

[Williams] D. Williams: **Probability with Martingales**. *Cambridge University Press*, 1991.

[Steward] W. J. Steward: **Probability, Markov Chains, Queues, and Simulation: The Mathematical Basis of Performance Modelling**. *Princeton University Press*, 2009.

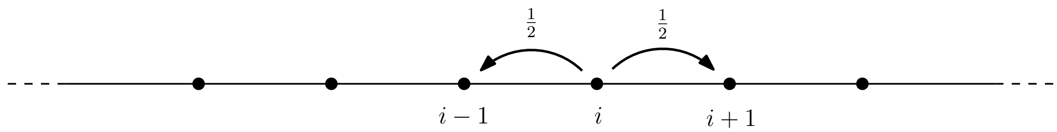


FIGURE 1 – *The simple random walk on* $\mathbb{Z}$.

1 Discrete time Markov chains

1.1 Definitions and first properties

Markov chains can be defined on any space: discrete or continuous. In this course, we will only treat with discrete state spaces, but one has to keep in mind that Markov chains exists as well on $\mathbb{R}$, for example. But in the following, E will always be a discrete space.

Definition 1.1

A matrix $P = (p_{x,y})_{x,y\in E}$ *is a* **stochastic matrix** *if, for all* $x \in E$,

$$\sum_{y\in E} p_{x,y} = 1.$$

Definition 1.2

Let P *be a stochastic matrix on* E. *A sequence* $(X_n)_{n\geq 1}$ *of random variables taking value in* E *is a* **Markov chain** *of initial law* μ_0 *and transition matrix* P *if*

(i) X_0 *has law* μ_0, *and,*

(ii) for all $n \geq 0$, *for all* $x \in E$,

$$\mathbb{P}(X_{n+1} = x \,|\, X_n, \dots, X_0) = \mathbb{P}(X_{n+1} = x \,|\, X_n) = p_{X_n,x}.$$

Proposition 1.3

Let $(X_n)_{n\geq 1}$ *be a* **Markov chain** *of initial law* μ_0 *and transition matrix* P. *Then, for all* $n \geq 0$, *for all* $x_0, \dots, x_n \in E$,

$$\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mu(x_0)p_{x_0,x_1}\dots p_{x_{n-1},x_n}.$$

Example 1.1: The simple random walk on $\mathbb{Z}$ (cf. Figure 1).

Wild Bill Hickok plays *heads or tails* against his banker. His honesty is so much renowned that his banker allows him an infinite credit: he will eventually pay his dept after arresting some wanted outlaw. At time 0, Bill owns x_0 dollars. Each time Bill tosses a coin, he earns one dollar if its heads and looses one if its tails.

If we denote by X_n the number of dollars Wild Bill owns after he has tossed his n^{th} coin, the sequence $(X_n)_{n\geq 0}$ is a Markov chain on $E = \mathbb{Z}$. Its initial law is $\mu_0 = \delta_{x_0}$ and its transition probabilities are defined as follows: for all $i \in \mathbb{Z}$,

$$\begin{array}{lcll} p_{i,i+1} & = & 1/2 & \\ p_{i,i-1} & = & 1/2 & \\ p_{i,j} & = & 0 & \text{for all} \quad j \notin \{i-1, i+1\}. \end{array}$$

Example 1.2: Umbrellas management in England.

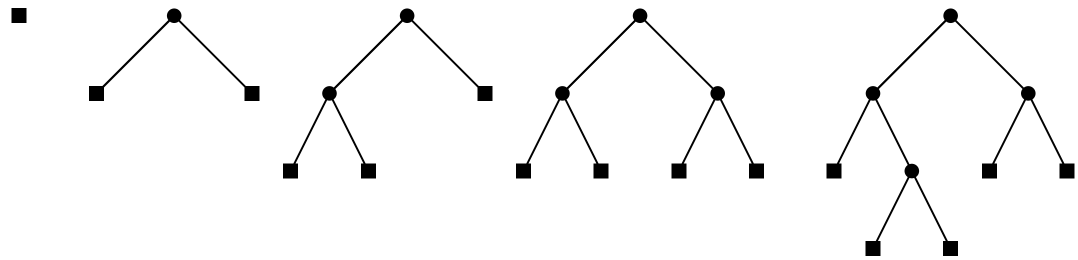


FIGURE 2 – *A realisation of the random BST from time 1 (on the left) to time 5 (on the right).*

I own n umbrellas (n is reasonably large because I live in England). At the beginning of the year, all my umbrellas are at home. Every morning, I go from home to work and every evening from work to home. If it rains when I leave home, and only if it rains, I take one umbrella with me. If it rains when I leave work, and only if it rains, I take one umbrella with me. And each time I leave a building, it rains with probability p (independently).

If we denote by X_n the number of umbrella I have at home at the n^{th} night of the year, then X_n is a Markov chain on $E = \{0, \ldots, n\}$. Can you find its probability transitions? For all $i \in \{1, \ldots, n-1\}$

$$\begin{array}{lcl} p_{i,i-1} & = & \\ p_{i,i+1} & = & \\ p_{i,i} & = & \\ p_{i,j} & = & 0 \text{ if } j \notin \{i-1, i, i+1\} \end{array}.$$

And don't forget the extremal cases $i = 0$ and $i = n$.

Example 1.3: Ehrenfest's urn

Snowy and Snoopy have fleas: in total, there are N fleas. Each day, a flea chosen at random amongst the N fleas jumps from one dog to the other.

Let us denote by X_n the number of fleas on Snowy on the n^{th} day. The sequence X_n is a Markov chain of transition probabilities

$$\begin{array}{lcl} p_{i,i-1} & = & i/N \\ p_{i,i+1} & = & 1 - i/N \\ p_{i,j} & = & 0 \text{ if } j \notin \{i-1, i+1\} \end{array}$$

Example 1.4: The Binary Search Tree (cf. Figure 2)

The random BST is defined as follows: At time 1, it is a single node. At each step, a leaf of the tree is picked up uniformly at random and becomes an internal node with two leaves as children.

If we denote by T_n the random binary search tree at time n, then $(T_n)_{n\geq 0}$ is a Markov chain on E, the space of binary trees. Can you understand its transition probabilities?

Theorem 1.4 (*Markov property*)

Let (X_n) be a Markov chain of transition matrix P and initial law μ_0. Then, for all $m \geq 1$, $(X_{m+n} \,|\, X_0, \ldots, X_m)_{n\geq 0}$ is a Markov chain of transition matrix P and initial law δ_{X_m}.

1.2 Stationary probability and reversibility

Definition 1.5

A probability measure π on E is a **stationary probability** *of a Markov chain of transition matrix P if and only if*

$$\pi P = \pi,$$

i.e. for all $x \in E$, $\sum_{y \in E} \pi_y p_{y,x} = \pi_x$.

The existence of such a stationary probability is not guaranteed; it is for example interesting to prove that the simple random walk on $\mathbb{Z}$ does not admit a stationary probability.

Example 1.5: Umbrellas management in England.

The probability transitions of the umbrellas management problem (cf. Example 1.2) are given by: for all $i \in \{1, \dots, N-1\}$

$$\begin{aligned} p_{i,i+1} &= p(1-p) & p_{0,1} &= p & p_{N,N-1} &= p(1-p) \\ p_{i,i-1} &= p(1-p) & p_{0,0} &= 1-p & p_{N,N} &= 1-p(1-p) \\ p_{i,i} &= 1-2p(1-p) \\ p_{i,j} &= 0 \text{ if } j \notin \{i-1, i, i+1\} \end{aligned}$$

Thus, to be a stationary probability of this Markov chain, π has to verify

$$\pi_0 = p(1-p)\pi_1 + (1-p)\pi_0$$

$$\pi_N = p(1-p)\pi_{N-1} + (1-p(1-p))\pi_N$$

and, for all $i \in \{1, \dots, N-1\}$,

$$\pi_i = p(1-p)\pi_{i-1} + (1-2p(1-p))\pi_i + p(1-p)\pi_{i+1}.$$

It implies that

$$\pi_0 = (1-p)\pi_1 \quad \text{and} \quad \pi_N = \pi_{N-1},$$

and, for all $i \in \{1, \dots, N-1\}$, $2\pi_i = \pi_{i-1} + \pi_{i+1}$, which implies

$$\pi_i = \frac{1}{N-p} \quad \text{for all } i \in \{1, \dots, N\} \text{ and } \pi_0 = \frac{1-p}{N-p}.$$

The unique stationary probability of this Markov chain is this *almost uniform law* on $\{0, \dots, N\}$.

Example 1.6: Ehrenfest's urn.

To be a probability distribution on the Ehrenfest's urn defined in Example 1.3, π has to verify:

$$\pi_0 = \frac{1}{N}\pi_1, \quad \pi_N = \frac{1}{N}\pi_{N-1},$$

and, for all $i \in \{1, \dots, N-1\}$,

$$\pi_i = \left(1 - \frac{i-1}{N}\right)\pi_{i-1} + \frac{i+1}{N}\pi_{i+1}.$$

One can check that if, for all $i \in \{0, \dots, N\}$,

$$\pi_i = \frac{1}{2^N}\binom{N}{i},$$

then, π is a stationary probability of the Ehrenfest's urn.

Definition 1.6

A Markov chain of transition matrix P is reversible according to a probability measure π if and only if, for all $x, y \in E$,

$$\pi_x p_{x,y} = \pi_y p_{y,x}.$$

Lemma 1.7

If a Markov chain is reversible according to a probability measure π, then π is an stationary probability of this Markov chain.

Proof. Recall that π is invariant for a Markov chain of transition matrix P if and only if $\pi P = \pi$. Consider a Markov chain of transition matrix P and assume it is reversible according to π. Then,

$$\sum_{y \in E} \pi_y p_{y,x} = \sum_{y \in E} \pi_x p_{x,y} = \pi_x,$$

which implies that π is invariant for the considered Markov chain. $\square$

If $(X_n)_{n\geq 0}$ is a Markov chain reversible according to π and with initial distribution π, then, for all $n \in \mathbb{N}$, the random vectors $(X_0, \dots, X_n)$ and $(X_n, \dots, X_0)$ have the same law.

1.3 Recurrence and transience

Definition 1.8

An **absorbing state** *of a Markov chain $(X_n)_{n\geq 0}$ is a state $x \in E$ such that $p_{x,x} = 1$.*

Let $(X_n)_{n\geq 0}$ be a Markov chain of initial law μ_0 and of transition matrix P. For all $n \geq 1$, let $p^{(n)}_{x,y} = \mathbb{P}(X_n = y|X_0 = x) = \mathbb{P}_x(X_n = y)$. Then, the n^{th} power of the transition matrix P is given by

$$P^n = \left(p^{(n)}_{x,y}\right)_{x,y\in E}.$$

Definition 1.9

A Markov chain of transition matrix $P = (p_{x,y})_{x,y\in E}$ is **irreducible** *if and only if, for all $x, y \in E$, the probability that a Markov chain starting from x eventually reaches y is positive, i.e. if and only if, for all $x, y \in E$, there exists $n \geq 0$ such that $p^{(n)}_{x,y} > 0$.*

The examples of Markov chain introduced in Section 1 are all irreducible, except the Binary Search Tree markov chain.

The reaching time of a state $x \in E$ is defined and denoted as follows:

$$\tau_x = \inf\{n \geq 1 \,|\, X_n = x\}.$$

Definition 1.10

Let $(X_n)_{n\geq 0}$ be a Markov chain, a state $x \in E$ is

- **recurrent** *for this Markov chain if $\mathbb{P}(\tau_x < +\infty) = 1$;*
- **transient** *for this Markov chain if $\mathbb{P}(\tau_x = +\infty) = 1$.*

A Markov chain is recurrent (resp. transient) if all its states are recurrent (resp. transient).

For all $x \in E$, let us denote by $N_x = \sum_{n\geq 0} \mathbb{1}_{X_n = x}$ the number of visits of the Markov chain $(X_n)_{n\geq 0}$ at state x.

Proposition 1.11

Let $(X_n)_{n\geq 0}$ be a Markov chain of transition matrix P. Then:

(i) If $x \in E$ is transient, then $\mathbb{P}_x(N_x < +\infty) = 1$, $\sum_{n\geq 0} p^{(n)}_{x,x} < +\infty$, and, conditioned on $\{X_0 = x\}$, N_x is a geometric random variable of parameter $\mathbb{P}_x(T_x = +\infty)$.

(ii) If x is recurrent then $\mathbb{P}_x(N_x = +\infty) = 1$ and $\sum_{n\geq 0} p^{(n)}_{x,x} = +\infty$.

(iii) If the Markov chain $(X_n)_{n\geq 0}$ is irreducible, then it is either recurrent or transient. In the first case, for all $x \in E$, $\mathbb{P}(N_x = +\infty) = 1$. In the second case, for all $x \in E$, $\mathbb{P}(N_x < +\infty) = 1$.

Proof. First of all, remark that

$$\mathbb{P}_x(\tau_x = +\infty) = \mathbb{P}_x(N_x = 1).$$

For all $m \geq 1$, let us denote by $\tau_x^{(m)}$ the time of the m^{th} visit of the chain into x: $\tau_x^{(1)} := \tau_x$, and

$$\tau_x^{(m)} := \inf\{i > \tau_x^{(m-1)} \mid X_i = x\}.$$

Remark that, for all $m \geq 1$,

$$\begin{aligned}
\mathbb{P}_x(N_x > m) &= \sum_{s\geq m} \mathbb{P}_x(N_x > m \text{ and } \tau_x^{(m)} = s) \\
&= \sum_{s\geq m} \mathbb{P}_x\left(\sum_{i=1}^{s} \mathbb{1}_{X_i=x} = m \text{ and } X_s = x \text{ and } \sum_{i\geq s+1} \mathbb{1}_{X_i=x} > 1\right) \\
&= \sum_{s\geq m} \mathbb{P}_x\left(\sum_{i=1}^{s} \mathbb{1}_{X_i=x} = m \text{ and } X_s = x\right) \mathbb{P}_x\left(\sum_{i\geq 1} \mathbb{1}_{X_i=x} > 1\right) \\
&= \mathbb{P}_x(N_x \geq m)\mathbb{P}_x(N_x > 1).
\end{aligned}$$

Thus, if we denote by $p := \mathbb{P}_x(\tau_x = +\infty) = \mathbb{P}_x(N_x = 1)$, we get, for all $m \geq 0$,

$$\mathbb{P}_x(N_x > m) = (1-p)^m.$$

it immediately implies that

$$\mathbb{P}_x(N_x = m) = p(1-p)^{m-1}.$$

Finally, note that

$$\mathbb{E}N_x = \sum_{i\geq 1} \mathbb{P}_x(X_i = x) = \sum_{i\geq 1} p^{(i)}_{x,x}.$$

(i) If $x \in E$ is transient, then $p > 0$, and conditioned on $\{X_0 = x\}$, N_x is geometrically distributed with parameter p, which implies that its expectation is finite.

(ii) If x is recurrent, then $p = 0$, $\mathbb{P}_x(N_x = +\infty) = 1$ and the expectation of N_x is infinite.

(iii) Let x and y in E. Note that since the chain is irreducible, there exist $n_1, n_2 > 0$ such that $p^{(n_1)}_{x,y} > 0$ and $p^{(n_2)}_{y,x} > 0$. In addition, for all $n \geq 0$,

$$p^{(n+n_1+n_2)}_{y,y} \geq p^{(n_2)}_{y,x} p^n_{x,x} p^{(n_1)}_{x,y},$$

which implies that the two series $\sum_{n\geq 1} p^{(n)}_{x,x}$ and $\sum_{n\geq 1} p^{(n)}_{y,y}$ have the same behaviour. Therefore, an irreducible chain is either recurrent or transient.

If the chain is transient, then, for all $x \in E$,

$$\mathbb{P}(N_x = +\infty) = \sum_{s\geq 0} \mathbb{P}(\tau_x = s)\mathbb{P}_x(N_x = +\infty) = 0.$$

The recurrent case is more complicated and left to the reader. □

Example 1.7: The simple random walk on $\mathbb{Z}$ is recurrent (cf. Example 1.1)

For all $n \geq 0$,

$$p_{0,0}^{(2n)} = \binom{2n}{n} \frac{1}{2^{2n}} = \mathtt{Cat}_n 4^{-n},$$

with $\mathtt{Cat}_n = \frac{1}{n+1}\binom{2n}{n}$. Recall that $\mathtt{Cat}_n \sim n^{-3/2} 4^n$ when $n \to +\infty$, thus,

$$\sum_{n \geq 0} p_{00}^{(n)} = +\infty,$$

which implies, by Proposition 1.11 lemma that 0 is recurrent. Since the simple random walk is irreducible, we can conclude that the whole chain is recurrent.

Remark: It can be proved that the simple random walk on $\mathbb{Z}^2$ is recurrent as well, but that the simple random walk on $\mathbb{Z}^3$ is transient. In fact, for all $d \geq 3$, the simple walk on $\mathbb{Z}^d$ is transient.

Definition 1.12

Let $(X_n)_{n \geq 0}$ be a Markov chain of transition matrix P. The **period** *of a state $x \in E$ is the* GCD *of $\{n > 0 \,|\, p_{x,x}^{(n)} > 0\}$. A state is said to be* **aperiodic** *if its period is 1 and periodic otherwise. A Markov chain is aperiodic if all its states are aperiodic.*

Proposition 1.13

Let $(X_n)_{n \geq 0}$ be a Markov chain of transition matrix P, then:

(i) If $x \in E$ is aperiodic, the $p_{x,x}^{(n)} > 0$ for all n large enough.

(ii) If $(X_n)_{n \geq 0}$ is irreducible, it is aperiodic as soon as one of its states is aperiodic.

Proof. (i) Assume that $x \in E$ is aperiodic. Met $I = \{n \geq 1 \mid p_{x,x}^{(n)} > 0\}$. Remark that I is stable by addition. There exists $K > 0$, $n_1, \ldots, n_K > 0$ and $a_1, \ldots, a_K \in \mathbb{Z}$ such that $n_i \in I$ for all $i \in \{1, \ldots, K\}$, and

$$1 = \sum_{i=1}^{K} a_i n_i.$$

Let $n_1 = \sum_{a_i > 0} a_i n_i$ and $n_2 = -\sum_{a_i < 0} a_i n_i$. We know that $n_1, n_2 \in I$ and $n_1 - n_2 = 1$.

Let $n \geq n_2^2$, then, there exists $q \geq n_2$ and $0 \leq r < n_2$ such that

$$n = qn_2 + r = qn_2 + r(n_1 - n_2) = (q - r)n_2 + rn_1,$$

which implies that any $n \geq n_2$ belongs to I.

(ii) Assume that $x \in E$ is aperiodic, then, for all n large enough, $p_{x,x}^{(n)} > 0$ For all $y \in E$, there exists $n_1, n_2 \geq 1$ such that $p_{x,y}^{(n_1)} > 0$ and $p_{y,x}^{(n_2)} > 0$. Thus, for all $n \geq 1$,

$$p_{y,y}^{(n+n_1+n_2)} \geq p_{y,x}^{(n_2)} p_{x,x}^{(n)} p_{x,y}^{(n_1)},$$

which implies that, for all n large enough, $p_{y,y}^{(n)} > 0$ and thus that y is also aperiodic. $\square$

Recall that $\tau_x = \inf\{n \geq 1 \,|\, X_n = x\}$. For all $x \in E$, we define $\nu(x) = \frac{1}{\mathbb{E}_x T_x} \in [0, 1]$. Remark that if $(X_n)_{n \geq 0}$ is an irreducible, transient Markov chain, then, for all $x \in E$, $\nu(x) = 0$.

Definition 1.14

A recurrent state x of the Markov chain is **positive recurrent** *if $\nu(x) > 0$ and* **null recurrent** *if $\nu(x) = 0$. A Markov chain is called positive (resp. null) recurrent if all its states are positive (resp. null) recurrent.*

1.4 Ergodic theorems

An event A is almost sure for a Markov chain if, for all state $x \in E$, $\mathbb{P}_x(A) = 1$, i.e. if $\mathbb{P}(A) = 1$ for any initial distribution μ_0.

Theorem 1.15

Let $(X_n)_{n \geq 0}$ be an irreducible Markov chain on E.

(i) $(X_n)_{n \geq 0}$ is either transient, either positive recurrent, or null recurrent.

(ii) If $(X_n)_{n \geq 0}$ is transient or null recurrent, then, she has no invariant probability, and $\nu = 0$.

(iii) For all $x \in E$, we have, almost surely when n tends to infinity,

$$\frac{1}{n} \sum_{m=0}^{n} \mathbb{1}_{X_m = x} \to \nu(x).$$

This result tells you the following: if you are able to exhibit an invariant probability for a Markov chain, then this Markov chain is recurrent. It thus apply for example for the Ehrenfest urn (cf. Example 1.3) or for the umbrellas Markov chain (cf. Example 1.2) which are thus both recurrent. Remark that knowing that a Markov chain admits no invariant probability is not enough to conclude that it is not recurrent: the simple random walk on $\mathbb{Z}$, for example is recurrent but has no stationary distribution.

Theorem 1.16 (*Ergodic Theorem*)

Let $(X_n)_{n \geq 0}$ be an irreducible, positive recurrent Markov chain on E, then:

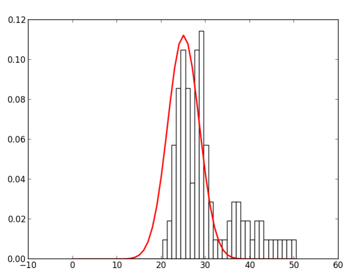
1. *ν is a probability distribution on E and is the unique invariant probability of $(X_n)_{n \geq 0}$. We moreover have that $\nu(x) > 0$ for all $x \in E$.*
2. *For all function $f : E \to \mathbb{R}$ such that $f \geq 0$ or $\int_E f(x) d\nu(x) < +\infty$, we have,*

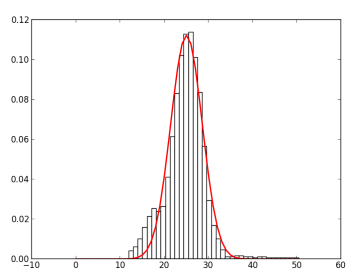
$$\frac{1}{n} \sum_{m=0}^{n} f(X_k) \to \int_E f(x) d\nu(x).$$

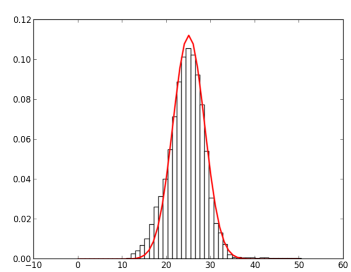
3. *If, in addition, $(X_n)_{n \geq 0}$ is aperiodic, then $X_n \to \nu$ in law when n tends to infinity, and thus, $\mathbb{P}(X_n = x) \to \nu(x)$ for all $x \in E$ when n tends to $+\infty$.*

Example 1.8: Ehrenfest's urn.

The Ehrenfest urn is an irreducible positive recurrent chain on $\{1, \dots, N\}$? Therefore, Theorem 1.16 applies as it can be seen on the following simulations: The figure below is the histogram of the number of fleas on Snoopy from time 0 to time 100 (resp. 2000, resp. 5000), when $N = 50$, starting from Snoopy having 50 fleas on it at time 0. The blue curve is the stationary distribution of this Markov chain.

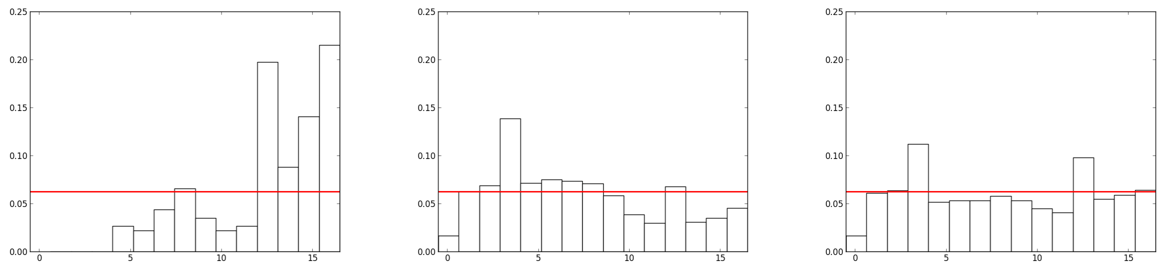






Example 1.9: Umbrellas.

The umbrellas Markov chain described in Example 1.2 is an irreducible positive recurrent chain on $\{1, \dots, N\}$? Therefore, Theorem 1.16 applies as it can be seen on the following simulations: The figure below is the histogram of the number of umbrellas at home between days 1 and 200 (resp. 5000, resp. 10000), when $N = 16$. The red curve is the uniform law on $\{0, \dots, N\}$.



Remark: One can prove that both for a transient and for a null recurrent irreducible Markov chain, $\lim_{n\to+\infty} \mathbb{P}(X_n = x) = 0$.

Corollary 1.17

An irreducible Markov chain on a **finite** *space E is positive recurrent and thus, ν is its unique invariant probability and Theorem 1.16 applies.*

2 Discrete time martingales

2.1 Definitions and first properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 2.1

Let $(\mathcal{F}_n)_{n\geq 0}$ a filtration of Ω, i.e. an increasing family of sub σ-algebras of $\mathcal{F}$. A sequence $(M_n)_{n\geq 0}$ of random variables is an $\mathcal{F}_n$-**martingale** *if, and only if, for all $n \geq 0$,*

(i) M_n is $\mathcal{F}_n$ measurable,

(ii) M_n is integrable, i.e. $\mathbb{E}M_n < +\infty$, and

(iii) $\mathbb{E}[M_{n+1}|\mathcal{F}_n] = M_n$ almost surely.

In most applications, the considered filtration is $\mathcal{F}_n = \sigma(M_1, \dots, M_n)$, i.e. contains all the information of the martingale before time n. More generally, given a sequence $(X_n)_{n\geq 0}$ of random variables, we call the filtration $(\mathcal{F}_n = \sigma(X_1, \dots, X_n))_{n\geq 0}$ its **natural filtration**.

Definition 2.2

If (iii) in Definition 2.1 is replaced by

- $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \leq M_n$ *a.s., we get the definition of a* **super-martingale**.
- $\mathbb{E}[M_{n+1}|\mathcal{F}_n] \geq M_n$ *a.s., we get the definition of a* **sub-martingale**.

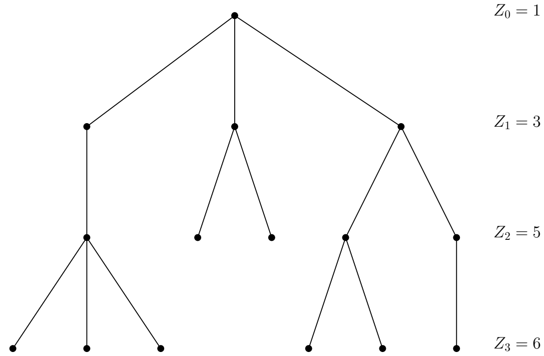


FIGURE 3 – *A realisation of a Galton-Watson tree.*

Proposition 2.3

Let $(M_n)_{n\geq 0}$ be a $\mathcal{F}_n$-martingale, then, for all $n \geq 0$, $\mathbb{E}M_n = \mathbb{E}M_0$.

What can we say of the sequence $(\mathbb{E}M_n)_{n\geq 0}$ for a super-martingale (resp. sub-martingale)?

Example 2.1: Simple random walk again

Let $(X_n)_{n\geq 0}$ be a sequence of integrable i.i.d. random variables, such that $\mathbb{E}X_1 = 0$. Can you check that $S_n = \sum_{i=1}^n X_i$ is a martingale?

Example 2.2: Galton-Watson tree (cf. Figure 3)

A Galton-Watson tree is described as follows: The first generation is composed of a unique root. Each individual of generation n gives birth to a random number ξ of individuals of generation $n+1$, independently from the rest of the process. We denote by Z_n the number of individuals in generation n: $Z_0 = 1$ and, for all $n \geq 0$,

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)},$$

where the $(\xi_i^{(n)})_{i,n}$ are i.i.d. copies of ξ.

Denote by $m = \mathbb{E}\xi$, then,

$$M_n = m^{-n} Z_n$$

is a martingale.

Example 2.3: The profile of the random Binary Search Tree (cf. Example 1.4)

This exercise is inspired by an article by Chauvin, Klein, Marckert and Rouault (2005): Martingales and Profile of Binary Search Trees, in which martingales are used to get precise information about the shape of the random BST.

Let $\mathcal{T}_n$ be the random BST at time n. For all $n, k \in \mathbb{N}$, let us denote by $N_k(n)$ the number of leaves of $\mathcal{T}_n$ that are at distance k from the root (i.e. at height k in the tree). We denote by $P_n(z)$ the profile polynomial of the BST at time n, given by

$$P_n(z) := \sum_{k\geq 0} U_k(n) z^n.$$

Remark that, if we denote by $|\ell|$ the height of a leaf ℓ of a tree, then

$$P_n(z) = \sum_{\ell \in \mathcal{T}_n} z^{|\ell|}.$$

Can you determine a sequence of rational functions $Z_n(z)$ such that $(M_n := Z_n P_n)_{n\geq 0}$ is a martingale?

Example 2.4: Pólya urn

A Pólya urn is a random process defined by two parameters: an initial composition vector ${}^t(\alpha,\beta)$, and a replacement matrix

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

where α, β, a, b, c and d are integers.

We define the sequence of random vectors $(U(n) = (X_n, Y_n))_{n\geq 0}$ representing the composition of a two-colour urn at time t, meaning that the urn contains X_n red balls and Y_n black balls at time n: The urn contains initially α red balls and β black balls. At each step, we pick up uniformly at random a ball in the urn. If the ball is red, we replace it in the urn together with a additional red balls and b black balls. If it is black, we replace it in the urn together with c red balls and d additional black balls.

Let us assume that the urn is balanced, meaning that $a+b = c+d = S$. It implies that the total number of the urn at time n is $X_n + Y_n = \alpha + \beta + nS$. Let

$$Z_n = \left(1 + \frac{A}{\alpha+\beta}\right)^{-1} \dots \left(1 + \frac{A}{\alpha+\beta+(n-1)S}\right)^{-1},$$

where $A =^t R$ (we assume that all the matrices involved in Z_n are indeed invertible). One can then prove that $(M_n := Z_n U(n))_{n\geq 0}$ is a martingale on $\mathbb{R}^2$ for its natural filtration.

2.2 Stopping theorems

Definition 2.4

A **stopping time** *with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$ is a random variable T such that, for all $n \geq 0$, the event $\{T \leq n\}$ is $\mathcal{F}_n$-measurable.*

Example 2.5: Back to Markov chains

Let $(X_n)_{n\geq 0}$ be a Markov chain on a discrete space E. Let $x \in E$, then $\tau_x := \inf\{n \geq 1 \mid X_n = x\}$ is a stopping time with respect to the natural filtration of $(X_n)_{n\geq 0}$.

Lemma 2.5

For all martingale $(M_n)_{n\geq 0}$, and for all stopping time T, the **stopped process** *$(M_n^T := M_{n\wedge T})_{n\geq 0}$ is a martingale, (where $\wedge$ denotes the minimum between its two terms).*

This lemma is also true for sub-martingales and super-martingales.

Proof. For all $n \geq 1$,

$$\mathbb{E}[M_{n+1}^T | \mathcal{F}_n] = \mathbb{E}[M_{n+1} \mathbb{1}_{T>n} | \mathcal{F}_n] + \mathbb{E}[M_T \mathbb{1}_{T\leq n} | \mathcal{F}_n].$$

Since $\{T > n\}$ and $\{T \leq n\}$ are both $\mathcal{F}_n$-measurable, we get

$$\mathbb{E}[M_{n+1}^T | \mathcal{F}_n] = \mathbb{E}[M_{n+1} | \mathcal{F}_n] \mathbb{1}_{T>n} + M_T \mathbb{1}_{T\leq n} = M_n \mathbb{1}_{T>n} + M_T \mathbb{1}_{T\leq n} = M_n^T.$$

□

Corollary 2.6

For all martingale $(M_n)_{n\geq 0}$ and for all bounded stopping time T, $\mathbb{E}M_T = \mathbb{E}M_0$.

Definition 2.7

Given a stopping time T, we define its σ-algebra

$$\mathcal{F}_T := \{A \in \mathcal{F} \mid \forall n \geq 0, A \cap \{T \leq n\} \in \mathcal{F}_n\}.$$

Of course, one has to check that $\mathcal{F}_T$ is a σ-algebra. We omit this proof.

Proposition 2.8

Let $(M_n)_{n\geq 1}$ be a $\mathcal{F}_n$ martingale and T a finite stopping time. Then M_T is $\mathcal{F}_T$-measurable.

Proposition 2.9

Let T and S two $(\mathcal{F}_n)$-stopping times such that $S \leq T$ almost surely. Then $\mathcal{F}_S \subseteq \mathcal{F}_S$.

Theorem 2.10 (*Doob's stopping theorem*)

Let $(M_n)_{n\geq 0}$ be a martingale, let S and T two bounded stopping times such that, $S \leq T$ almost surely. Then, almost surely,

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S.$$

Proof. It is enough to prove that, for all $A \in \mathcal{F}_S$,$\mathbb{E}[M_T \mathbb{1}_A] = \mathbb{E}[M_S \mathbb{1}_A]$. Let $A \in \mathcal{F}_S$. Define

$$R = S\mathbb{1}_A + T\mathbb{1}_{^cA}.$$

Remark that for all $n \geq 1$,

$$\{R \leq n\} = (A \cap \{S \leq n\}) \cup (A \cap \{T \leq n\}) \in \mathcal{F}_n.,$$

which implies that R is a bounded stopping time. We thus have $\mathbb{E}M_R = \mathbb{E}M_0 = \mathbb{E}M_T$. Since

$$\mathbb{E}M_T = \mathbb{E}[M_T\mathbb{1}_A + M_T\mathbb{1}_{^cA}]$$

$$\mathbb{E}M_R = \mathbb{E}[M_S\mathbb{1}_A + M_T\mathbb{1}_{^cA}],$$

we get

$$\mathbb{E}[M_T\mathbb{1}_A] = \mathbb{E}[M_S\mathbb{1}_A].$$

□

2.3 Doob's inequalities

Proposition 2.11

Let $(M_n)_{n\geq 0}$ a non-negative sub-martingale such that $\mathbb{E}M_0 < +\infty$. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{i\leq n} M_i \geq \alpha) \leq \frac{\mathbb{E}M_n}{\alpha}$$

Proof. Let us denote $A = \{\max_{i\leq n} M_i \geq \alpha\}$, and define, for all $k \geq 0$,

$$A_k := \{\max_{i<k} M_i < \alpha \leq M_k\}.$$

Then

$$\mathbb{E}[M_n\mathbb{1}_A] = \sum_{k=0}^{n} \mathbb{E}[\mathbb{1}_{A_k} M_n] = \sum_{k=0}^{n} \mathbb{E}\left[\mathbb{1}_A \mathbb{E}[M_n|\mathcal{F}_k]\right] \geq \sum_{k=0}^{n} \mathbb{E}\left[\mathbb{1}_A M_k\right] \geq \alpha\mathbb{P}(A).$$

Thus,

$$\mathbb{P}(A) \leq \frac{1}{\alpha}\mathbb{E}[M_n\mathbb{1}_A] \leq \mathbb{E}M_n,$$

since M_n is non-negative. □

The following corollary is a consequence of the following fact: let $(M_n)_{n\geq 0}$ be a martingale and ϕ be a convex function. Then, $(\phi(M_n))_{n\geq 0}$ is a sub-martingale. Apply this propertu to the convex function $(x \mapsto x^2)$ to get the corollary:

Corollary 2.12

Let $(M_n)_{n\geq 0}$ be a square integrable martingale. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{i\leq n} M_i \geq \alpha) \leq \frac{\mathbb{E}M_n^2}{\alpha^2}.$$

2.4 Convergence of martingales

Definition 2.13

A sequence of random variables $(X_n)_{n\geq 0}$ is **bounded in** L^p *if and only if*

$$\sup \mathbb{E}|X_n|^p < +\infty.$$

The sequence is **uniformly integrable** *if and only if*

$$\lim_{x\to+\infty} \mathbb{E}[X_n \mathbb{1}_{X_n > x}] \to 0,$$

when $x \to +\infty$.

Theorem 2.14

A martingale bounded in L^2 converges in L^2, meaning that there exists a random variable M_∞ such that

$$\lim_{n\to+\infty} \mathbb{E}[|M_n - M_\infty|^2] = 0.$$

Example 2.6: Super-critical Galton-Watson process (cf. Example 2.2).

Let us recall that is Z_n is the number of individuals composing the n^{th} generation in a Galton-Watson process, then $M_n = m^{-n}Z_n$ is a martingale. Let us prove that this martingale is bounded in L^2:

$$\mathbb{E}[Z_{n+1}^2|\mathcal{F}_n] = \mathbb{E}\left[\left(\sum_{i=1}^{Z_n} \xi_n^{(i)}\right)^2 \middle| \mathcal{F}_n\right] = Z_n^2(\mathbb{E}\xi)^2 + Z_n \mathtt{Var}\xi.$$

It implies that

$$\mathbb{E}Z_{n+1}^2 = m^2\mathbb{E}Z_n^2 + m^n \mathtt{Var}\xi,$$

and thus,

$$\mathbb{E}M_{n+1}^2 = \mathbb{E}M_n^2 + m^{-n-2}\mathtt{Var}\xi,$$

which implies that the martingale is bounded in L^2 as soon as $m > 1$, i.e, as soon as the process is super-critical, and assuming that ξ is square-integrable.

Theorem 2.15 (*Doob's Theorem*)

Let $(M_n)_{n\geq 0}$ be a sub-martingale such that

$$\sup_{n\geq 0} \mathbb{E}X_n \mathbb{1}_{X_n \geq 0} < +\infty.$$

Then, M_n converges almost surely to an integrable random variable M_∞.

Corollary 2.16

Any martingale bounded in L^1 converges almost surely to an integrable random variable.

It is very important to note that, in the corollary above, even if the martingale is bounded in L^1 and its almost sure limit is integrable, there is, a priori, no convergence in L^1!

The following cororllary is maybe the most useful in practise:

Corollary 2.17

Any non negative super-martingale converges almost surely to an integrable random variable M_∞ and

$$\mathbb{E}M_\infty \le \liminf_{n\to+\infty} \mathbb{E}M_n.$$

Proof. If $(M_n)_{n\ge 0}$ is a super-martingale, then $(-M_n)_{n\ge 0}$ is a sub-martingale. Moreover, it is a non-positive sub-martingale, which implies that

$$\sup_{n\ge 0} \mathbb{E}X_n \mathbb{1}_{X_n\ge 0} = 0 < +\infty.$$

The Doob's Theorem thus applies and $(-M_n)_{n\ge 0}$ converges almost surely to an integrable random variable $-M_\infty$, which concludes the proof. The last inequality is an application of Fatou's lemma. □

Example 2.7: Galton-Watson process (cf. Example 2.2).

Let us recall that is Z_n is the number of individuals composing the n^{th} generation in a Galton-Watson process, then $M_n = m^{-n}Z_n$ is a martingale. It is non-negative and therefore converges almost surely to a random variable M_∞ by Corollary 4.11.

`Exercise: calculate the probability of extinction of a Galton-Watson process.`

Theorem 2.18

Let $(M_n)_{n\ge 0}$ be a martingale. The three following propositions are equivalent:

(i) M_n converges in L^1 to an integrable random variable M_∞;

(ii) $(M_n)_{n\ge 0}$ is bounded in L^1 and there exists a random variable M_∞ such that

$$\mathbb{E}[M_\infty|\mathcal{F}_n] = M_n \qquad (\textit{for all } n \ge 0);$$

(iii) $(M_n)_{n\ge 0}$ is uniformly integrable.

Such a martingale is called **regular**. *It implies in particular that, for all $n \ge 0$, $\mathbb{E}M_n = \mathbb{E}M_\infty$.*

Corollary 2.19

Any martingale bounded in L^p ($p > 1$) converges almost surely and in L^p.

Proof. Let $(M_n)_{n\ge 0}$ be a martingale bounded in L^p: then, for all $x \ge 0$

$$\mathbb{E}[|M_n|^p] \ge \mathbb{E}[|M_n|^p \mathbb{1}_{M_n\ge x}] + \mathbb{E}[|M_n|^p \mathbb{1}_{M_n> x}] \ge \mathbb{E}[M_n^p \mathbb{1}_{M_n\ge x}] \ge x^{p-1}\mathbb{E}[M_n \mathbb{1}_{M_n\ge x}].$$

Since $(M_n)_{n\ge 0}$ is bounded in L^p, there exists a constant $C > 0$ such that

$$\mathbb{E}[M_n \mathbb{1}_{M_n\ge x}] \le \frac{C}{x^{p-1}} \to 0$$

when $x \to +\infty$, because $p > 1$. Thus $(M_n)_{n\ge 0}$ is uniformly-integrable and Theorem 2.18 applies: $(M_n)_{n\ge 0}$ is bounded in L^1 and there exists a random variable M_∞ such that

$$\mathbb{E}[M_\infty|\mathcal{F}_n] = M_n \qquad (\text{for all } n \ge 0).$$

By Fatou's lemma,

$$\mathbb{E}[|M_n|^p] \le \liminf_{n\to+\infty} \mathbb{E}|M_n|^p < +\infty,$$

and, by dominated convergence, M_n converges to M_∞ in L^p. □

3 Continuous time Markov processes

The aim of this section is not to introduce Markov processes in full generality: we will only focus on jump Markov processes and their main application to queuing theory.

3.1 Definitions

Let E be a discrete state space. Let $(Z_n)_{n\geq 0}$ and $(T_n)_{n\geq 0}$ be two sequences of random variables such that $0 = T_0 \leq T_2 \leq \ldots$, $T_n \to +\infty$ when $n \to +\infty$ and $Z_n \in E$ for all $n \geq 0$.

Definition 3.1

The random function

$$X_t := \sum_{n\geq 0} Z_n \mathbb{1}_{[T_n, T_{n+1}[}(t)$$

is called the **random jump function** *associated to the sequences* $(Z_n)_{n\geq 0}$ *and* $(T_n)_{n\geq 0}$.

Definition 3.2

A random jump function $(X_t)_{t\geq 0}$ *is a* **jump Markov process** *if, for all* $0 < s < t$, *for all* $n \geq 0$, *for all* $t_0 < t_1, \ldots, t_n < s$, *for all* $x_0, x_1, \ldots, x_n, x, y \in E$,

$$\mathbb{P}(X_t = y \mid X_{t_0} = x_0, \ldots, X_{t_n} = x_n \text{ and } X_s = x) = \mathbb{P}(X_t = y \mid X_s = x).$$

If, in addition, $P(X_t = y \mid X_s = x)$ *only depends on* x, y *and* $(t - s)$, *then the jump Markov process is called* **homogeneous**.

In the following, we will only consider **homogeneous jump Markov processes**, and we will denote

$$P_{x,y}(t-s) := \mathbb{P}(X_t = y \mid X_s = x).$$

For all $t \geq 0$, the matrix $P(t) = (P_{x,y}(t))_{x,y\in E}$ is the **transition matrix** of the process $(X_t)_{t\geq 0}$ at time t. We denote by $(\mu(t))$ the law of the random variable X_t, for all $t \geq 0$.

Proposition 3.3

Let $(X_t)_{t\geq 0}$ *be a (homogeneous) Markov jump process on* E, *with initial law* $\mu(0) = \mu$ *and transition matrix* $(P(t))_{t\geq 0}$. *Then, for all* $0 < s < t$,

(i) $\mu(t) = \mu(t)P(t)$

(ii) $P(s+t) = P(s)P(t)$ *(semi-group condition)*

Example 3.1: Poisson process.

A Poisson process $(N_t)_{t\geq 0}$ is a Markov jump process on $\mathbb{N}$, with transition matrix

$$P_{x,y}(t) = \begin{cases} \dfrac{(\lambda t)^{y-x}}{(y-x)!} \mathrm{e}^{-\lambda t} & \text{if } y \geq x, \\ 0 & \textit{otherwise.} \end{cases}$$

Example 3.2: Let $(T_n)_{n\geq 0}$ be a Poisson point process on $[0, +\infty[$ with intensity λ and let $(Z_n)_{n\geq 0}$ be a discrete time Markov chain on E, of transition matrix P, independent of $(T_n)_{n\geq 0}$. Then, the continuous time process

$$X_t := \sum_{n\geq 0} Z_n \mathbb{1}_{[T_n, T_{n+1}[}$$

is a Markov jump process. Can you determine its transition matrix?

The semi-group property tells us that the transition matrix $(P(t))_{t\geq 0}$ is determined by its values for small $t \geq 0$. Said differently, it is determined by its derivative at 0:

Definition 3.4

Let $(P(t))_{t\geq 0}$ be the transition matrix of a Markov jump process $(X_t)_{t\geq 0}$. Then, there exists $Q = (Q_{x,y})_{x,y\in E}$ called the **generator** *of $(X_t)_{t\geq 0}$, such that*

(i) $Q_{x,y} \geq 0$ *if* $x \neq y$,

(ii) $Q_{x,x} = -\sum_{y\neq x} Q_{x,y} \leq 0$,

(iii) $P_{x,y}(h) = hQ_{x,y} + o(h)$ *when* $h \to 0$, *if* $x \neq y$,

(iv) $P_{x,x}(h) = 1 + hQ_{x,x} + o(h)$ *when* $h \to 0$.

One can see $Q_{x,y}$ as the rate with which the Markov jump process will jump from site x to site y.

Theorem 3.5

Markov property Let $(X_t)_{t\geq 0}$ be a jump Markov process of generator Q. For all real t_0, the process $(X_{t_0+t})_{t\geq 0}$ is a Markov process of initial law $\delta_{X_{t_0}}$.

If we forget time and just focus on the successive positions of the process, we exhibit the underlying Markov chain of the process. Let us denote by τ_n the time of the n^{th} jump of the process: then, the discrete time process $M_n := X_{\tau_n}$ is a Markov chain and its transition matrix $P = (p_{i,j})_{i,j\in E}$ is given by

$$p_{x,y} = \begin{cases} \frac{Q_{x,y}}{q_x} & \text{if} \quad i \neq j \\ 0 & \text{if} \quad i = j \end{cases},$$

where $q_x := -Q_{x,x}$ for all $x \in E$.

3.2 Ergodicity

A jump Markov process is **irreducible** as soon as its underlying Markov chain is irreducible. It implies that, for all $t > 0$, for all $x, y \in E$, $P_{x,y}(t) > 0$. A state $x \in E$ is recurrent (resp. transient) for the Markov jump process $(X_t)_{t\geq 0}$ if it is recurrent (resp. transient) for its underlying Markov chain.

Theorem 3.6

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and recurrent, with generator $Q = (Q_{x,y})_{x,y\in E}$ and transition matrix $(P(t))_{t\geq 0}$. Then, there exists a unique measure (up to a constant factor) π such that $\pi Q = 0$ and $\pi P(t) = \pi$ for all $t \geq 0$.

Definition 3.7

For all $x \in E$, we denote by $\tau_x := \inf\{t > 0 \mid X_t = x\}$. A state $x \in E$ is **positive recurrent** *(resp.* **null recurrent***) for $(X_t)_{t\geq 0}$ if x is recurrent and if $\mathbb{E}_x\tau_x < +\infty$ (resp. $\mathbb{E}_x\tau_x = +\infty$)*

Theorem 3.8

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and recurrent. Then, the following assumptions are equivalent:

(i) $x \in E$ *is positive recurrent,*

(ii) all states are positive recurrent,

(iii) there exists a unique invariant probability distribution π.

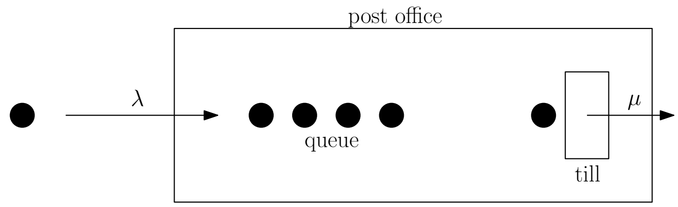


FIGURE 4 – *The $M/M/1$ queue.*

If these assumptions are verified, then, for all $x \in E$,

$$\mathbb{E}_x \tau_x = \frac{1}{\pi_x q_x}.$$

Theorem 3.9

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and positive recurrent. Denote by π its invariant probability. Then, for all bounded function $f : E \to \mathbb{R}$, almost surely, when $t \to +\infty$,

$$\frac{1}{t}\int_0^t f(X_s)ds \to \sum_{x\in E} f(x)\pi_x.$$

Proposition 3.10

Let $(X_t)_{t\geq 0}$ be a Markov jump process, irreducible and positive recurrent. Denote by π its invariant probability. Then, for all probability distribution μ on E, for all $x \in E$, asymptotically when $t \to +\infty$,

$$(\mu P(t))_x \to \pi_x.$$

3.3 Queues

The example we will study in the whole section it the queuing theory. It is very important in computer science, since it permits to model routers activity.

The idea is the following: in my post office, there are N tills. People enter the post office according to a Poisson process of rate λ, meaning that the interval between a client and the next one is exponentially distributed with parameter λ, independently from the rest of the process. The time needed to serve a client is exponentially distributed with parameter μ, independently from the rest of the process.

When a client enters the post office: either all tills are occupied and he joins the queue, or one till is free, and he bigins to be served as soon as he enters.

This model is usually called $M/M/N$ meaning that the arrivals and service times are exponentially distributed, with respective parameters λ and μ, and that there are N tills.

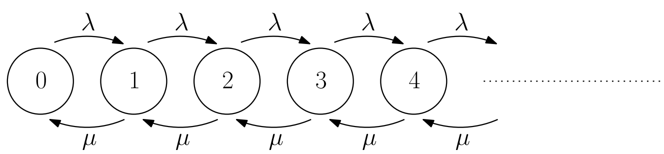
The question is the following: do you need to add more tills so that the length of the queue does not explode? Quite an important question for router, post office or server management.

Example 3.3: The $M/M/1$ queue (cf. Figure 4)

Let us first focus on the case where there is a unique till in the post office. Let X_t be the number of clients inside the post office (queue + till) at time t. Then $(X_t)_{t\geq 0}$ is indeed a Markov process and its generator is the following infinite matrix:

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

This information can be represented as follows:



It is possible to prove that $\pi_x := \rho^x(1-\rho)$, where $\rho := {}^{\lambda}/_{\mu}$, is an invariant probability of the queue, as soon as $\rho < 1$. If $\rho \geq 1$, then, the queue admits no invariant probability and is thus transient. It means that our queue will explode. Can you calculate the probability that a newly arrived client will have to queue before being served?

Exercise 3.1: Can you give the generator of the queue $M/M/\infty$?

Example 3.4: In the queues described above, the $M/M/N$, the capacity of the queue is infinite, meaning that the queue can become arbitrarily large. One can also describe queues with finite capacity K: the queues $M/M/N/K$. It behaves as the $M/M/N$, except that when the queue is full (i.e. contains K clients), any client arriving to the shop cannot enter the shop and evaporates.

Can you give the generator of such a queue? What is its invariant probability?

4 Continuous time martingales

4.1 Definitions and first properties

Let $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space.

Definition 4.1

A continuous time process $(M_t)_{t\geq 0}$ is a martingale for the filtration $(\mathcal{F}_t)_{t\geq 0}$ if and only if, for all $t \geq 0$,

(i) M_t is $\mathcal{F}_t$-measurable;

(ii) M_t is integrable; and

(iii) for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] = M_s$.

Definition 4.2

Replacing (iii) in the above definition by

- *for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] \leq M_s$ gives the definition of a* **super-martingale**.
- *for all $s < t$, $\mathbb{E}[M_t|\mathcal{F}_s] \geq M_s$ gives the definition of a* **sub-martingale**.

Example 4.1: The Yule tree (cf. Figure 5)

Let us consider the stochastic process $(Y_t)_{t\geq 0}$ defined as follows. At time zero, there is one particle in the system: $Y_0 = 1$. Each particle dies and gives birth to two new particles after an exponentially distributed random time, independently from the other particles. Let us denote by Y_t the number of particles alive at time t.

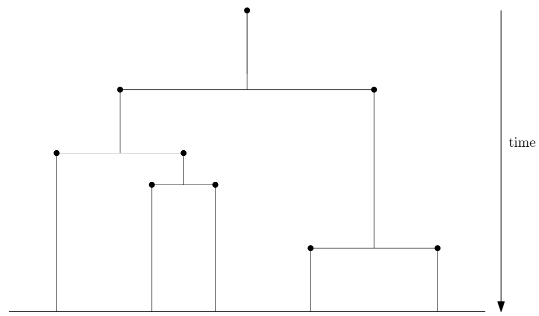


Figure 5 – *A realisation of the Yule tree*

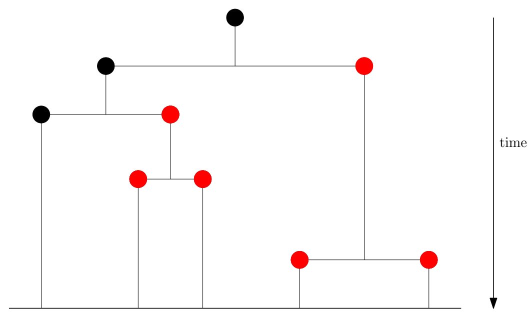


Figure 6 – *A realisation of the multi-type branching process defined by the initial composition* ${}^t(0,1)$ *and the replacement matrix* $R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

Can you find $(m_t)_{t\geq 0}$ a function such that $M_t := m_t^{-1} Y_t$ is a martingale?

Example 4.2: Multi-type branching process (cf. Figure 6)

A multi-type branching process is the embedding in continuous time of a Pólya urn. It is defined by an initial composition $U(0) =^t (\alpha, \beta)$ and a replacement matrix

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The vector composition of the urn at time t is given by $U(t) =^t (X_t, Y_t)$, where X_t is the number of red balls and Y_t the number of black balls at time t in the urn. Each ball in the urn will split after an exponentially distributed random time into

- $a+1$ red balls and b black balls if it is a red ball;
- or c red balls and $d+1$ black balls if it is a black ball,

independently for the other balls.

Assume that the replacement matrix is balanced: $a+b = c+d = S$. What can you say about the total number of balls in the urn at time t? Can you prove that $M_t := \mathsf{e}^{-tA} U(t)$ is a vector valued martingale, where $A =^t R$?

4.2 Stopping times

Definition 4.3

A random variable T is a stopping time for the filtration $(\mathcal{F}_t)_{t\geq 0}$ if and only if, for all $t \geq 0$, the event $\{T \leq t\}$ is $\mathcal{F}_t$-measurable.

Lemma 4.4

For all martingale $(M_t)_{t\geq 0}$, and for all stopping time T, the **stopped process** $(M_t^T := M_{t\wedge T})_{t\geq 0}$ *is a martingale, (where $\wedge$ denotes the minimum between its two terms).*

Theorem 4.5

Stopping theorem Let $(M_t)_{t\geq 0}$ be a martingale, let S and T two bounded stopping times such that, $S \leq T$ almost surely. Then, almost surely,

$$\mathbb{E}[M_T|\mathcal{F}_S] = M_S.$$

4.3 Doob's inequalities

Proposition 4.6

Let $(M_t)_{t\geq 0}$ a non-negative sub-martingale such that $\mathbb{E}M_0 < +\infty$. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{s\leq t} M_s \geq \alpha) \leq \frac{\mathbb{E}M_t}{\alpha}$$

Corollary 4.7

Let $(M_t)_{t\geq 0}$ be a square integrable martingale. Then, for all $\alpha > 0$,

$$\mathbb{P}(\max_{s\leq t} M_s \geq \alpha) \leq \frac{\mathbb{E}M_t^2}{\alpha^2}.$$

4.4 Convergence of continuous time martingales

Definition 4.8

A sequence of random variables $(X_t)_{n\geq 0}$ is **bounded in** L^p *if and only if*

$$\sup_{t\geq 0} \mathbb{E}|X_t|^p < +\infty.$$

The sequence is uniformly integrable if and only if

$$\lim_{x\to+\infty} \sup_{t\geq 0} \mathbb{E}[X_t \mathbb{1}_{X_t > x}] \to 0,$$

when $x \to +\infty$.

Theorem 4.9

A martingale bounded in L^2 converges in L^2, meaning that there exists a random variable M_∞ such that

$$\lim_{t\to+\infty} \mathbb{E}[|M_t - M_\infty|^2] = 0.$$

Theorem 4.10 (*Doob's Theorem*)

Let $(M_t)_{t\geq 0}$ be a sub-martingale such that

$$\sup_{t\geq 0} \mathbb{E}X_t \mathbb{1}_{X_t\geq 0} < +\infty.$$

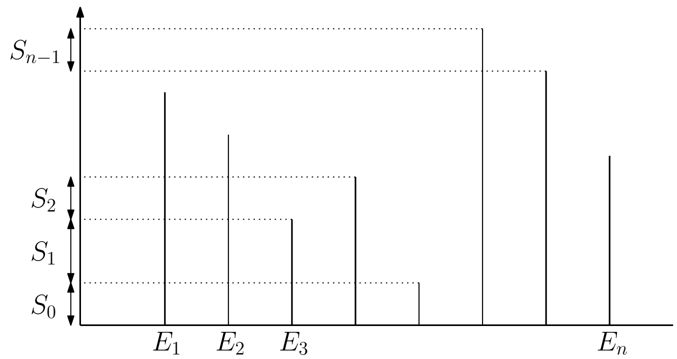


FIGURE 7 – *The n i.i.d. random variables $E_1, \ldots, E_n$ are represented by the length of the vertical sticks. The S_i are independent random variables exponentially distributed, of respective parameters $n-i$.*

Then, M_t converges almost surely to an integrable random variable M_∞.

Corollary 4.11

All non negative super-martingale $(M_t)_{t\geq 0}$ converges almost surely to an integrable random variable M_∞ and

$$\mathbb{E}M_\infty \leq \liminf_{t\to+\infty} \mathbb{E}M_t.$$

Example 4.3: The Yule tree martingale (cf. Example 4.1)

The process $(M_t := \mathrm{e}^{-t}Y_t$ is a non negative martingale and thus converges almost surely to a limit random variable W. Let us prove that this random variable is exponentially distributed.

For all $t \geq 0$, $\mathbb{P}(Y_t \geq n) = \mathbb{P}(\tau_n \leq t)$ where τ_n is the time of the n^{th} split in the Yule process. Remark that, by definition, $\tau_n = \sum_{i=1}^n T_i$ where T_i is exponentially distributed of parameter i and the $(T_i)_{i=1..n}$ are independent of each other.

Let us consider $E_1, \ldots, E_n$ being n i.i.d. random variable exponentially distributed of parameter 1. Let us denote by m_n their maximum. Remark that $m_n = \sum_{i=0}^{n-1} S_i$ where S_i is exponentially distributed of parameter $n-i$ and the $(S_i)_{i=1..n}$ are independent of each other (see Figure 7).

Thus,

$$\mathbb{P}(\tau_n \leq t) = \mathbb{P}(m_n \leq t) = \mathbb{P}(E_i \leq t\ \forall 1 \leq i \leq n) = (1-\mathrm{e}^{-t})^n.$$

It implies that, for all $t \geq 0$, for all $x \geq 0$,

$$\mathbb{P}(M_t \geq x) = \mathbb{P}(Y_t \geq x\mathrm{e}^t) = (1-\mathrm{e}^{-t})^{x\mathrm{e}^t} \to \mathrm{e}^{-x},$$

when $t \to +\infty$. Thus, for all $x \geq 0$,

$$\mathbb{P}(W \geq x) = \mathrm{e}^{-x},$$

and W is exponentially distributed of parameter 1.

Theorem 4.12

Let $(M_t)_{t\geq 0}$ be a martingale. The three following propositions are equivalent:

(i) *M_t converges in L^1 to an integrable random variable M_∞;*

(ii) *$(M_t)_{t\geq 0}$ is bounded in L^1 and there exists a random variable M_∞ such that*

$$\mathbb{E}[M_\infty|\mathcal{F}_t] = M_t \qquad (\textit{for all } t \geq 0);$$

(iii) $(M_t)_{t\geq 0}$ *is uniformly integrable.*

Such a martingale is called **regular**. *It implies in particular that, for all* $t \geq 0$, $\mathbb{E}M_t = \mathbb{E}M_\infty$.

Corollary 4.13

Any martingale bounded in L^p $(p > 1)$ *converges in* L^p.

5 Exercises

Exercise 5.1: Simple random walk

Let us consider the biased random walk on $\mathbb{Z}$ defined as follows: choose $p \in (0,1)$ and denote $q = 1-p$, when the walker is in state x, it jumps to $x+1$ with probability p and to $x-1$ with probability q.

(1) Prove that the unbiased random walk on $\mathbb{Z}$ is recurrent but has no invariant probability: it is thus null recurrent.

(2) A gambler enters a casino with a GBP and begins to play *heads or tails* with the casino. The casino has b GBP when the gambler begins to play. The coin is biased and gives *heads* with probability p and *tails* with probability q. The gambler gives one pound to the casino when its *heads* and the casino gives him one pound when its *tails*. The game ends when either the gambler or the casino is ruined. What is the probability that the gambler gets ruined?

Hint: Denote by X_n *the wealth of the gambler at time* n, $\tau_0 := \inf\{s \geq 0 \mid X_s = 0\}$ *and* $\tau_{a+b} := \inf\{s \geq 0 \mid X_s = a+b\}$. *It is a good idea to define* $u_x := \mathbb{P}(\tau_0 < \tau_{a+b} \mid X_0 = x\}$, *for all* $x \in \mathbb{Z}$.

Exercise 5.2: The original Pólya urns

Consider the Pólya urn with initial composition vector ${}^t(1,1)$ and replacement matrix I_2. Let us denote by ${}^t(X_n, Y_n)$ the composition vector of the urn process at time n.

(1) Prove that X_n is a Markov chain and give its transition probabilities.

(2) Let $\bar{X}_n = \frac{X_n}{X_n+Y_n} = \frac{X_n}{n+2}$ be the proportion of balls of type 1 in the urn at time n. Prove that $(\bar{X}_n)_{n\geq 0}$ is a martingale.

(3) Prove that $(\bar{X}_n)_{n\geq 0}$ converges almost surely and in L^1 to a limit X_∞.

(4) Let
$$Z_n^{(k)} := \frac{X_n(X_n+1)\cdots(X_n+k-1)}{(n+2)(n+3)\cdots(n+k+1)}.$$
Prove that $(Z_n^{(k)})_{n\geq 0}$ is a martingale for all $k \geq 1$.

(5) Prove that, for all $k \geq 1$, $\mathbb{E}X_\infty^k = \mathbb{E}Z_0^{(k)} = \frac{1}{k+1}$ and deduce from it that X_∞ has uniform law on $[0,1]$.

Exercise 5.3: Queue with finite capacity

Let us study the queue $M/M/1/K$, corresponding to a queue with arrivals of rate λ, service times of rate μ, with 1 tills and K maximum places in the queue. The number of costumers in the post office is a Markov jump process on $\{0,\ldots,K\}$:

(1) write its generator Q and its transition matrix $(P(t))_{t\geq 0}$;

(2) convince yourself that the process is irreducible, and calculate its invariant probability;

(3) what is the average number of costumers in the system?