Discrete-time Martingales

François Chapon

Contents

1. Introduction

By way of introduction, we quote Joseph Leo Doob: ”Martingale theory illustrates the history of mathematical probability: the basic definitions are inspired by crude notions of gambling, but the theory has become a sophisticated tool of modern abstract mathematics, drawing from and contributing to other fields.” [Doob, *What is a Martingale?*, The American Mathematical Monthly, 78(5), (1971).]

2. Definitions

2.1. Filtrations.

Definition 2.1. *A sequence of random variables $X = (X_n)_{n\geq 0}$ defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a stochastic process.*

Definition 2.2. *A filtration on $(\Omega, \mathcal{F}, \mathbb{P})$ is a non-decreasing sequence of sub-σ-algebras of $\mathcal{F}$:*

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_n \subset \cdots \subset \mathcal{F}.$$

One says that $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$ is a filtered probability space.

Example 2.1. Let $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1[, \mathcal{B}([0,1[), \lambda)$, where λ is Lebesgue measure. The filtration $(\mathcal{F}_n)_{n\geq 0}$ defined by

$$\mathcal{F}_n = \sigma\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right[, i = 0, \ldots, 2^n - 1\right), \quad n \geq 0$$

is called the dyadic filtration.

If the parameter n denotes time, then $\mathcal{F}_n$ is interpreted as available information up to time n.

Example 2.2. For a stochastic process $(X_n)_{n\geq 0}$, we define its natural filtration $\mathcal{F}^X = (\mathcal{F}^X_n)_{n\geq 0}$ by: for all $n \geq 0$,

$$\mathcal{F}^X_n = \sigma(X_0, X_1, \cdots, X_n),$$

which is the smallest σ-algebra such that $X_0, \ldots, X_n$ are measurable.

Definition 2.3. *We say that a stochastic process $X = (X_n)_{n\geq 0}$ is adapted to the filtration $(\mathcal{F}_n)_{n\geq 0}$, if for all $n \geq 0$, X_n is $\mathcal{F}_n$-measurable. We say that a stochastic process $(X_n)_{n\geq 0}$ is adapted if it is adapted to some filtration.*

A stochastic process is obviously adapted to its natural filtration.

Remark 2.1. If $(\mathcal{F}_n)_{n\geq 0}$ and $(\mathcal{G}_n)_{n\geq 0}$ are two filtrations such that $\mathcal{G}_n \subset \mathcal{F}_n$ for all $n \geq 0$, and if $(X_n)_{n\geq 0}$ is adapted to $(\mathcal{G}_n)_{n\geq 0}$, then $(X_n)_{n\geq 0}$ is adapted to $(\mathcal{F}_n)_{n\geq 0}$.

2.2. Martingales. In the sequel, we fix a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_n)_{n\geq 0}, \mathbb{P})$.

Definition 2.4. *Let $(X_n)_{n\geq 0}$ be an adapted stochastic process such that $\mathbb{E}(|X_n|) < \infty$, for all $n \geq 0$. The process $(X_n)_{n\geq 0}$ is said to be*

- *a martingale: if for any $n \geq 0$,*
$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = X_n;$$
- *a supermartingale: if for any $n \geq 0$,*
$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \leq X_n;$$
- *a submartingale: if for any $n \geq 0$,*
$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \geq X_n.$$

Remark 2.2. An easy consequence of the definition, is that if $(X_n)_{n\geq 0}$ is a martingale, then for all $m \geq n$, one has

$$\mathbb{E}(X_m \mid \mathcal{F}_n) = X_n.$$

We use induction: this is true for $m = n+1$ by definition. Suppose it is true for some $m > n+1$. Then

$$\mathbb{E}(X_{m+1} \mid \mathcal{F}_n) = \mathbb{E}\left(\mathbb{E}(X_{m+1} \mid \mathcal{F}_m) \mid \mathcal{F}_n\right),$$

by the tower property of conditional expectations, since $\mathcal{F}_n \subset \mathcal{F}_m$. Hence, by the martingale property,

$$\mathbb{E}(X_{m+1} \mid \mathcal{F}_n) = \mathbb{E}(X_m \mid \mathcal{F}_n) = X_n,$$

by induction hypothesis. Moreover, by taking the expectation, one has that the sequence of expectations $(\mathbb{E}(X_n))_{n\geq 0}$ is constant.

An analogous statement holds for a supermartingale and a submartingale by replacing the sign $=$ by $\leq$ and $\geq$ respectively: if $(X_n)_{n\geq 0}$ is a supermatingale, then

$$\mathbb{E}(X_m \mid \mathcal{F}_n) \leq X_n, \quad \text{for all } m \geq n,$$

and $\mathbb{E}(X_n)$ is non-increasing, while if $(X_n)_{n\geq 0}$ is a submatingale, then

$$\mathbb{E}(X_m \mid \mathcal{F}_n) \geq X_n, \quad \text{for all } m \geq n,$$

and $\mathbb{E}(X_n)$ is non-decreasing.

Remark 2.3. If $(X_n)_{n\geq 0}$ is a submartingale, then $(-X_n)_{n\geq 0}$ is a supermartingale. Thus, most of the results concerning supermartingales are immediately deduced from the case of submartingales (and vice versa). Furthermore, $(X_n)_{n\geq 0}$ is a martingale if and only if it is both a supermartingale and a submartingale

Remark 2.4. If one interprets the random variable X_n as a gambler's holdings at time n, then the σ-algebra $\mathcal{F}_n$ is the information available to the gambler up to time n, and in particular the results of all the previous games. The martingale property

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = X_n$$

thus reflects the fact that the mean value of the gain at time $n+1$, when the past is known up to time n, is exactly X_n: on average the player neither loses nor wins. A martingale thus corresponds to a fair game. Similarly, a supermartingale corresponds to an unfair game (unfair to the player), and a submartingale to a favourable game.

Example 2.3. The random walk on $\mathbb{Z}$. Let $(\varepsilon_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, with Bernoulli distribution on $\{-1, 1\}$ with parameter p ($p \in (0,1)$). Let $x \in \mathbb{Z}$ and define the random walk $(X_n)_{n\geq 0}$ by $X_0 = x$ and for $n \geq 1$,

$$X_n = x + \varepsilon_1 + \cdots + \varepsilon_n.$$

We consider the filtration $(\mathcal{F}_n)_{n\geq 0}$ defined by

$$\mathcal{F}_0 = \{\varnothing, \Omega\} \quad \text{and} \quad \mathcal{F}_n = \sigma(X_1, \ldots, X_n), \quad n \geq 1.$$

Note that it is also the natural filtration of $(\varepsilon_n)_{n\geq 1}$ (exercise). Then,

$$\mathbb{E}(X_{n+1} \mid \mathcal{F}_n) = \mathbb{E}(X_n + \varepsilon_{n+1} \mid \mathcal{F}_n) = X_n + \mathbb{E}(\varepsilon_{n+1} \mid \mathcal{F}_n) = X_n + \mathbb{E}(\varepsilon_{n+1}),$$

since ε_{n+1} is independent of $\mathcal{F}_n$. Since $\mathbb{E}(\varepsilon_{n+1}) = 2p - 1$, we have that $(X_n)_{n\geq 0}$ is a martingale if $p = \frac{1}{2}$, a supermartingale if $p < \frac{1}{2}$, and a submartingale if $p > \frac{1}{2}$.

Example 2.4. Let $(Y_n)_{n\geq 0}$ be a sequence of bounded i.i.d. random variables with $\mathbb{E}(Y_0) = 1$. Define $Z_n = \prod_{k=0}^n Y_k$. Then, the stochastic process $(Z_n)_{n\geq 0}$ is a $(\mathcal{F}_n^Y)_{n\geq 0}$-martingale: for all $n \geq 0$,

$$\mathbb{E}(Z_{n+1} \mid \mathcal{F}_n^Y) = \mathbb{E}\left(Y_{n+1} Z_n \mid \mathcal{F}_n^Y\right) = Z_n \, \mathbb{E}(Y_{n+1}) = Z_n.$$

Not all martingales come from an i.i.d. sequence of random variables, as the following example shows:

Example 2.5 (Closed martingale). Let $Y \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and let $(\mathcal{F}_n)_{n\geq 0}$ be a filtration. Define, for all $n \geq 0$,

$$X_n = \mathbb{E}(Y \,|\, \mathcal{F}_n).$$

Then, $(X_n)_{n\geq 0}$ is obviously $(\mathcal{F}_n)_{n\geq 0}$-adapted, and is a martingale: for all $n \geq 0$,

$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}\left(\mathbb{E}(Y \,|\, \mathcal{F}_{n+1}) \,|\, \mathcal{F}_n\right) = \mathbb{E}(Y \,|\, \mathcal{F}_n) = X_n,$$

where we use the tower property of conditional expectations. A martingale of this form is said to be a closed martingale.

2.3. Martingale transforms.

Proposition 2.1. *Let $X = (X_n)_{n\geq 0}$ be a $(\mathcal{F}_n)_{n\geq 0}$-martinagle (resp. a supermartingale, resp. a submartingale). Let $(\mathcal{G}_n)_{n\geq 0}$ be a smaller filtration (i.e. $\mathcal{G}_n \subset \mathcal{F}_n$ for all $n \geq 0$) such that X is adapted to $(\mathcal{G}_n)_{n\geq 0}$. Then X is a $(\mathcal{G}_n)_{n\geq 0}$-martinagle (resp. a supermartingale, resp. a submartingale).*

Proof. It suffices to apply the tower property of conditional expectations:

$$\mathbb{E}(X_{n+1} \,|\, \mathcal{G}_n) = \mathbb{E}\left(\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) \,|\, \mathcal{G}_n\right) = \mathbb{E}(X_n \,|\, \mathcal{G}_n) = X_n,$$

since X_n is $\mathcal{G}_n$-measurable. □

In particular, a $(\mathcal{F}_n)_{n\geq 0}$-(super,sub)-martingale is a (super,sub)-martingale relative to its natural filtration.

Proposition 2.2. *Let $\varphi\colon \mathbb{R} \to [0, +\infty[$ be a convex function. If $(X_n)_{n\geq 0}$ is a martingale, then $(\varphi(X_n))_{n\geq 0}$ is a submartingale. If moreover φ is non-decreasing, and if $(X_n)_{n\geq 0}$ is a submartingale, then $(\varphi(X_n))_{n\geq 0}$ is a submartingale.*

Remark 2.5. In particular, if $(X_n)_{n\geq 0}$ is a martingale, $(|X_n|)_{n\geq 0}$ and $(X_n^+)_{n\geq 0}$ are submartingales.[1] If moreover $\mathbb{E}(X_n^2) < \infty$, for all $n \geq 0$, then $(X_n^2)_{n\geq 0}$ is a submartingale.

Proof. This follows immediately from the conditional Jensen inequality. □

Definition 2.5. *A stochastic process $(H_n)_{n\geq 1}$ is called a predictable process if H_n is $\mathcal{F}_{n-1}$-measurable for each $n \geq 1$.*

Example 2.6. Show that a predictable martingale X satisfies: $X_n = X_0$ a.s. for all $n \geq 0$.

The following transform is called a martingale transform of X. It can be seen as a discrete stochastic integral of X.

Proposition 2.3. *Let $X = (X_n)_{n\geq 0}$ be an adapted process and $(H_n)_{n\geq 0}$ a bounded predictable process. Define the process $(H \cdot X)$ by $(H \cdot X)_0 = 0$, and for all $n \geq 1$,*

$$(H \cdot X)_n = \sum_{k=1}^{n} H_k \left(X_k - X_{k-1}\right).$$

Then,

(i) if (X_n) is a martingale, then $((H \cdot X)_n)$ is a martingale;
(ii) if (X_n) is a supermartingale (resp. a submartingale) and $H_n \geq 0$ for all n, then $((H \cdot X)_n)$ is a supermartingale (resp. a submartingale).

[1] Recall that f^+ denotes the positive part of the function f, i.e. $f^+(x) = \max(f(x), 0)$, and f^- its negative part, i.e. $f^-(x) = \max(-f(x), 0)$.

Proof. (i) Since H_n is bounded, $(H \cdot X)_n$ is in L^1. Since (H_n) is predictable and (X_n) is adapted, $(H \cdot X)_n$ is also adapted. Now,

$$\begin{aligned}\mathbb{E}\left((H \cdot X)_{n+1} - (H \cdot X)_n \mid \mathcal{F}_n\right) &= \mathbb{E}\left((H_{n+1}(X_{n+1} - X_n) \mid \mathcal{F}_n\right) \\ &= H_{n+1}\,\mathbb{E}\left((X_{n+1} - X_n) \mid \mathcal{F}_n\right),\end{aligned}$$

since H_{n+1} is $\mathcal{F}_n$-measurable. Hence, the conclusion follows since (X_n) is a martingale.

Item (ii) is left as an exercise. □

Remark 2.6. In a casino gambling game, if $X_n - X_{n-1}$ is the gambler's gain in the n-th round, one can interpret $(H_n)_n$ as a gambling strategy, that is H_n is the amount the player bet in the n-th round. Hence, $(H_n)_n$ must be predictable: the value of H_n has to be decided at time $n-1$, before the result of X_n is known. The martingale transform $(H \cdot X)_n$ is thus the total gain of the gambler at time n.

Example 2.7 (St. Petersburg game). Consider the following roulette game in a casino. They are 37 pockets in the roulette, 18 of which are red, 18 are black and one is green. If the player bets on "red" and the balls lands in a red pocket, the gain is the double of the bet. Otherwise, the bet is lost. The chance of winning is thus $p = \frac{18}{37} < \frac{1}{2}$. Consider a sequence $(\varepsilon_n)_{n\geq 1}$ of i.i.d random variable with distribution $p\delta_1 + (1-p)\delta_{-1}$. Then, $(X_n)_{n\geq 1}$ where $X_n = \varepsilon_1 + \cdots + \varepsilon_n$, is a supermartingale. If the gambler bets a random amount of H_n at the n-th round, the total amount of profit at time n is then

$$G_n = \sum_{k=1}^{n} H_k \varepsilon_k = (H \cdot X)_n.$$

The gambler adopts the following strategy: in the first round, the bet is $H_1 = 1$. If she wins, she leaves the casino. If she loses, she doubles the stake: $H_2 = 2$. If she wins, she leaves the casino, otherwise she doubles again the stake, and so on. The gambling strategy $(H_n)_{n\geq 1}$ is thus defined by the predictable process:

$$\begin{aligned} H_n &= \begin{cases} 0 & \text{if there is } k \in \{1, \ldots, n-1\} \text{ such that } \varepsilon_k = 1, \\ 2^{n-1} & \text{else} \end{cases} \\ &= 2^{n-1} \mathbb{1}_{\{\varepsilon_1 = -1, \ldots, \varepsilon_{n-1} = -1\}}. \end{aligned}$$

Since $(G_n)_n$ is a supermartingale, we have $\mathbb{E}(G_n) \leq \mathbb{E}(G_1) < 0$. You can't beat the system...

3. Stopping times

3.1. Definition.

Definition 3.1. *A random variable $T\colon \Omega \to \mathbb{N} \cup \{+\infty\}$ is called a stopping time (with respect to the filtration $(\mathcal{F}_n)_{n\geq 0}$) if for all $n \geq 0$,*

$$\{T \leq n\} \in \mathcal{F}_n.$$

Remark 3.1. Since $\{T = n\} = \{T \leq n\} \setminus \{T \leq n-1\}$, T is a stopping time if and only if for all $n \geq 0$,

$$\{T = n\} \in \mathcal{F}_n.$$

Remark 3.2. A stopping time is thus a random time, which can be interpreted as a stopping rule for deciding whether to continue or stop a process on the basis of the present information and past events, for instance playing until you go broke or you break the bank, etc. . .

Example 3.1. (i) If $T = n$ a.s., then clearly T is a stopping time.

(ii) Let $(X_n)_{n\geq 0}$ be an adapted stochastic process, and consider the first time X_n reaches the borel set A:
$$T_A = \inf\{n \geq 0 \,|\, X_n \in A\},$$
with the convention that $\inf \varnothing = +\infty$. It is called the hitting time of A. Then T_A is a stopping time. Indeed,
$$\begin{aligned}\{T_A = n\} &= \{X_0 \notin A, X_1 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}\\ &= \bigcap_{k=0}^{n-1}\{X_k \notin A\} \cap \{X_n \in A\} \in \mathcal{F}_n.\end{aligned}$$

(iii) Show that $\tau_A = \sup\{n \geq 1 \,|\, X_n \in A\}$ the last passage time in A is not a stopping time.

Recall the notations: $x \wedge y = \inf(x, y)$ and $x \vee y = \max(x, y)$.

Proposition 3.1. *If S and T are two stopping times, then $S \wedge T$, $S \vee T$ and $S + Y$ are also stopping times.*

Proof. Writing
$$\{S \wedge T \leq n\} = \{S \leq n\} \cup \{T \leq n\}$$
and
$$\{S \vee T \leq n\} = \{S \leq n\} \cap \{T \leq n\}$$
gives the result for $S \wedge T$ and $S \vee T$. For $S + T$, we write:
$$\{S + T \leq n\} = \bigcup_{k\leq n}\{S = k\} \cap \{T \leq n - k\} \in \mathcal{F}_n,$$
since $\mathcal{F}_k \subset \mathcal{F}_n$ for all $k \leq n$. □

Remark 3.3. In particular, if T is a stopping time, then for all $n \geq 0$, $T \wedge n$ is a bounded stopping time.

Proposition 3.2. *If $(T_k)_k$ is a sequence of stopping times, then $\inf_k T_k$, $\sup_k T_k$, $\liminf_k T_k$ and $\limsup_k T_k$ are also stopping times.*

Proof. Exercise. □

Proposition 3.3. *Let T be a stopping time. Then,*
$$\mathcal{F}_T = \{A \in \mathcal{F} \,|\, \forall n \geq 0, A \cap \{T = n\} \in \mathcal{F}_n\}$$
is a σ-algebra, called the σ-algebra of T-past.

Remark 3.4. Obviously, T is $\mathcal{F}_T$-measurable.

Proof. It is obvious that $\Omega \in \mathcal{F}_T$. If $A \in \mathcal{F}_T$, then for all n,
$$A^c \cap \{T = n\} = \{T = n\} \setminus A = \{T = n\} \setminus (A \cap \{T = n\}) \in \mathcal{F}_n,$$
hence $A^c \in \mathcal{F}_T$. If $(A_k)_k$ is countable collection of $\mathcal{F}_T$-mesurable set, then
$$\left(\bigcup_k A_k\right) \cap \{T = n\} = \bigcup_k (A_k \cap \{T = n\}) \in \mathcal{F}_n,$$
hence $\bigcup_k A_k \in \mathcal{F}_T$. □

Proposition 3.4. *Let S and T be two stopping times such that $S \leq T$. Then, $\mathcal{F}_S \subset \mathcal{F}_T$.*

PROOF. Let $A \in \mathcal{F}_S$. Then, for all $n \geq 0$,

$$A \cap \{T = n\} = A \cap \{S \leq n\} \cap \{T = n\} = \bigcup_{k=0}^{n} A \cap \{S = k\} \cap \{T = n\} \in \mathcal{F}_n. \qquad \square$$

DEFINITION 3.2. *Let $(X_n)_{n\geq 0}$ be an adapted stochastic process and T a stopping time. If $T < \infty$ a.s., we define the random variable X_T by*

$$X_T(\omega) = X_{T(\omega)}(\omega) = X_n(\omega) \quad \textit{if } T(\omega) = n.$$

Note that X_T is $\mathcal{F}_T$-measurable, since

$$\{X_T \in B\} \cap \{T = n\} = \{X_n \in B\} \cap \{T = n\} \in \mathcal{F}_n,$$

for any Borel set B.

3.2. Optional stopping theorem.

THEOREM 3.1 (Doob's optional stopping theorem). *Let $X = (X_n)_{n\geq 0}$ be a martingale (resp. a supermartingale, resp. a submartingale) and let T be a stopping time. Then $(X_{n\wedge T})_{n\geq 0}$ is also a martingale (resp. a supermartingale, resp. a submartingale). Moreover, if $S \leq T$ are two bounded stopping times, then*

$$\mathbb{E}(X_T \,|\, \mathcal{F}_S) = X_S \quad (\textit{resp. } \mathbb{E}(X_T \,|\, \mathcal{F}_S) \leq X_S, \textit{ resp. } \mathbb{E}(X_T \,|\, \mathcal{F}_S) \geq X_S.)$$

PROOF. We only deal with the case of martingales, the super and sub-martingales cases being similar. For all $n \geq 1$, define

$$H_n = \mathbb{1}_{\{T\geq n\}}.$$

Since $\{T \geq n\} = \{T \leq n-1\}^c \in \mathcal{F}_{n-1}$, the process $(H_n)_n$ is predictable. Now remark that

$$X_0 + (H\cdot X)_n = X_0 + \sum_{k=1}^{n} H_k(X_k - X_{k-1}) = X_0 + \sum_{k=1}^{n}(X_k - X_{k-1})\mathbb{1}_{\{T\geq k\}} = \begin{cases} X_n & \text{if } T \geq n \\ X_T & \text{if } T < n. \end{cases}$$

Hence,

$$X_{n\wedge T} = X_0 + (H\cdot X)_n,$$

so $(X_{n\wedge T})_{n\geq 0}$ is a martingale. The same argument applies if X is a supermartingale (resp. a submartingale), since the predictable process H is non-negative.

Now suppose that T is bounded, say $T \leq N$ a.s., and let $A \in \mathcal{F}_T$. Then,

$$\mathbb{E}(X_T\mathbb{1}_A) = \sum_{k\leq N} \mathbb{E}(X_k\mathbb{1}_{A\cap\{T=k\}}) = \sum_{k\leq N} \mathbb{E}\left(\mathbb{E}(X_N \,|\, \mathcal{F}_k)\mathbb{1}_{A\cap\{T=k\}}\right) = \sum_{k\leq N} \mathbb{E}\left(X_N\mathbb{1}_{A\cap\{T=k\}}\right),$$

since $A \cap \{T = k\} \in \mathcal{F}_k$, for all k. Hence,

$$\mathbb{E}(X_T\mathbb{1}_A) = \mathbb{E}(X_N\mathbb{1}_A),$$

for all $A \in \mathcal{F}_T$, thus $X_T = \mathbb{E}(X_N \,|\, \mathcal{F}_T)$ since X_T is $\mathcal{F}_T$-measurable.

Now if $S \leq T$ are two stopping times bounded by N, we apply the above to the martingale $(X_{n\wedge T})_{n\geq 0}$, hence

$$X_S = \mathbb{E}(X_{N\wedge T} \,|\, \mathcal{F}_S) = \mathbb{E}(X_T \,|\, \mathcal{F}_S).$$

$\square$

REMARK 3.5. As a consequence, if $S \leq T$ are two bounded stopping times, then

- if X is a martingale, then $\mathbb{E}(X_S) = \mathbb{E}(X_T)$;
- if X is a submartingale, then $\mathbb{E}(X_S) \leq \mathbb{E}(X_T)$.

DEFINITION 3.3. *The stochastic process $(X_{n\wedge T})_{n\geq 0}$ is called the stopped process, and is denoted by $X^T = (X^T_n)_{n\geq 0}$, i.e. $X^T_n = X_{n\wedge T}$, for all $n \geq 0$.*

Example 3.2. The assumption that the the stopping time is bounded in the optional stopping theorem cannot be weakened without further assumptions on the process. Indeed, consider the symmetric random walk $(X_n)_{n\geq 0}$ starting at 0. Let T_1 be the hitting time of 1. By the optional stopping theorem, for all $n \geq 0$,

$$\mathbb{E}(X_n^{T_1}) = \mathbb{E}(X_0) = 0.$$

Since the walk is a recurrent Markov chain, we known that $T_1 < \infty$ a.s. Hence, $n \wedge T_1$ converges to T_1, but we cannot taking the limit in the expectation since $\mathbb{E}(X_{T_1}) = 1$. There is no contradiction since the stopping time T_1 is not bounded (as the walk is null recurrent).

We have a partial converse to the previous theorem:

Proposition 3.5. *Let $X = (X_n)_{n\geq 0}$ be an integrable and adapted stochastic process. Then X is a martingale if and only if for all bounded stopping time T, we have*

$$\mathbb{E}(X_T) = \mathbb{E}(X_0).$$

Proof. The "only if" part follows from the optional stopping theorem. So, suppose that X is integrable and adapted and that $\mathbb{E}(X_T) = \mathbb{E}(X_0)$, for all bounded stopping time. Let $k < n$. It suffices to prove that for all $A \in \mathcal{F}_k$, $\mathbb{E}(X_k \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A)$. Let $A \in \mathcal{F}_k$. Define

$$T = k\mathbb{1}_A + n\mathbb{1}_{A^c}.$$

Then T is a obviously a bounded stopping time since $A \in \mathcal{F}_k \subset \mathcal{F}_n$. Hence,

$$\begin{aligned}\mathbb{E}(X_T) &= \mathbb{E}(X_k\mathbb{1}_A) + \mathbb{E}(X_n\mathbb{1}_{A^c}) = \mathbb{E}(X_k\mathbb{1}_A) + \mathbb{E}(X_n) - \mathbb{E}(X_n\mathbb{1}_A)\\ &= \mathbb{E}(X_0)\end{aligned}$$

by assumption, and since $\mathbb{E}(X_n) = \mathbb{E}(X_0)$ since n is also a stopping time, we get

$$\mathbb{E}(X_k\mathbb{1}_A) = \mathbb{E}(X_n\mathbb{1}_A).$$

□

3.3. The gambler ruin's problem. A gambler is playing heads or tails against the bank. Her initial fortune is $a \geq 1$ and the initial fortune of the bank is $b \geq 1$. We are interested to determine the probability of ruin of the player or of the bank. Let $(\varepsilon_n)_{n\geq 1}$ be a sequence of i.i.d. random variables with distribution $p\delta_1 + (1-p)\delta_{-1}$. Define $X_0 = a$, and $X_n = a + \varepsilon_1 + \cdots + \varepsilon_n$, for $n \geq 1$. Consider the stopping times

$$\tau_0 = \inf\{n \geq 0 \,|\, X_n = 0\}, \quad \tau_{a+b} = \inf\{n \geq 0 \,|\, X_n = a+b\}, \quad \tau = \tau_0 \wedge \tau_{a+b}.$$

First consider a fair game, i.e. $p = \frac{1}{2}$. Then $X = (X_n)_{n\geq 0}$ is a martingale, and by Doob's optional stopping theorem, the stopped process $X^\tau = (X_{n\wedge\tau})_{n\geq 0}$ is also a martingale, which is bounded by $a+b$.

Since $\tau < \infty$ a.s., (but not bounded, as $(X_n)_{n\geq 0}$ is null recurrent), we have that

$$X_{n\wedge\tau} \underset{n\to\infty}{\longrightarrow} X_\tau \quad \text{a.s.}$$

Thus by the dominated convergence theorem, we get that

$$\mathbb{E}\left(X_{n\wedge\tau}\right) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\left(X_\tau\right).$$

Since X^τ is a martingale, we have $\mathbb{E}\left(X_{n\wedge\tau}\right) = \mathbb{E}(X_0) = a$, hence $\mathbb{E}\left(X_\tau\right) = a$. We deduce that

$$\mathbb{P}(X_\tau = a+b) = \frac{a}{a+b}, \quad \text{and} \quad \mathbb{P}(X_\tau = 0) = \frac{b}{a+b}.$$

The ruin probability of the gambler is thus $\mathbb{P}(\tau = \tau_0) = \mathbb{P}(X_\tau = 0) = \frac{b}{a+b}$ (note that if $b \gg a$, then $\mathbb{P}(X_\tau = 0) \approx 1$, so don't play against the bank...).

Now consider an unfair game, i.e. $p \neq \frac{1}{2}$. Let $q = 1 - p$, and put $r = \frac{q}{p}$. Then, the process Z defined by

$$Z_n = r^{X_n}, \quad n \geq 0,$$

is a martingale. Indeed,

$$\mathbb{E}\left(r^{X_{n+1}} \,|\, \mathcal{F}_n\right) = r^{X_n}\,\mathbb{E}\left(r^{\varepsilon_{n+1}}\right) = r^{X_n}(rp + r^{-1}q) = r^{X_n}.$$

Note that $\tau < \infty$ a.s., as either τ_0 or τ_{a+b} (not both!) is a.s. finite by the law of large numbers. Hence, since $Z_{n\wedge\tau} \to Z_\tau$ a.s. and since Z^τ is bounded, one gets using again the dominated convergence theorem and the optional stopping theorem that

$$r^a = \mathbb{E}(Z_\tau) = \mathbb{P}(Z_\tau = 0) + r^{a+b}\,\mathbb{P}(Z_\tau = a + b),$$

and the ruin's probabilities are thus

$$\mathbb{P}(\tau = \tau_{a+b}) = \frac{r^a - 1}{r^{a+b} - 1} \quad \text{and} \quad \mathbb{P}(\tau = \tau_0) = \frac{r^{a+b} - r^a}{r^{a+b} - 1}.$$

4. Doob's decomposition

Definition 4.1. *Let $X = (X_n)_{n\geq 0}$ be a stochastic process. We denote by $(\Delta X_n)_{n\geq 0}$ the process given by the increments of X:*

$$\Delta X_0 = X_0 \quad \textit{and} \quad \Delta X_n = X_n - X_{n-1}, \quad \textit{for all } n \geq 1.$$

Remark 4.1. An adapted process $X = (X_n)_{n\geq 0}$ is thus a martingale if and only if

$$\mathbb{E}(\Delta X_{n+1} \,|\, \mathcal{F}_n) = 0, \quad \text{for all } n \geq 1.$$

Theorem 4.1 (Doob's decomposition). *Let $(X_n)_{n\geq 0}$ be a submartingale. Then, $(X_n)_{n\geq 0}$ can be written*

$$X_n = M_n + A_n,$$

where $(M_n)_{n\geq 0}$ is a martingale and $(A_n)_{n\geq 0}$ is a predictable non-decreasing process. This decomposition is unique almost surely. It is called the Doob's decomposition of $(X_n)_{n\geq 0}$.

Proof. We start with the existence. Define:

$$M_0 = X_0 \quad \text{and} \quad \Delta M_n = X_n - \mathbb{E}(X_n \,|\, \mathcal{F}_{n-1}), \text{ for all } n \geq 1.$$

Obviously,

$$\mathbb{E}(\Delta M_n \,|\, \mathcal{F}_{n-1}) = 0,$$

hence $(M_n)_{n\geq 0}$ is a martingale. Now define:

$$A_0 = 0 \quad \text{and} \quad \Delta A_n = \mathbb{E}(X_n \,|\, \mathcal{F}_{n-1}) - X_{n-1}, \text{ for all } n \geq 1.$$

Clearly $(A_n)_{n\geq 0}$ is predictable. Since $(X_n)_{n\geq 0}$ is a submartingale,

$$\mathbb{E}(X_n \,|\, \mathcal{F}_{n-1}) \geq X_{n-1},$$

hence one has $\Delta A_n \geq 0$, and A_n is non-decreasing. By construction, it is clear that

$$X_n = M_n + A_n.$$

Now we prove the unicity a.s. Consider a second decomposition

$$X_n = M'_n + A'_n,$$

where $(M'_n)_{n\geq 0}$ is a martingale and $(A'_n)_{n\geq 0}$ is predictable and non-decreasing. Then,

$$\Delta A'_n = \Delta X_n - \Delta M'_n.$$

By taking the conditional expectation given $\mathcal{F}_{n-1}$, and using the fact that $(M'_n)_{n\geq 0}$ is a martingale and $(A'_n)_{n\geq 0}$ is predictable, we get that

$$\Delta A'_n = \mathbb{E}(\Delta X_n \,|\, \mathcal{F}_{n-1}) = \Delta A_n.$$

Hence, $A_n = A'_n$ a.s. and thus $M'_n = M_n$ a.s. $\square$

Recall that if $(X_n)_n$ is a martingale, then $(X_n^2)_n$ is a submartingale by Jensen's inequality. We make the following definition.

Definition 4.2. *Let $(X_n)_{n\geq 0}$ be a L^2-martingale. The unique predictable process $(A_n)_n$ such that $(X_n^2 - A_n)_n$ is a martingale is called the square variation process of $(X_n)_n$ and is denoted by $\langle X\rangle = (\langle X\rangle_n)_n$.*

Proposition 4.1. *Let $(X_n)_{n\geq 0}$ be a L^2-martingale. Then, for all $n \geq 0$,*

$$\langle X\rangle_n = \sum_{k=1}^n \mathbb{E}\left((X_k - X_{k-1})^2 \,|\, \mathcal{F}_{k-1}\right)$$
$$\mathbb{E}(\langle X\rangle_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_0^2).$$

Proof. From the proof of Doob's decomposition, one has that the square variation process of $(X_n)_n$ is given by

$$\langle X\rangle_n = \sum_{k=1}^n \left(\mathbb{E}\left(X_k^2 \,|\, \mathcal{F}_{k-1}\right) - X_{k-1}^2\right).$$

On the other hand, we have:

$$\begin{aligned}\sum_{k=1}^n \mathbb{E}\left((X_k - X_{k-1})^2 \,|\, \mathcal{F}_{k-1}\right) &= \sum_{k=1}^n \mathbb{E}\left(X_k^2 + X_{k-1}^2 - 2X_kX_{k-1} \,|\, \mathcal{F}_{k-1}\right) \\ &= \sum_{k=1}^n \left(\mathbb{E}\left(X_k^2 \,|\, \mathcal{F}_{k-1}\right) + X_{k-1}^2 - 2X_{k-1}\,\mathbb{E}(X_k \,|\, \mathcal{F}_{k-1})\right) \\ &= \sum_{k=1}^n \left(\mathbb{E}\left(X_k^2 \,|\, \mathcal{F}_{k-1}\right) - X_{k-1}^2\right)\end{aligned}$$

since $(X_n)_n$ is a martingale. The first formula follows. Now taking the expectation in the definition of $\langle X\rangle_n$ gives the second one. $\square$

Example 4.1. We continue the example of the symmetric random walk:

$$X_n = \varepsilon_1 + \cdots + \varepsilon_n,$$

where $(\varepsilon_k)_k$ are i.i.d. random variables with uniform distribution on $\{-1, 1\}$. The process $(X_n)_{n\geq 0}$ is a martingale and since $X_k - X_{k-1} = \varepsilon_k$ is independent of $\mathcal{F}_{k-1}$, its square variation process is

$$\langle X\rangle_n = \sum_{k=1}^n \mathbb{E}(\varepsilon_k^2) = n.$$

Hence, $(X_n^2 - n)_{n\geq 0}$ is a martingale.

Now, $(|X_n|)_{n\geq 0}$ is also a submartingale and we can apply Doob's decomposition. Its associated increasing predictable process is then

$$A_n = \sum_{k=1}^n \left(\mathbb{E}\left(|X_k| \,|\, \mathcal{F}_{k-1}\right) - |X_{k-1}|\right).$$

Now remark that

$$|X_k| = \begin{cases} |X_{k-1}| + \varepsilon_k & \text{if } X_{k-1} > 0, \\ |X_{k-1}| - \varepsilon_k & \text{if } X_{k-1} < 0, \\ 1 & \text{if } X_{k-1} = 0. \end{cases}$$

Hence, since ε_k is independent of $\mathcal{F}_{k-1}$, and $\mathbb{E}(\varepsilon_k) = 0$, we have

$$\mathbb{E}\left(|X_k| \,|\, \mathcal{F}_{k-1}\right) = \begin{cases} |X_{k-1}| & \text{if } X_{k-1} \neq 0, \\ 1 & \text{if } X_{k-1} = 0, \end{cases}$$

which gives

$$A_n = \sum_{k=1}^{n} \mathbb{1}_{\{X_{k-1}=0\}} = \#\{0 \leq k \leq n-1 \,|\, X_k = 0\}.$$

This is the local time of $(X_n)_n$ at 0.

5. Martingale convergence theorems

5.1. L^2 convergence. We start with an easy result concerning the convergence in L^2 of martingales.

Let X be a square-integrable martingale, that is $X_n \in L^2$, for all $n \geq 0$. The following proposition states that the increments of a square-integrable martingale are orthogonal.

Proposition 5.1. *Let X be a square-integrable martingale, then for all $n \neq k$,*

$$\mathbb{E}(\Delta X_n \Delta X_k) = 0.$$

Proof. Suppose that $n > k$. Since $\Delta X_k = X_k - X_{k-1}$ is $\mathcal{F}_k$-measurable,

$$\mathbb{E}(\Delta X_n \Delta X_k \,|\, \mathcal{F}_k) = \Delta X_k \, \mathbb{E}(\Delta X_n \,|\, \mathcal{F}_k) = 0,$$

by the martingale property. Then we take the expectation. □

Furthermore, if the martingale is bounded in L^2, we have the following convergence theorem.

Theorem 5.1. *Let $X = (X_n)_{n\geq 0}$ be a martingale, which is bounded in L^2, that is*

$$\sup_{n\geq 0} \mathbb{E}\left(X_n^2\right) < +\infty.$$

Then X converges in L^2.

Remark 5.1. The L^2-boundedness condition is equivalent to $\sup_n \mathbb{E}(\langle X\rangle_n) < \infty$ since $\mathbb{E}(\langle X\rangle_n) = \mathbb{E}(X_n^2) - \mathbb{E}(X_0^2)$ (see Proposition 4.1).

Proof. We have,

$$\begin{aligned} \mathbb{E}(X_n^2) &= \mathbb{E}\left(\left(X_0 + \sum_{k=1}^{n} \Delta X_n\right)^2\right) \\ &= \mathbb{E}(X_0^2) + \sum_{k=1}^{n} \mathbb{E}\left((\Delta X_n)^2\right), \end{aligned}$$

by the orthogonal property of the increments. Hence, since $\sup_{n\geq 0} \mathbb{E}(X_n^2) < \infty$ by assumption, one has that the series

$$\sum_{n\geq 1} \mathbb{E}\left((\Delta X_n)^2\right) < \infty.$$

Moreover, for all $n, p \geq 0$,

$$\mathbb{E}\left((X_{n+p} - X_n)^2\right) = \sum_{k=n+1}^{n+p} \mathbb{E}\left((\Delta X_k)^2\right),$$

since the increments are orthogonal. Thus, as $p \to \infty$ and $n \to \infty$, we get that

$$\lim_{n\to\infty} \lim_{p\to\infty} \mathbb{E}\left((X_{n+p} - X_n)^2\right) = 0,$$

hence $(X_n)_{n\geq 0}$ is a Cauchy sequence in the Hilbert space L^2, hence it converges in L^2. $\square$

We will now study the almost sure convergence and the convergence in L^p for $p > 1$.

5.2. Almost sure convergence.

5.2.1. *Upcrossing lemma.* Let $a < b$. Introduce the family of stopping times:

$$\tau_1 = \inf\{n \geq 0 \,|\, X_n \leq a\}$$
$$\sigma_1 = \inf\{n > \tau_1 \,|\, X_n \geq b\},$$

and for $k \geq 2$,

$$\tau_k = \inf\{n > \sigma_{k-1} \,|\, X_n \leq a\}$$
$$\sigma_k = \inf\{n > \tau_k \,|\, X_n \geq b\}.$$

We thus have

$$0 \leq \tau_1 < \sigma_1 < \tau_2 < \sigma_2 < \cdots < \sigma_{k-1} < \tau_k < \sigma_k < \cdots$$

For $n \geq 1$, define the number of upcrossings of X accross $[a, b]$ until time n by

$$U_n^{a,b} = \sup\{k \geq 1 \,|\, \sigma_k \leq n\} = \sum_{k\geq 1} \mathbb{1}_{\{\sigma_k \leq n\}}.$$

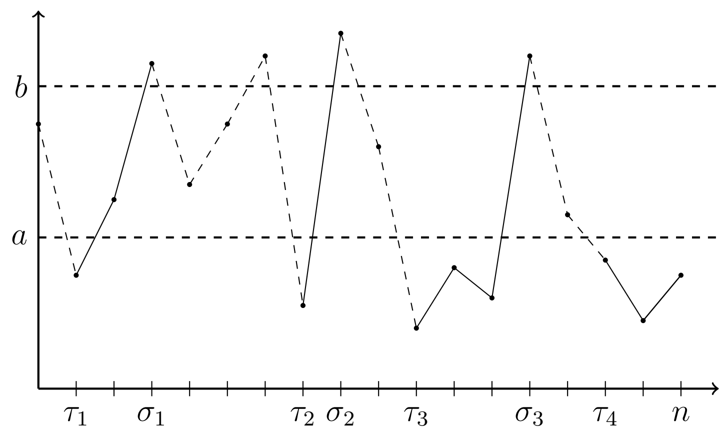

FIGURE 1. A stochastic process with 3 upcrossings over $[a, b]$.

Note that by construction, for all $k \geq 1$, $X_{\tau_k} \leq a < b \leq X_{\sigma_k}$ on $\{\sigma_k < \infty\}$, and that $\{U_n^{a,b} \geq k\} = \{\sigma_k \leq n\}$, see Figure 1.

If we think of X as a stock price, a trading strategy would be to buy the stock when its price has fallen below a and to sell it when it exceeds b. Each time the price makes an upcrossing from a to b, we make a profit of at least $b - a$. The maximal profit up to time n will thus be given by the number of such upcrossing times $b - a$ minus the possible loss at n. This idea is formalized in the lemma below.

LEMMA 5.1 (Doob's upcrossing inequality). *Let $X = (X_n)_{n\geq 0}$ be a supermartingale. Then*

$$(b - a)\,\mathbb{E}\left(U_n^{a,b}\right) \leq \mathbb{E}((X_n - a)^-).$$

PROOF. Define the process H by

$$H_n = \sum_{k\geq 1} \mathbb{1}_{\{\tau_k < n \leq \sigma_k\}} = \begin{cases} 1 & \text{if } n \in \{\tau_k + 1, \ldots, \sigma_k\} \text{ for some } k \geq 1, \\ 0 & \text{else.} \end{cases}$$

Then H is non-negative, bounded by 1, and predictable since:

$$\{H_n = 1\} = \bigcup_{k\geq 1} \{\tau_k \leq n-1\} \cap \{\sigma_k \leq n-1\}^c \in \mathcal{F}_{n-1}.$$

Now we claim that the martingale transform $H \cdot X$ satisfies

$$(H \cdot X)_n \geq U_n^{a,b}(b-a) - (X_n - a)^-.$$

This can be understood from the above picture. Each upcrossing of $[a,b]$ results in a profit of at least $b-a$ and the term $(X_n-a)^-$ overestimates the eventual loss due to the possible incomplete upcrossing at the end. Formally,

$$\begin{aligned}
(H \cdot X)_n &= \sum_{k=1}^n H_k\,(X_k - X_{k-1}) \\
&= \sum_{k=1}^n \sum_{j\geq 1} \mathbb{1}_{\{\tau_j+1\leq k\leq \sigma_j\}}\,(X_k - X_{k-1}) \\
&= \sum_{j\geq 1} \sum_{k=1}^n \mathbb{1}_{\{\tau_j+1\leq k\leq \sigma_j\}}\,(X_k - X_{k-1}) \\
&= \sum_{j\geq 1} \mathbb{1}_{\{\tau_j<n\}} \sum_{k=\tau_j+1}^{\sigma_j\wedge n} (X_k - X_{k-1}) \\
&= \sum_{j\geq 1} \left(X_{\sigma_j} - X_{\tau_j}\right) \mathbb{1}_{\{\sigma_j\leq n\}} + \sum_{j\geq 1} \left(X_n - X_{\tau_j}\right) \mathbb{1}_{\{\sigma_{j-1}\leq n<\sigma_j\}} \mathbb{1}_{\{\tau_j<n\}} \\
&= \sum_{j=1}^{U_n^{a,b}} \left(X_{\sigma_j} - X_{\tau_j}\right) + \mathbb{1}_{\{\tau_{U_n^{a,b}+1}<n\}}(X_n - X_{\tau_{U_n^{a,b}+1}})
\end{aligned}$$

since $\{\sigma_j \leq n\} = \{U_n^{a,b} \geq j\}$ and $\{\sigma_{j-1} \leq n < \sigma_j\} = \{U_n^{a,b} = j-1\}$. Thus,

$$(H \cdot X)_n \geq U_n^{a,b}(b-a) - (X_n - a)^-,$$

since $X_{\sigma_j} - X_{\tau_j} \geq b-a$, and on $\{\tau_{U_n^{a,b}+1} < n\}$, $X_n - X_{\tau_{U_n^{a,b}+1}} \geq X_n - a \geq -(X_n-a)^-$.

Since $H \cdot X$ is a supermartingale, we have $\mathbb{E}\left((H \cdot X)_n\right) \leq \mathbb{E}\left((H \cdot X)_0\right) = 0$, so we get the inequality. □

5.2.2. *Almost sure convergence.*

THEOREM 5.2. *Let $X = (X_n)_{n\geq 0}$ be a supermartingale with $\sup_{n\geq 0} \mathbb{E}(X_n^-) < \infty$. There exists a random variable X_∞ with $\mathbb{E}(|X_\infty|) < \infty$ such that*

$$X_n \underset{n\to\infty}{\longrightarrow} X_\infty \quad a.s.$$

PROOF. Let $a < b$ and $U_n^{a,b}$ the number of upcrossings over $[a,b]$ of X up to time n. By Doob's upcrossing inequality, and using that $(x-a)^- \leq |a| + x^-$, we have

$$(b-a)\,\mathbb{E}(U_n^{a,b}) \leq \mathbb{E}((X_n-a)^-) \leq |a| + \mathbb{E}(X_n^-) \leq |a| + \sup_n \mathbb{E}(X_n^-) < \infty,$$

by assumption. Define

$$U^{a,b} = \sup_n U_n^{a,b}.$$

By monotone convergence theorem, one gets that $\mathbb{E}(U^{a,b}) < \infty$, hence $U^{a,b} < \infty$ a.s. Now define,

$$C^{a,b} = \left\{\liminf_n X_n < a < b < \limsup_n X_n\right\}.$$

Then,

$$C^{a,b} \subset \left\{U^{a,b} = +\infty\right\},$$

hence $C^{a,b}$ is a null set. But,

$$C = \left\{\liminf_n X_n < \limsup_n X_n\right\} = \bigcup_{\substack{a,b\in\mathbb{Q}\\ a<b}} C^{a,b},$$

hence $\mathbb{P}(C) = 0$, and $(X_n)_n$ converges in $\overline{\mathbb{R}}$ a.s. It remains to prove that the limit X_∞ is in L^1 which will imply that X_∞ is finite a.s. By Fatou's lemma, one has

$$\mathbb{E}(|X_\infty|) \le \liminf_n \mathbb{E}(|X_n|) \le \sup_n \mathbb{E}(|X_n|).$$

Writing

$$|X_n| = X_n^+ + X_n^- = X_n + 2X_n^-$$

and since $(X_n)_n$ is a supermartingale $\mathbb{E}(X_n) \le \mathbb{E}(X_0)$, we get that

$$\sup_n \mathbb{E}(|X_n|) \le \mathbb{E}(X_0) + 2\sup_n \mathbb{E}(X_n^-) < \infty$$

by assumption. Hence $X_\infty \in L^1$, and in particular $X_\infty < \infty$ a.s. □

REMARK 5.2. If $(X_n)_n$ is a non-negative supermartingale, then the assumption of the theorem is automatically verified, hence a non-negative supermartingale converges almost surely!

If $(X_n)_n$ is a submartingale, $(-X_n)_n$ is a supermartingale, so one deduces:

COROLLARY 5.1. *Let $X = (X_n)_{n\ge0}$ be a submartingale with $\sup_{n\ge0}\mathbb{E}(X_n^+) < \infty$. There exists a random variable X_∞ with $\mathbb{E}(|X_\infty|) < \infty$ such that*

$$X_n \underset{n\to\infty}{\longrightarrow} X_\infty \quad a.s.$$

The analogous statement for martingales is:

COROLLARY 5.2. *Let $X = (X_n)_{n\ge0}$ be a martingale which is non-negative or bounded in L^1, i.e. $\sup_{n\ge0}\mathbb{E}(|X_n|) < \infty$. There exists a random variable X_∞ with $\mathbb{E}(|X_\infty|) < \infty$ such that*

$$X_n \underset{n\to\infty}{\longrightarrow} X_\infty \quad a.s.$$

5.3. L^p convergence, $p > 1$.

5.3.1. *Maximal Inequalities.*

LEMMA 5.2 (Doob's maximal inequality). *Let $X = (X_n)_{n\ge0}$ be a non-negative submartingale and denote*

$$X_n^* = \sup_{0\le k\le n} X_k, \quad n \ge 0.$$

For all $n \ge 0$, and all $\lambda > 0$, one has

$$\lambda\,\mathbb{P}\left(X_n^* \ge \lambda\right) \le \mathbb{E}\left(X_n \mathbb{1}_{\{X_n^*\ge\lambda\}}\right) \le \mathbb{E}(X_n).$$

PROOF. The second inequality is obvious (since X is non-negative). We introduce the stopping time $\tau_\lambda = \inf\{n \ge 0 \,|\, X_n \ge \lambda\}$. Note that

$$\{\tau_\lambda \le n\} = \{X_n^* \ge \lambda\}.$$

Introduce also the bounded stopping time $\tau = \tau_\lambda \wedge n$. Using the optional stopping theorem applied to $\tau \le n$, one has

$$\begin{aligned}\mathbb{E}(X_n) &\ge \mathbb{E}(X_\tau)\\ &= \mathbb{E}\left(X_\tau \mathbb{1}_{\{X_n^*\ge\lambda\}}\right) + \mathbb{E}\left(X_\tau \mathbb{1}_{\{X_n^*<\lambda\}}\right)\\ &\ge \lambda\,\mathbb{P}\left(X_n^* \ge \lambda\right) + \mathbb{E}\left(X_n \mathbb{1}_{\{X_n^*<\lambda\}}\right),\end{aligned}$$

since on $\{X_n^* \geq \lambda\}$, $\tau = \tau_\lambda$ and thus $X_\tau \geq \lambda$, and on $\{X_n^* < \lambda\}$, $\tau = n$. Therefore,

$$\lambda\,\mathbb{P}\,(X_n^* \geq \lambda) \leq \mathbb{E}(X_n) - \mathbb{E}\left(X_n \mathbb{1}_{\{X_n^* < \lambda\}}\right) = \mathbb{E}\left(X_n \mathbb{1}_{\{X_n^* \geq \lambda\}}\right). \qquad \square$$

Corollary 5.3. *Let $X = (X_n)_{n\geq 0}$ be a martingale. Then, for all $n \geq 0$ and all $\lambda > 0$,*

$$\lambda\,\mathbb{P}\left(\sup_{0\leq k\leq n} |X_n| \geq \lambda\right) \leq \mathbb{E}\left(|X_n| \mathbb{1}_{\{\sup_{0\leq k\leq n} |X_k| \geq \lambda\}}\right) \leq \mathbb{E}\,|X_n|.$$

Proof. This follows immediately using Doob's maximal inequality since if $(X_n)_{n\geq 0}$ is a martingale, then $(|X_n|)_{n\geq 0}$ is a non-negative submartingale. $\square$

Example 5.1 (Kolmogorov's inequality). The following inequality, due to Kolmogorov, can be proved directly, but the proof becomes immediate using martingale theory. Let $(X_n)_{n\geq 0}$ be a sequence of square-integrable independent random variables such that $\mathbb{E}(X_1) = 0$. Put $S_n = \sum_{k=1}^n X_k$. Then, for all $n \geq 0$, and all $\lambda > 0$,

$$\mathbb{P}\left(\sup_{0\leq k\leq n} |S_k| \geq \lambda\right) \leq \frac{1}{\lambda^2}\,\mathbb{V}\mathrm{ar}(S_n).$$

Indeed, it suffices to apply Doob's maximal inequality to the submartingale $(S_n^2)_{n\geq 0}$.

Proposition 5.2 (Doob L^p maximal inequality). *Let $p > 1$. Let $X = (X_n)_{n\geq 0}$ be a non-negative submartingale such that $X_n \in L^p$ for all $n \geq 0$. Denote*

$$X_n^* = \sup_{0\leq k\leq n} X_k, \quad n \geq 0.$$

Then for all $n \geq 0$, $X_n^* \in L^p$, and

$$||X_n^*||_p \leq \frac{p}{p-1}||X_n||_p.$$

Proof. Since X is non-negative,

$$(X_n^*)^p \leq \left(\sum_{k=0}^n X_k\right)^p \leq (n+1)^{p-1} \sum_{k=0}^n X_k^p,$$

hence X_n^* is in L^p since $X_n \in L^p$ for all $n \geq 0$.

Writing $x^p = \int_0^x p\lambda^{p-1} d\lambda$ and using Fubini's theorem, one has

$$\begin{aligned}\mathbb{E}\left((X_n^*)^p\right) &= \int_0^{+\infty} p\lambda^{p-1}\,\mathbb{P}(X_n^* \geq \lambda) d\lambda \\ &\leq p \int_0^{+\infty} \lambda^{p-2}\,\mathbb{E}\left(X_n \mathbb{1}_{\{X_n^* \leq \lambda\}}\right) d\lambda\end{aligned}$$

using Doob's maximal inequality. Hence,

$$\begin{aligned}\mathbb{E}\left((X_n^*)^p\right) &\leq p\,\mathbb{E}\left(X_n \int_0^{X_n^*} \lambda^{p-2} d\lambda)\right) = \frac{p}{p-1}\,\mathbb{E}\left(X_n (X_n^*)^{p-1}\right) \\ &\leq \frac{p}{p-1}\left(\mathbb{E}(X_n^p)\right)^{1/p}\left(\mathbb{E}\left((X_n^*)^p\right)\right)^{(p-1)/p}\end{aligned}$$

by Hölder's inequality (with conjugates p and $q = \frac{p}{p-1}$). Dividing by $\left(\mathbb{E}\left((X_n^*)^p\right)\right)^{(p-1)/p}$ gives the result. $\square$

5.3.2. *L^p convergence, $p > 1$.*

THEOREM 5.3. *Let $p > 1$. Let $X = (X_n)_{n\geq 0}$ be a martingale which is bounded in L^p, i.e. $\sup_{n\geq 0} \mathbb{E}(|X_n|^p) < \infty$. Then, X converges a.s. and in L^p towards a random variable X_∞ such that*

$$\mathbb{E}(|X_\infty|^p) = \sup_{n\geq 0} \mathbb{E}(|X_n|^p).$$

PROOF. Since the martingale X is bounded in L^1, we already know that X converges a.s. to some X_∞. By Doob's L^p maximal inequality, we have

$$\mathbb{E}\left((X_n^*)^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{n\geq 0} \mathbb{E}(|X_n|^p),$$

where $X_n^* = \sup_{0\leq k\leq n} |X_k|$. By monotone convergence, we get

$$\mathbb{E}\left((X_\infty^*)^p\right) \leq \left(\frac{p}{p-1}\right)^p \sup_{n\geq 0} \mathbb{E}(|X_n|^p) < \infty.$$

Since $|X_n| \leq X_\infty^*$ for all n, by the dominated convergence theorem, we get that $(X_n)_n$ converges to X_∞ in L^p. Since $x \mapsto |x|^p$ is a convex function, by Jensen's inequality for conditional expectations, we have that the sequence $(E(|X_n|^p))_n$ is non-decreasing, hence by monotone convergence,

$$\mathbb{E}(|X_\infty|^p) = \sup_{n\geq 0} \mathbb{E}(|X_n|^p). \qquad \square$$

REMARK 5.3. Note that the theorem does not hold for $p = 1$, and we will need the notion of uniform integrability to obtain L^1 convergence. For instance, if $(X_n)_{n\geq 1}$ is a sequence of i.i.d. random variables with Gaussian distribution $\mathcal{N}(0, 1)$, define $S_0 = 0$, and $S_n = X_1 + \cdots + X_n$, for $n \geq 1$. Hence,

$$M_n = e^{S_n - n/2}, \quad n \geq 0,$$

is non-negative (and in particular bounded in L^1) martingale which converges to 0 a.s. but the convergence does not hold in L^1 (exercise).

6. Uniformly integrable martingales

6.1. Definition, examples. We start with the definition of uniform integrability of a family of random variables.

DEFINITION 6.1. *A family of random variables $X = (X_i)_{i\in I}$ is said to be uniformly integrable if*

$$\lim_{a\to\infty} \sup_{i\in I} \mathbb{E}\left(|X_i| \mathbb{1}_{\{|X_i|\geq a\}}\right) = 0.$$

EXAMPLE 6.1. (i) A random variable $X \in L^1$ is uniformly integrable by the dominated convergence theorem. Likewise, a finite family $\{X_0, X_1, \ldots, X_N\}$ is uniformly integrable.

(ii) If $(X_n)_{n\geq 0}$ is uniformly integrable, then it is bounded in L^1. Indeed, choose $a > 0$ large enough such that

$$\sup_n \mathbb{E}\left(|X_n| \mathbb{1}_{\{|X_n|\geq a\}}\right) \leq 1,$$

and write

$$\mathbb{E}(|X_n|) = \mathbb{E}\left(|X_n| \mathbb{1}_{\{|X_n|\geq a\}}\right) + \mathbb{E}\left(|X_n| \mathbb{1}_{\{|X_n|< a\}}\right) \leq 1 + a,$$

Then $\sup_n \mathbb{E}(|X_n|) < \infty$. The converse is wrong as shown by the following example.

(iii) A non uniformly integrable example: let X_n defined on the probability space $([0,1],\mathcal{B}([0,1]),\lambda)$, where λ denotes Lebesgue measure, by

$$X_n(\omega) = \begin{cases} n & \text{if } \omega \in (0, \frac{1}{n}) \\ 0 & \text{else,} \end{cases}$$

that is, $X_n = n\mathbb{1}_{(0,\frac{1}{n})}$. Then, for all n, $\mathbb{E}(|X_n|) = 1$, and $\mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}) = 1$ for $n \geq a$, so (X_n) is not uniformly integrable. Moreover, it is easy to see that (X_n) converges in probability to 0, but does not converges in L^1.

(iv) A dominated family of random variables is easily seen to be uniformly integrable: let $(X_n)_n$ such that there exists an integrable random variable $Z > 0$ such that $|X_n| \leq Z$ for all n. Then,

$$\sup_n \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) \leq \mathbb{E}\left(Z\mathbb{1}_{\{Z\geq a\}}\right) \underset{a\to\infty}{\longrightarrow} 0,$$

by the dominated convergence theorem.

(v) If $(X_n)_{n\geq 0}$ is bounded in L^p for some $p > 1$, that is $\sup_n \mathbb{E}(|X_n|^p) < \infty$, then it is uniformly integrable. Indeed, using Hölder's inequality, and Markov inequality, one has

$$\begin{aligned} \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) &\leq (\mathbb{E}(|X_n|^p))^{1/p}\,\mathbb{P}\left(|X_n| \geq a\right)^{1-1/p} \\ &\leq \frac{1}{a^{1-1/p}}\left(\mathbb{E}(|X_n|^p)\right)^{1/p}\left(\mathbb{E}(|X_n|)\right)^{1-1/p}. \end{aligned}$$

Hence, since $(X_n)_{n\geq 0}$ is bounded in L^p, there exists a constant $M > 0$, such that

$$\sup_n \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) \leq \frac{M}{a^{1-1/p}} \underset{a\to\infty}{\longrightarrow} 0.$$

Proposition 6.1. *A family of random variables $(X_n)_{n\geq 0}$ is uniformly integrable if and only if it is bounded in L^1 and equicontinuous: for all $\varepsilon > 0$, there exists $\delta > 0$ such that*

$$\mathbb{P}(A) < \delta \;\Rightarrow\; \sup_n \mathbb{E}(|X_n|\mathbb{1}_A) < \varepsilon.$$

Proof. The "only if" part. Writing

$$\begin{aligned} \mathbb{E}(|X_n|\mathbb{1}_A) &= \mathbb{E}\left(|X_n|\mathbb{1}_A\mathbb{1}_{\{|X_n|<a\}}\right) + \mathbb{E}\left(|X_n|\mathbb{1}_A\mathbb{1}_{\{|X_n|\geq a\}}\right) \\ &\leq a\,\mathbb{P}(A) + \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right), \end{aligned}$$

for $A = \Omega$, we get

$$\sup_n \mathbb{E}(|X_n|) \leq a + \sup_n \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right)$$

hence, $(X_n)_{n\geq 0}$ is bounded in L^1 since it is uniformly integrable. Now, let $\varepsilon > 0$. By uniform integrability, for a large enough, one has

$$\sup_n \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) < \varepsilon/2.$$

Choose $\delta = \varepsilon/2a$. Then if $\mathbb{P}(A) < \delta$, by the above inequality, we get

$$\sup_n \mathbb{E}(|X_n|\mathbb{1}_A) \leq a\frac{\varepsilon}{2a} + \frac{\varepsilon}{2} = \varepsilon,$$

hence $(X_n)_{n\geq 0}$ is equicontinuous.

The "if" part. Put $M = \sup_n \mathbb{E}(|X_n|) < \infty$. By Markov inequality, for all $n \geq 0$,

$$\mathbb{P}(|X_n| \geq a) \leq \frac{M}{a}.$$

For $\varepsilon > 0$, let $\delta > 0$ such that $\mathbb{P}(A) < \delta$ implies $\sup_n \mathbb{E}(|X_n|\mathbb{1}_A) < \varepsilon$. Then for a large enough such that $a > M/\delta$, one has for all $n \geq 0$,

$$\mathbb{P}(|X_n| \geq a) < \delta,$$

hence, for all $n \geq 0$,

$$\mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) < \varepsilon,$$

hence $(X_n)_n$ is uniformly integrable. □

We have the following refinement of the dominated convergence theorem.

THEOREM 6.1. *A family of random variables $X = (X_n)_{n\geq 0}$ converges in L^1 if and only if X is uniformly integrable and converges in probability.*

PROOF. If $(X_n)_{n\geq 0}$ converges in L^1 then it converges in probability (by Markov inequality). Let $\varepsilon > 0$. Choose N large enough such that for all $n \geq N$,

$$\mathbb{E}\left(|X_n - X_N|\right) < \frac{\varepsilon}{2}.$$

Since $X_0, X_1, \ldots, X_N$ is a finite family of random variables, it is uniformly integrable. Hence, by equicontinuity, we can choose $\eta > 0$ such that for all measurable set A such that $\mathbb{P}(A) < \eta$,

$$\sup_{0\leq k\leq N} \mathbb{E}\left(|X_k|\mathbb{1}_A\right) < \frac{\varepsilon}{2}.$$

Moreover, for all $n > N$,

$$\mathbb{E}\left(|X_n|\mathbb{1}_A\right) \leq \mathbb{E}\left(|X_N|\mathbb{1}_A\right) + \mathbb{E}\left(|X_n - X_N|\right) < \varepsilon.$$

This proves the equicontinuity of the family $(X_n)_{n\geq 0}$, hence the uniform integrability.

Conversely, suppose that $(X_n)_{n\geq 0}$ converges in probability to say X and that it is uniformly integrable. The family $(X_n - X_m)_{n,m\geq 0}$ is also uniformly integrable (this follows easily from the equicontinuity of $(X_n)_n$). Hence, for $\varepsilon > 0$, we can choose a large enough such that

$$\sup_{n,m} \mathbb{E}\left(|X_n - X_m|\mathbb{1}_{\{|X_n - X_m|\geq a\}}\right) \leq \varepsilon.$$

Hence, for all $m, n \geq 0$,

$$\begin{aligned}\mathbb{E}\left(|X_n - X_m|\right) &\leq \mathbb{E}\left(|X_n - X_m|\mathbb{1}_{\{|X_n - X_m|\leq\varepsilon\}}\right) + \mathbb{E}\left(|X_n - X_m|\mathbb{1}_{\{\varepsilon<|X_n - X_m|\leq a\}}\right) \\ &\qquad + \mathbb{E}\left(|X_n - X_m|\mathbb{1}_{\{|X_n - X_m|\geq a\}}\right) \\ &\leq \varepsilon + a\,\mathbb{P}\left(|X_n - X_m| > \varepsilon\right) + \varepsilon.\end{aligned}$$

The convergence in probability of $(X_n)_n$ implies that it is a Cauchy sequence for the convergence in probability:

$$\mathbb{P}\left(|X_n - X_m| > \varepsilon\right) \leq \mathbb{P}\left(|X_n - X| > \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|X - X_m| > \frac{\varepsilon}{2}\right) \underset{n,m\to\infty}{\longrightarrow} 0.$$

Hence, we have that

$$\lim_{n,m\to\infty} \mathbb{E}\left(|X_n - X_m|\right) = 0,$$

so $(X_n)_n$ is a Cauchy sequence in L^1 which is complete, hence converges in L^1. □

6.2. L^1 convergence. We recall the definition of a closed martingale.

Definition 6.2. *A martingale $X = (X_n)_{n\geq 0}$ is said to be a closed martingale if there exists an integrable random variable Y such that*

$$X_n = \mathbb{E}(Y \,|\, \mathcal{F}_n), \quad \textit{for all } n \geq 0.$$

We now have all the ingredients to prove the following theorem.

Theorem 6.2. *Let $X = (X_n)_{n\geq 0}$ be a martingale. The following assertions are equivalent:*

(i) X is a closed martingale;
(ii) X is uniformly integrable;
(iii) X converges almost surely and in L^1.

Proof. $(i) \Rightarrow (ii)$. Let $(X_n)_n$ be a closed martingale: there exists $Y \in L^1$, such that $X_n = \mathbb{E}(Y \,|\, \mathcal{F}_n)$, for all $n \geq 0$. Let $\varepsilon > 0$. Since Y is uniformly integrable, there exists $\delta > 0$ such that

$$\mathbb{P}(A) < \delta \;\Rightarrow\; \mathbb{E}(|Y|\mathbb{1}_A) < \varepsilon.$$

Now by Markov inequality,

$$\mathbb{P}(|X_n| \geq a) \leq \frac{1}{a}\,\mathbb{E}(|X_n|) \leq \frac{1}{a}\;\mathbb{E}(|Y|) < \delta,$$

for a large enough. Hence, for all $n \geq 0$,

$$\begin{aligned} \mathbb{E}\left(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\right) &\leq \mathbb{E}\left(\mathbb{E}(|Y| \,|\, \mathcal{F}_n)\mathbb{1}_{\{|X_n|\geq a\}}\right) \\ &= \mathbb{E}\left(\mathbb{E}(|Y|\mathbb{1}_{\{|X_n|\geq a\}} \,|\, \mathcal{F}_n)\right) \\ &= \mathbb{E}\left(|Y|\mathbb{1}_{\{|X_n|\geq a\}}\right), \end{aligned}$$

where we use the fact that $\{|X_n| \geq a\}$ is $\mathcal{F}_n$-measurable. Since $\mathbb{P}(|X_n| \geq a) < \delta$, we conclude using the uniform integrability of Y.

$(ii) \Rightarrow (iii)$. Since $(X_n)_n$ is uniformly integrable, it is bounded in L^1. Hence, we have already seen that (X_n) converges a.s. The convergence in L^1 follows by Theorem 6.1.

$(iii) \Rightarrow (i)$. Suppose that $X_n \to X_\infty$ a.s. and in L^1. Let $n \geq 0$ be fixed and let $A \in \mathcal{F}_n$. By the martingale property, for $m > n$, we have

$$\mathbb{E}\left(X_m\mathbb{1}_A\right) = \mathbb{E}\left(X_n\mathbb{1}_A\right).$$

Since X_m converges in L^1 to X_∞, letting $m \to \infty$ in the above equality gives,

$$\mathbb{E}(X_\infty\mathbb{1}_A) = \mathbb{E}(X_n\mathbb{1}_A).$$

Since X_n is $\mathcal{F}_n$-measurable, we deduce that $X_n = \mathbb{E}(X_\infty \,|\, \mathcal{F}_n)$. □

Remark 6.1. By the previous theorem, a closed martingale $X_n = \mathbb{E}(Y \,|\, \mathcal{F}_n)$ converges a.s. and in L^1 to some X_∞. Denote by

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$$

the σ-algebra limit of the filtration $(\mathcal{F}_n)_{n\geq 0}$. Then $X_\infty = \mathbb{E}(Y \,|\, \mathcal{F}_\infty)$. Indeed, X_∞ is $\mathcal{F}_\infty$-measurable as the limit of $(X_n)_n$, so it suffices to prove that for all $A \in \mathcal{F}_\infty$,

$$\mathbb{E}(Y\mathbb{1}_A) = \mathbb{E}(X_\infty\mathbb{1}_A).$$

But, since X_∞ converges in L^1 to X_∞, we have seen that for all $n \geq 0$, and all $A \in \mathcal{F}_n$,

$$\mathbb{E}(X_n\mathbb{1}_A) = \mathbb{E}(X_\infty\mathbb{1}_A),$$

hence for all $n \geq 0$, and all $A \in \mathcal{F}_n$,

$$\mathbb{E}(Y \mathbb{1}_A) = \mathbb{E}(X_\infty \mathbb{1}_A).$$

But it is easy to see that

$$\mathcal{M} = \{A \in \mathcal{F} \mid \mathbb{E}(Y \mathbb{1}_A) = \mathbb{E}(X_\infty \mathbb{1}_A)\}$$

is a λ-system[2] which contains the π-system[3] $\bigcup_{n\geq 0} \mathcal{F}_n$. By Dynkin's π-λ theorem, $\mathcal{M}$ contains $\sigma\left(\bigcup_{n\geq 0} \mathcal{F}_n\right)$ and the conclusion follows.

REMARK 6.2. Consider the symmetric random walk $(X_n)_{n\geq 0}$. Since, $\limsup_n X_n = +\infty$ and $\liminf_n X_n = -\infty$, the martingale $(X_n)_{n\geq 0}$ does not converge, which is consistent with the fact that it is not uniformly integrable.

EXAMPLE 6.2 (Kolmogorov's 0–1 law). Let $(X_n)_{n\geq 1}$ be a sequence of random variables. Define, for all $n \geq 1$,

$$\mathcal{F}_n = \sigma(X_k, 1 \leq k \leq n)$$

the σ-algebra generated by $X_1, \ldots, X_n$. Define also, for all $n \geq 0$,

$$\mathcal{G}_n = \sigma(X_{n+1}, X_{n+2}, \ldots).$$

The tail σ-algebra $\mathcal{T}$ is defined as

$$\mathcal{T} = \bigcap_{n\geq 0} \mathcal{G}_n.$$

Intuitively, the σ-algebra of tail events consists of events which do not depend on the first finitely many times of the process $(X_n)_{n\geq 1}$. For instance, the random variables $\limsup_n X_n$ and $\liminf_n X_n$ are $\mathcal{T}$-measurable. The classical Kolmogorov's 0–1 law states that if $(X_n)_{n\geq 1}$ is a sequence of independent random variables, then the tail σ-algebra $\mathcal{T}$ is trivial: for all $A \in \mathcal{T}$, $\mathbb{P}(A) \in \{0, 1\}$. We can give an alternative proof using martingale theory (exercise: prove it directly!).

Let $A \in \mathcal{T}$, and consider the closed martingale defined by

$$X_n = \mathbb{E}\left(\mathbb{1}_A \mid \mathcal{F}_n\right),$$

for all $n \geq 1$. By the previous theorem, $(X_n)_{n\geq 1}$ converges a.s. and in L^1 to $\mathbb{E}\left(\mathbb{1}_A \mid \mathcal{F}_\infty\right)$, where $\mathcal{F}_\infty = \sigma(\cup_{n\geq 1}\mathcal{F}_n)$. But since obviously $\mathcal{T} \subset \mathcal{F}_\infty$, one has $\mathbb{E}\left(\mathbb{1}_A \mid \mathcal{F}_\infty\right) = \mathbb{1}_A$.

But $A \in \mathcal{G}_n$ for all n, so A is independent of $\mathcal{F}_n$ for all n, hence,

$$\mathbb{E}\left(\mathbb{1}_A \mid \mathcal{F}_n\right) = \mathbb{E}\left(\mathbb{1}_A\right),$$

so we get $\mathbb{P}(A) = \mathbb{1}_A$, so $\mathbb{P}(A) \in \{0, 1\}$.

6.3. Optional stopping theorem for uniformly integrable martingales. We now generalize the Doob's optional theorem to non-necessarily bounded stopping times, but for uniformly integrable martingale.

PROPOSITION 6.2. *Let $X = (X_n)_{n\geq 0}$ be a uniformly integrable submartingale and T a finite stopping time. Then $X^T = (X_{n\wedge T})_{n\geq 0}$ is a uniformly integrable submartingale.*

PROOF. By the optional stopping theorem applied to the bounded stopping time $n \wedge T$ and to the submartingale $(X^+_{n\wedge T})_n$, one has that $\mathbb{E}(X^+_{n\wedge T}) \leq \mathbb{E}(X^+_n)$. Hence,

$$\sup_n \mathbb{E}(X^+_{n\wedge T}) \leq \sup_n \mathbb{E}(X^+_n) < \infty,$$

[2]Recall that a λ-system is a collection of subsets which contains Ω and is closed under complements of subsets in supersets and under countable increasing unions.

[3]Recall that a π-system is a collection of subsets closed under finite intersections.

since (X_n) is uniformly integrable, so it is bounded in L^1. Hence, by the martingale convergence theorem (Theorem 5.2), $X_{n\wedge T} \to X_T$ a.s. and $\mathbb{E}(|X_T|) < \infty$. Now, write

$$\begin{aligned}\mathbb{E}(|X_{n\wedge T}|\mathbb{1}_{\{|X_{n\wedge T}|\geq a\}}) &= \mathbb{E}(|X_T|\mathbb{1}_{\{|X_{n\wedge T}|\geq a\}}\mathbb{1}_{\{T\geq n\}}) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}\mathbb{1}_{\{T>n\}})\\ &\leq \mathbb{E}(|X_T|\mathbb{1}_{\{T\geq n\}}) + \mathbb{E}(|X_n|\mathbb{1}_{\{|X_n|\geq a\}}).\end{aligned}$$

But one has $\sup_n \mathbb{E}(|X_T|\mathbb{1}_{\{T\geq n\}}) = 0$ by the dominated convergence theorem. Hence the uniform integrability of $(X_{n\wedge T})_n$ follows from the uniform integrability of $(X_n)_n$. □

Theorem 6.3 (Optional stopping theorem for uniformly integrable martingales). *Let $X = (X_n)_{n\geq 0}$ be a uniformly integrable submartingale. Let $S \leq T$ be two finite stopping time. Then, we have*

$$X_S \leq \mathbb{E}(X_T \,|\, \mathcal{F}_S).$$

Proof. Consider the two bounded stopping times $S\wedge n \leq T\wedge n$. By the optional stopping theorem, one has

$$\mathbb{E}(X_{S\wedge n}) \leq \mathbb{E}(X_{T\wedge n}).$$

Since X^T (resp. X^S) is a uniformly integrable submartingale by the previous proposition, it converges a.s. and in L^1 to X_T (resp. X_S). Hence, letting $n\to\infty$ in the above equality, we get

$$\mathbb{E}(X_S) \leq \mathbb{E}(X_T).$$

Now, let $A \in \mathcal{F}_S$. We have to prove that

$$\mathbb{E}(X_S\mathbb{1}_A) \leq \mathbb{E}(X_T\mathbb{1}_A).$$

Consider the stopping time

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

It is indeed a stopping time since

$$\{R=n\} = (\{S=n\}\cap A)\cup(\{T=n\}\cap A^c) \in \mathcal{F}_n$$

since $A\in\mathcal{F}_S\subset\mathcal{F}_T$. Hence,

$$\mathbb{E}(X_S\mathbb{1}_A) = \mathbb{E}(X_R\mathbb{1}_A) = \mathbb{E}(X_R) - \mathbb{E}(X_R\mathbb{1}_{A^c}) = \mathbb{E}(X_R) - \mathbb{E}(X_T\mathbb{1}_{A^c}),$$

and since $R\leq T$, we have $\mathbb{E}(X_R)\leq\mathbb{E}(X_T)$, hence,

$$\mathbb{E}(X_S\mathbb{1}_A) \leq \mathbb{E}(X_T) - \mathbb{E}(X_T\mathbb{1}_{A^c}) = \mathbb{E}(X_T\mathbb{1}_A),$$

proving that $X_S \leq \mathbb{E}(X_T\,|\,\mathcal{F}_S)$. □

7. Backwards martingales

Definition 7.1. *A backwards filtration is a family of σ-algebras $\mathcal{F}_{-n}$, indexed by non-positive integers $-n\in-\mathbb{N}$ such that*

$$\cdots\subset\mathcal{F}_{-n}\subset\mathcal{F}_{-n+1}\subset\cdots\subset\mathcal{F}_{-1}\subset\mathcal{F}_0$$

that is for all $n\in\mathbb{N}$, $\mathcal{F}_{-n}\subset\mathcal{F}_{-n+1}$. We aslo define:

$$\mathcal{F}_{-\infty} = \bigcap_{n\in-\mathbb{N}}\mathcal{F}_n,$$

which is again a σ-algebra.

Note that here, the σ-algebra $\mathcal{F}_n$ becomes smaller and smaller as $n\to-\infty$.

Definition 7.2. *The stochastic process $X = (X_n)_{n\in-\mathbb{N}}$ is called a backwards martingale (resp. supermartingale, submartingale), if for all $n\in-\mathbb{N}$, X_n is $\mathcal{F}_n$-measurable, $\mathbb{E}(|X_n|)<\infty$, and for all $n\leq m\leq 0$,*

$$X_n = \mathbb{E}(X_m\,|\,\mathcal{F}_n) \quad (\textit{resp. } X_n\geq\mathbb{E}(X_m\,|\,\mathcal{F}_n),\ \textit{resp. } X_n\leq\mathbb{E}(X_m\,|\,\mathcal{F}_n).)$$

Convergence is particularly simple for backwards martingales:

THEOREM 7.1. *Let $(X_n)_{n\in -\mathbb{N}}$ be a backwards martingale. Then, $(X_n)_{n\in -\mathbb{N}}$ is uniformly integrable and converges a.s. and in L^1 as $n \to -\infty$ to some $X_{-\infty}$. Moreover, for all $n \in -\mathbb{N}$,*

$$\mathbb{E}(X_n \,|\, \mathcal{F}_{-\infty}) = X_{-\infty}$$

PROOF. The proof parallels that of the "forward" case. Fixed $K \geq 0$, and define, for all $0 \leq n \leq K$,

$$Y_n = X_{-K+n} \quad \text{and} \quad \mathcal{G}_n = \mathcal{F}_{-K+n},$$

and for $n > K$, we put $Y_n = X_0$ and $\mathcal{G}_n = \mathcal{F}_0$.

Then, $(Y_n)_{n\geq 0}$ is a "forward" martingale with respect to the filtration $(\mathcal{G}_n)_{n\geq 0}$. By Doob's upcrossing inequality, we have

$$(b-a)\,\mathbb{E}\left(U_K^{a,b}\right) \leq \mathbb{E}((Y_K - a)^-) \leq |a| + \mathbb{E}(|X_0|),$$

where $U_n^{a,b}$ denotes the number of upcrossing across $[a,b]$ of Y_n. Letting $K \to \infty$, we find, by monotone convergence theorem, that the number $U^{a,b}$ of upcrossing across $[a,b]$ of $(X_n)_{n\in -\mathbb{N}}$ satisfies

$$\mathbb{E}\left(U^{a,b}\right) < \infty,$$

hence is finite a.s. As in the "forward" case, this implies that $(X_n)_{n\in -\mathbb{N}}$ converges a.s. and Fatou's inequality implies that $X_{-\infty} \in L^1$ (the details are left as an exercise). Moreover, by the backwards martingale property, one has that for all $n \geq 0$,

$$X_{-n} = \mathbb{E}\left(X_0 \,|\, \mathcal{F}_{-n}\right).$$

Hence, one proves that $(X_n)_{n\in -\mathbb{N}}$ is uniformly integrable exactly as in the proof of Thm 6.2. The convergence in L^1 follows by Theorem 6.1.

Now let $A \in \mathcal{F}_{-\infty}$. Since $\mathcal{F}_{-\infty} \subset \mathcal{F}_n$, for all $n \leq 0$, A is also $\mathcal{F}_n$-measurable, and the martingale property gives, that for all $m \leq n \leq 0$,

$$\mathbb{E}(X_m \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A).$$

Hence, letting $m \to -\infty$, and using the L^1-convergence, we obtain that

$$\mathbb{E}(X_{-\infty} \mathbb{1}_A) = \mathbb{E}(X_n \mathbb{1}_A),$$

that is, since $X_{-\infty}$ is $\mathcal{F}_{-\infty}$-measurable, $X_{-\infty} = \mathbb{E}(X_n \,|\, \mathcal{F}_{-\infty})$. □

COROLLARY 7.1. *Let $Y \in L^1$ and $(\mathcal{G}_n)_{n\geq 0}$ a non-increasing sequence of σ-algebras. Let $\mathcal{G}_\infty = \bigcap_{n\geq 0} \mathcal{G}_n$. Then we have,*

$$\mathbb{E}(Y \,|\, \mathcal{G}_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}(Y \,|\, \mathcal{G}_\infty), \quad \textit{a.s. and in } L^1.$$

PROOF. For all $n \geq 0$, define $X_{-n} = \mathbb{E}(Y \,|\, \mathcal{G}_n)$ and $\mathcal{F}_{-n} = \mathcal{G}_n$. Then $(X_{-n})_{n\geq 0}$ is a backwards martingale relative to the backwards filtration $(\mathcal{F}_{-n})_{n\geq 0}$ and we can apply the previous theorem. □

EXAMPLE 7.1 (The strong law of large numbers). Let $(X_n)_{n\geq 1}$ be a sequence of independent and identically distributed random variables, such that $\mathbb{E}(|X_1|) < \infty$. Define $S_0 = 0$, and

$$S_n = X_1 + \cdots + X_n.$$

The strong law of large numbers states that

$$\frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}(X_1) \quad \text{a.s.}$$

We give a new proof of the law of large numbers using backwards martingale. Define, for all $n \geq 0$,

$$Y_{-n} = \frac{S_n}{n},$$

and

$$\mathcal{F}_{-n} = \sigma(S_n, X_{n+1}, X_{n+2}, \ldots).$$

Then the family $(\mathcal{F}_{-n})_{n\geq 0}$ is a backwards filtration, and $(Y_{-n})_{n\geq 0}$ a backwards martingale: We first remark that, for all $1 \leq k \leq n$,

$$\mathbb{E}\left(X_k \,|\, \mathcal{F}_{-n}\right) = \mathbb{E}\left(X_k \,|\, S_n\right) = \mathbb{E}\left(X_n \,|\, S_n\right).$$

Indeed, the first equality uses the fact that X_k is independent of $\sigma(X_{n+1}, X_{n+1}, \ldots)$, and the second one the fact that (X_k, S_n) has the same distribution that (X_n, S_n), for all $1 \leq k \leq n$. Then, writing $S_{n-1} = S_n - X_n$, one has

$$\begin{aligned}\mathbb{E}(Y_{-n+1} \,|\, \mathcal{F}_{-n}) &= \frac{1}{n-1}\,\mathbb{E}(S_{n-1} \,|\, S_n) = \frac{1}{n-1}\Big(S_n - \mathbb{E}(X_n \,|\, S_n)\Big) = \frac{1}{n-1}\left(S_n - \frac{1}{n}S_n\right) \\ &= \frac{S_n}{n} = Y_{-n}.\end{aligned}$$

Hence, $(Y_{-n})_{n\geq 0}$ is a backwards martingale, so it is uniformly integrable, and converges almost surely to

$$Y_{-\infty} = \mathbb{E}\left(X_1 \,|\, \mathcal{F}_{-\infty}\right).$$

Since $Y_{-\infty}$ is measurable relative to the tail σ-algebra

$$\mathcal{T} = \bigcap_{n\geq 0} \sigma(X_n, X_{n+1}, \ldots),$$

which is trivial by Kolmogorov's zero-one law, we have that $Y_{-\infty}$ is constant a.s., so we deduce that

$$\mathbb{E}\left(X_1 \,|\, \mathcal{F}_{-\infty}\right) = \mathbb{E}(X_1).$$

Thus,

$$\frac{S_n}{n} \underset{n\to\infty}{\longrightarrow} \mathbb{E}(X_1) \quad \text{a.s.}$$

8. Examples and applications

8.1. Pólya urns. We consider an urn containing initially 1 red ball and 1 black ball. We draw a ball uniformly at random, put it back in the urn with an additional ball of same color. We repeat the procedure and denotes by X_n the number of red balls at time n. Initially, $X_0 = 1$. At each time n, the total number of balls is $n+2$. It is easy to see that the process $(X_n)_{n\geq 0}$ is a (inhomogeneous) Markov chain with transition probabilities:

$$\begin{aligned}\mathbb{P}(X_{n+1} = k \,|\, X_n = k) &= 1 - \frac{k}{n+2} \\ \mathbb{P}(X_{n+1} = k+1 \,|\, X_n = k) &= \frac{k}{n+2}.\end{aligned}$$

Hence, the conditional distribution of X_{n+1} given X_n is given by the Markov kernel $\nu(X_n, dy)$:

$$\nu(X_n, dy) = \left(1 - \frac{X_n}{n+2}\right)\delta_{X_n} + \frac{X_n}{n+2}\delta_{X_n+1}.$$

Thus, computing the conditional expectation of X_{n+1} given $\mathcal{F}_n$ gives

$$\mathbb{E}(X_{n+1} \,|\, \mathcal{F}_n) = \mathbb{E}(X_{n+1} \,|\, X_n) = \left(1 - \frac{X_n}{n+2}\right)X_n + \frac{X_n}{n+2}(X_n + 1) = \frac{n+3}{n+2}X_n.$$

Hence, if we denote by M_n the proportion of red balls at time n, i.e. $M_n = \frac{X_n}{n+2}$, we get

$$\mathbb{E}(M_{n+1} \,|\, \mathcal{F}_n) = M_n,$$

that is $(M_n)_{n\geq 0}$ is a martingale. Since it is positive and bounded by 1, $(M_n)_{n\geq 0}$ converges a.s. and in every L^p, $p>1$, to some $M_\infty \in [0,1]$ a.s. We now compute the distribution of M_∞. First, the distribution of X_n is uniform on $\{1,\dots,n+1\}$: this is done by induction. It is obvious for $n=0$. Suppose it is true at time n. Then, for all $k=1,\dots,n+1$,

$$\begin{aligned}
\mathbb{P}(X_{n+1}=k) &= \mathbb{P}(X_{n+1}=k \,|\, X_n = k)\,\mathbb{P}(X_n=k) + \mathbb{P}(X_{n+1}=k \,|\, X_n=k-1)\,\mathbb{P}(X_n=k-1)\\
&= \left(1-\frac{k}{n+2}\right)\frac{1}{n+1} + \frac{k-1}{n+2}\frac{1}{n+1}\\
&= \frac{1}{n+2},
\end{aligned}$$

and

$$\mathbb{P}(X_{n+1}=n+2) = \mathbb{P}(X_{n+1}=n+2 \,|\, X_n=n+1)\,\mathbb{P}(X_n=n+1) = \frac{1}{n+2}.$$

Thus, X_{n+1} is uniformly distributed on $\{1,\dots,n+2\}$. One easily deduces that M_∞ is uniformly distributed on $[0,1]$ (exercise).

8.2. Galton-Watson branching process. Let $(X_{n,i})_{n\geq 0, i\geq 1}$ be independent and identically distributed random variables with discrete distribution $\mathbb{P}(X_{1,1}=k)=p_k$, for $k\geq 0$. We assume that $p_0>0$ and $p_0+p_1<1$ to avoid trivialities. We also suppose that $X_{1,1}\in L^2$ and put $m=\mathbb{E}(X_{1,1})$ and $\sigma^2 = \mathbb{V}\mathrm{ar}(X_{1,1})$. We define the branching process $(Z_n)_{n\geq 0}$ by $Z_0=1$, and for $n\geq 0$,

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}.$$

We interpret $X_{n,i}$ as the number of children of the i-th individual of the n-th generation, and Z_n as the size of the population at time n. It is easy to see (exercise) that $(Z_n)_{n\geq 0}$ is a Markov chain with transition kernel $Q(x,y)=\mu^{*x}(y)$, where μ denotes the distribution of $X_{1,1}$ and μ^{*x} the x-times convolution of μ with itself.

Now, define, for all $n\geq 0$, $W_n = m^{-n}Z_n$. Then, $(W_n)_n$ is a martingale relative to the filtration defined by $\mathcal{F}_n = \sigma(X_{k,i}, k<n, i\geq 1)$:

$$\begin{aligned}
\mathbb{E}(W_{n+1} \,|\, \mathcal{F}_n) &= m^{-(n+1)}\,\mathbb{E}(Z_{n+1} \,|\, \mathcal{F}_n)\\
&= m^{-(n+1)}\,\mathbb{E}\left(\sum_{i=1}^{Z_n} X_{n,i} \,|\, \mathcal{F}_n\right)\\
&= m^{-(n+1)}\sum_{k\geq 1}\mathbb{E}\left(\mathbb{1}_{\{Z_n=k\}}\sum_{i=1}^{k} X_{n,i} \,|\, \mathcal{F}_n\right)\\
&= m^{-(n+1)}\sum_{k\geq 1}\mathbb{1}_{\{Z_n=k\}}\,\mathbb{E}\left(\sum_{i=1}^{k} X_{n,i} \,|\, \mathcal{F}_n\right)\\
&= m^{-(n+1)}\sum_{k\geq 1}\mathbb{1}_{\{Z_n=k\}}\,\mathbb{E}\left(\sum_{i=1}^{k} X_{n,i}\right),
\end{aligned}$$

where we use the fact that Z_n is $\mathcal{F}_n$-measurable in the penultimate equality, and the independence of $X_{n,i}$ and $\mathcal{F}_n$ in the last one. Hence, since the random variables $X_{n,i}$ have the same distribution, $\mathbb{E}\left(\sum_{i=1}^{k} X_{n,i}\right) = km$, hence

$$\mathbb{E}(W_{n+1} \,|\, \mathcal{F}_n) = m^{-n} \sum_{k\geq 1} \mathbb{1}_{\{Z_n=k\}} k = m^{-n} Z_n = W_n,$$

proving that $(W_n)_n$ is a martingale. Since $(W_n)_n$ is non-negative, it converges almost surely to some finite random variable W_∞. Moreover, by Fatou's inequality, we have $\mathbb{E}(W_\infty) \leq \mathbb{E}(W_0) = 1$.

A central question in branching process theory is to establish when extinction of the population occurs. The probability of extinction is defined by

$$\pi = \lim_n \mathbb{P}(Z_n = 0).$$

Observe that since Z_n is integer valued, Z_n goes to zero if and only if there exists n such that $Z_n = 0$, and thus for all $k \geq n$, $Z_k = 0$. Moreover, one can see that if Z_n converges, then it is to 0 or $+\infty$. Indeed, let $j \geq 1$. Then, using the Markov property, one has

$$\begin{aligned}\mathbb{P}\left(\bigcap_{n=k}^{N} \{X_n = j\}\right) &= \mathbb{P}(X_N = j \,|\, X_{N-1} = j)\, \mathbb{P}(X_{N-1} = j \,|\, X_{N-2} = j) \cdots \mathbb{P}(X_k = j) \\ &= \left(Q(j,j)\right)^{N-k} \mathbb{P}(X_k = j).\end{aligned}$$

Since by assumption $p_0 > 0$, we have $\mu^{*j}(0) > 0$ and thus $Q(j,j) < 1$. Hence,

$$\mathbb{P}\left(\bigcap_{n\geq k} \{X_n = j\}\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n=k}^{N} \{X_n = j\}\right) = 0,$$

and we deduce that $\mathbb{P}(\liminf_n X_n = j) = 0$. Hence, all states $j \geq 1$ are visited only a finite number of times, so Z_n can not converge to $j \neq 0$. We thus have:

- (Subcritical case). Let $m < 1$. then $Z_n = m^n W_n$ converges a.s. to 0. We have extinction of the population with probability one.
- (Critical case) Let $m = 1$. Then $Z_n = W_n$ converges a.s. to W_∞, which is finite a.s. and $\mathbb{E}(W_\infty) \leq 1$. Again, we must have $W_\infty = 0$ and we have extinction of the population with probability one. Note that since $\mathbb{E}(Z_n) = 1$ for all $n \geq 1$, the convergence does not hold in L^1.
- (Supercritical case) Let $m > 1$. We prove that W_n is bounded in L^2. Indeed,

$$\begin{aligned}\mathbb{E}\left((Z_n)^2 \,|\, \mathcal{F}_{n-1}\right) &= \sum_{k\geq 0} \mathbb{1}_{\{Z_{n-1}=k\}}\, \mathbb{E}\left(\left(\sum_{i=1}^{k} X_{n,i}\right)^2\right) \\ &= \sum_{k\geq 0} \mathbb{1}_{\{Z_{n-1}=k\}} \left(k\sigma^2 + k^2 m^2\right) \\ &= Z_{n-1}\sigma^2 + (Z_{n-1})^2 m^2,\end{aligned}$$

  hence

$$\mathbb{E}\left((Z_n)^2\right) = \mathbb{E}(Z_{n-1})\sigma^2 + \mathbb{E}\left((Z_{n-1})^2\right) m^2$$

  and thus, since $\mathbb{E}(Z_{n-1}) = m^{n-1}$ since W_n is a martingale, we get

$$\mathbb{E}\left((W_n)^2\right) = \frac{1}{m^{n+1}}\sigma^2 + \mathbb{E}\left((W_{n-1})^2\right).$$

  Hence,

$$\mathbb{E}\left((W_n)^2\right) = \sum_{k=0}^{n+1} \frac{1}{m^k}\sigma^2 \leq \sigma^2 \frac{1}{1-\frac{1}{m}}.$$

So $\sup_n \mathbb{E}\left((W_n)^2\right) < \infty$, and $(W_n)_n$ converges in L^2 and also in L^1 to W_∞. In particular, $\mathbb{E}(W_\infty) = \mathbb{E}(W_0) = 1$. Hence, $\mathbb{P}(W_\infty > 0) > 0$, and thus the population has an exponential growth with positive probability.

8.3. Martingale proof of the Radon-Nikodym theorem. Recall that the Radon-Nikodym theorem states that if $(E, \mathcal{E})$ is a measurable space and λ and ν are two σ-finite measures on $(E, \mathcal{E})$ such that $\nu \ll \lambda$ (i.e. for all $A \in \mathcal{E}$, $\lambda(A) = 0 \Rightarrow \nu(A) = 0$), there exists a measurable function $f\colon E \to [0, +\infty[$ such that for all $A \in \mathcal{E}$,

$$\nu(A) = \int_A f d\lambda.$$

The function f is unique λ-almost everywhere, and is called the Radon-Nikodym derivative, and is usually denoted by $\frac{d\nu}{d\lambda}$.

We give below a proof of this theorem, using martingales, in the special case of the probability space $([0,1[, \mathcal{B}([0,1[), \lambda)$, where λ denotes Lebesgue measure, but the argument can be generalized in a straightforward way to probability spaces $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathcal{F}$ is generated by a countable collection of sets.

Recall that the dyadic filtration on $\mathcal{B}([0,1[)$ is defined by,

$$\mathcal{F}_n = \sigma\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[, k = 0, 1, \dots, 2^n - 1\right), \quad n \geq 0.$$

Let ν be another probability measure on $([0,1[, \mathcal{B}([0,1[)$ which is absolutely continuous with respect to λ. Define $(f_n)_{n\geq 0}$ by

$$f_n(\omega) = \sum_{k=0}^{2^n-1} 2^n \nu\left(\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[\right) \mathbb{1}_{\left[\frac{k}{2^n}, \frac{k+1}{2^n}\right[}(\omega).$$

Then, $(f_n)_{n\geq 0}$ is a $(\mathcal{F}_n)_{n\geq 0}$-martingale: we have to prove that for all $A \in \mathcal{F}_n$,

$$\int f_{n+1} \mathbb{1}_A d\lambda = \int f_n \mathbb{1}_A d\lambda.$$

Since $\mathcal{F}_n$ is finitely generated, it suffices to prove the above equality for A of the form $\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right[$, for some $i = 0, 1, \dots, 2^n - 1$. Hence, writing $A = \left[\frac{2i}{2^{n+1}}, \frac{2i+1}{2^{n+1}}\right[\cup \left[\frac{2i+1}{2^{n+1}}, \frac{2i+2}{2^{n+1}}\right[$, we have,

$$\begin{aligned}
\int f_{n+1} \mathbb{1}_A d\lambda &= \sum_{k=0}^{2^{n+1}-1} 2^{n+1} \nu\left(\left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right[\right) \int \mathbb{1}_{\left[\frac{k}{2^{n+1}}, \frac{k+1}{2^{n+1}}\right[} \mathbb{1}_A d\lambda \\
&= \nu\left(\left[\frac{2i}{2^{n+1}}, \frac{2i+1}{2^{n+1}}\right[\right) + \nu\left(\left[\frac{2i+1}{2^{n+1}}, \frac{2i+2}{2^{n+1}}\right[\right) \\
&= \nu\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n}\right[\right) \\
&= \int f_n \mathbb{1}_A d\lambda.
\end{aligned}$$

Eventually, we get that for all $A \in \mathcal{F}_n$,

$$\int f_{n+1} \mathbb{1}_A d\lambda = \nu(A) = \int f_n \mathbb{1}_A d\lambda.$$

Thus, $(f_n)_{n\geq 0}$ is a martingale and is non-negative, hence it converges a.s. to some f_∞. Moreover, we show that $(f_n)_{n\geq 0}$ is uniformly integrable. First, since ν is absolutely continuous with respect to λ, then for all $\varepsilon > 0$, there exist $\delta > 0$ such that $\lambda(A) < \delta \Rightarrow \nu(A) < \varepsilon$. Indeed, if it were not the case, there exists A_n such that $\lambda(A_n) < \frac{1}{2^n}$ and $\nu(A_n) > \varepsilon$. By Borel-Cantelli lemma, we get $\lambda(\limsup_n A_n) = 0$, and since

$$\nu(\limsup_n A_n) \geq \limsup_n \nu(A_n) \geq \varepsilon > 0,$$

this contradicts the fact that $\nu \ll \lambda$. Now, we have, $\{f_n \geq a\}$ being in $\mathcal{F}_n$,

$$\int f_n \mathbb{1}_{\{f_n \geq a\}} d\lambda = \nu(f_n \geq a),$$

and Markov inequality gives

$$\lambda(f_n \geq a) \leq \frac{\int f_n d\lambda}{a} = \frac{1}{a}.$$

Hence, choosing a large enough, we get $\sup_n \int f_n \mathbb{1}_{\{f_n \geq a\}} d\lambda < \varepsilon$, hence $(f_n)_{n\geq 0}$ is uniformly integrable. Therefore, $(f_n)_{n\geq 0}$ is a closed martingale such that

$$f_n = \mathbb{E}(f_\infty \,|\, \mathcal{F}_n),$$

where $\mathbb{E}$ denotes expectation with respect to λ. As such, we get that for all $A \in \mathcal{F}_n$,

$$\nu(A) = \int f_n \mathbb{1}_A d\lambda = \int f_\infty \mathbb{1}_A d\lambda.$$

Since $\mathcal{B}([0,1[)$ is generated by the filtration $(\mathcal{F}_n)_{n\geq 0}$, Dynkin's π-λ theorem implies that for all $A \in \mathcal{B}([0,1[)$,

$$\nu(A) = \int f_\infty \mathbb{1}_A d\lambda,$$

giving that the Radon-Nikodym derivative is $f_\infty = \frac{d\nu}{d\lambda}$.

8.4. Classical convergence theorems. Martingales allow us to relax the independence assumption in classical theorems such as the law of large numbers and the central limit theorem by considering them in place of sums of independent random variables.

8.4.1. *The law of large numbers.*

THEOREM 8.1. *Let $X = (X_n)_{n\geq 0}$ be a square integrable martingale. On the set $\{\langle X\rangle_\infty = +\infty\}$, we have*

$$\frac{X_n}{\langle X\rangle_n} \underset{n\to\infty}{\longrightarrow} 0, \quad a.s.$$

REMARK 8.1. If $S_n = \varepsilon_1 + \cdots \varepsilon_n$ is a sum of n i.i.d. random variables, centered with variance equal to 1, the square variation process of S_n is $\langle S\rangle_n = n$. Hence, from the above theorem, one recovers the strong law of large numbers under the condition of a finite second moment.

PROOF. Assume without loss of generality that $X_0 = 0$. Define the process H by

$$H_n = \frac{1}{1 + \langle X\rangle_n},$$

for $n \geq 0$. Then, H is predictable (since $\langle X\rangle$ is predictable) and bounded by 1 on $\{\langle X\rangle_\infty = +\infty\}$. Now define Y by the martingale transform $Y = H \cdot X$. We are going to show that Y is bounded in L^2. We have, by definition of Y, $(\Delta Y_k)^2 = H_k^2 (\Delta X_k)^2$, hence

$$\mathbb{E}\left((\Delta Y_k)^2\right) = \mathbb{E}\left(H_k^2\, \mathbb{E}\left((\Delta X_k)^2 \,|\, \mathcal{F}_{k-1}\right)\right),$$

since H_k is $\mathcal{F}_{k-1}$-measurable. Now, recall the formula

$$\langle X\rangle_k = \sum_{j=1}^{k} \mathbb{E}\left((X_j - X_{j-1})^2 \,|\, \mathcal{F}_{j-1}\right).$$

Hence,

$$\Delta\langle X\rangle_k = \mathbb{E}\left((\Delta X_k)^2 \,|\, \mathcal{F}_{k-1}\right),$$

so we get

$$\sum_{k=1}^{n} \mathbb{E}\left((\Delta Y_k)^2\right) = \sum_{k=1}^{n} \mathbb{E}\left(H_k^2 \Delta\langle X\rangle_k\right) = \sum_{k=1}^{n} \mathbb{E}\left(\frac{\langle X\rangle_k - \langle X\rangle_{k-1}}{(1+\langle X\rangle_k)^2}\right).$$

Since $\langle X\rangle$ is non-decreasing, $\langle X\rangle_{k-1} \leq \langle X\rangle_k$, thus

$$\frac{\langle X\rangle_k - \langle X\rangle_{k-1}}{(1+\langle X\rangle_k)^2} \leq \frac{\langle X\rangle_k - \langle X\rangle_{k-1}}{(1+\langle X\rangle_k)(1+\langle X\rangle_{k-1})} = \frac{1}{1+\langle X\rangle_k} - \frac{1}{1+\langle X\rangle_{k-1}}.$$

Taking the sum over k, one gets

$$\sum_{k=1}^{n} \mathbb{E}\left((\Delta Y_k)^2\right) \leq \mathbb{E}\left(1 - \frac{1}{1+\langle X\rangle_n}\right) \leq 1.$$

Recall that the increments of a square integrable martingale are orthogonal in L^2. Thus,

$$\mathbb{E}(Y_n^2) = \sum_{k=1}^{n} \mathbb{E}\left((\Delta Y_k)^2\right) \leq 1.$$

Hence, Y is bounded in L^2, and thus converges a.s. and in L^2 by the martingale convergence theorem. We conclude using Kronecker's lemma: if $(x_n)_n$ is a real sequence such that

$$\sum_n x_n < \infty,$$

then for all positive and non-decreasing sequence $(b_n)_n$ such that $b_n \to \infty$,

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} b_k x_k = 0.$$

We apply Kronecker's lemma to $x_n = \frac{\Delta X_n}{1+\langle X\rangle_n}$ and $b = 1 + \langle X\rangle_n$, giving that on $\{\langle X\rangle_\infty = +\infty\}$,

$$\frac{1}{1+\langle X\rangle_n} \sum_{k=1}^{n} \Delta X_n \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

which proves the theorem. □

8.4.2. *The central limit theorem.*

THEOREM 8.2 (Central limit theorem). *Let $(X_n)_{n\geq 1}$ be a sequence of random variables such that: for all $k \geq 1$,*

- $\mathbb{E}(X_k \,|\, \mathcal{F}_{k-1}) = 0$*;*
- $\mathbb{E}(X_k^2 \,|\, \mathcal{F}_{k-1}) = 1$*;*
- $\mathbb{E}(|X_k|^3 \,|\, \mathcal{F}_{k-1}) \leq K$ *for some constant $K > 0$.*

Define, $(S_n)_{n\geq 0}$ by $S_0 = 0$, and $S_n = X_1 + \cdots + X_n$, for $n \geq 1$. Then, we have

$$\frac{S_n}{\sqrt{n}} \xrightarrow[n\to\infty]{(d)} \mathcal{N}(0,1).$$

PROOF. Let $u \in \mathbb{R}$. By Taylor-Lagrange formula, we have

$$e^{iu\frac{1}{\sqrt{n}}X_k} = 1 + iu\frac{1}{\sqrt{n}}X_k - \frac{u^2}{2n}X_k^2 + R_k(u),$$

where the remainder term satisfies $|R_k(u)| \leq \frac{|u|^3}{6n^{3/2}}|X_k|^3$. Now, we write,

$$\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_k}\right) = \mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_{k-1}}e^{iu\frac{1}{\sqrt{n}}X_k}\right) = \mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_{k-1}}\,\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}X_k} \,|\, \mathcal{F}_{k-1}\right)\right),$$

since S_{k-1} is $\mathcal{F}_{k-1}$-measurable. By assumption, we have

$$\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}X_k}\,|\,\mathcal{F}_{k-1}\right) = 1 - \frac{u^2}{2n} + \mathbb{E}\left(R_k(u)\,|\,\mathcal{F}_{k-1}\right),$$

hence,

$$\left|\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_k} - \left(1-\frac{u^2}{2n}\right)e^{iu\frac{1}{\sqrt{n}}S_{k-1}}\right)\right| \leq \frac{|u|^3}{6n^{3/2}}\,\mathbb{E}(|X_k|^3\,|\,\mathcal{F}_{k-1}) \leq \frac{|u|^3}{6n^{3/2}}K.$$

For n large enough, we have $0 \leq 1 - \frac{u^2}{2n} \leq 1$, hence

$$\left|\left(1-\frac{u^2}{2n}\right)^{n-k}\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_k}\right) - \left(1-\frac{u^2}{2n}\right)^{n-p+1}\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_{k-1}}\right)\right| \leq \frac{|u|^3}{6n^{3/2}}K.$$

Writing

$$\begin{aligned}\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_n}\right) - \left(1-\frac{u^2}{2n}\right)^n = \sum_{k=1}^n \Bigg[&\left(1-\frac{u^2}{2n}\right)^{n-k}\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_k}\right)\\ &-\left(1-\frac{u^2}{2n}\right)^{n-p+1}\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_{k-1}}\right)\Bigg],\end{aligned}$$

and using the triangle inequality, we get

$$\left|\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_n}\right) - \left(1-\frac{u^2}{2n}\right)^n\right| \leq \frac{|u|^3}{6\sqrt{n}}K.$$

Hence,

$$\left|\mathbb{E}\left(e^{iu\frac{1}{\sqrt{n}}S_n}\right) - \left(1-\frac{u^2}{2n}\right)^n\right| \underset{n\to\infty}{\longrightarrow} 0,$$

and since $\left(1-\frac{u^2}{2n}\right)^n \to e^{-u^2/2}$, we conclude using Levy's theorem. □

8.5. Martingales and Markov chains. Let A be a finite subset of $\mathbb{Z}^d$. The boundary of A is defined by

$$\partial A = \left\{x \in \mathbb{Z}^d \setminus A \,|\, \exists\, y \in A, |y-x| = 1\right\},$$

where $|\cdot|$ denotes the usual L^1-norm. The closure of A is defined as $\overline{A} = A \cup \partial A$, and A is said to be the interior of $\overline{A}$.

The discrete Laplacian Δ on $\mathbb{Z}^d$ is defined by

$$\Delta f(x) = \frac{1}{2d}\sum_{i=1}^d \left(f(x+e_i) + f(x-e_i) - f(x)\right),$$

where $(e_i)_{i=1,\ldots,d}$ is the canonical basis. A function h on $\overline{A}$ is said to be harmonic on A if

$$\Delta h(x) = 0, \quad \text{for } x \in A,$$

that is to say that h has the discrete mean-value property:

$$h(x) = \frac{1}{2d}\sum_{\substack{y\in\mathbb{Z}^d\\ |y-x|=1}} h(y), \quad \text{for } x \in A,$$

i.e. for all x in the interior A of the domain, $h(x)$ is the average of the function values on the boundary of A. Now, let g be a bounded function on ∂A. Then, h is a solution of the Dirichlet problem on A if

$$\begin{cases} \Delta h = 0 & \text{on } A, \\ h = g & \text{on } \partial A. \end{cases}$$

Now if $(X_n)_{n\geq 0}$ is the simple random walk on $\mathbb{Z}^d$, its transition kernel is given by

$$Qf(x) = \frac{1}{2d}\sum_{i=1}^{d} \left(f(x+e_i) + f(x-e_i)\right).$$

Hence, the discrete Laplacian Δ corresponds to the operator $\Delta = I - Q$. This motivates the following.

Let E be a countable state space and let $(X_n)_{n\geq 0}$ be the canonical Markov chain on E with transition probability Q, starting from $x \in E$. Recall that for a non-negative function on E, one has

$$Qf(x) = \sum_{y\in E} f(y)Q(x,y) = \mathbb{E}_x\left(f(X_1)\right).$$

Definition 8.1. *A non-negative function $f\colon E \to \mathbb{R}_+$ is called harmonic if*

$$Qf = f,$$

that is, for all $x \in E$,

$$\sum_{y\in E} f(y)Q(x,y) = f(x).$$

A non-negative function $f\colon E \to \mathbb{R}_+$ is called superharmonic if

$$Qf \geq f.$$

More generally, for a non-empty subset $A \subset E$, the function f is said to be harmonic (resp. superharmonic) on A if for all $x \in A$, $Qf(x) = f(x)$ (resp. $Qf(x) \geq f(x)$).

A link between martingales and Markov chains can be made through harmonic functions:

Proposition 8.1. *Let $f\colon E \to \mathbb{R}_+$. The function f is harmonic (resp. superharmonic) if and only if for all $x \in E$, the stochastic process $(f(X_n))_{n\geq 0}$ is a martingale (resp. supermartingale) under $\mathbb{P}_x$ relative to its natural filtration.*

Proof. Let f be harmonic. Hence, $Qf = f$, and by induction $Q^n f = f$. Hence, since $\mathbb{E}_x\left(f(X_n)\right) = Q^n f(x) < \infty$, $f(X_n)$ is integrable for all $n \geq 0$. Moreover, by the Markov property, the conditional distribution of X_{n+1} given $\mathcal{F}_n$ is given by the kernel Q, that is

$$\mathbb{P}_x(X_{n+1} \in dy \,|\, \mathcal{F}_n) = Q(X_n, dy),$$

hence:

$$\mathbb{E}_x\left(f(X_{n+1}) \,|\, \mathcal{F}_n\right) = \sum_{y\in E} f(y)Q(X_n,y) = Qf(X_n) = f(X_n),$$

since f is harmonic. Hence, $(f(X_n))_{n\geq 0}$ is a martingale.

Conversely, for all $x \in E$, the martingale property implies that

$$\mathbb{E}_x\left(f(X_1)\right) = \mathbb{E}_x\left(f(X_0)\right),$$

that is, $Qf(x) = f(x)$, for all $x \in E$. $\square$

In the same way, we have:

Proposition 8.2. *Let $A \subset E$ be a non-empty subset and let*

$$\tau_A = \inf\{n \geq 0 \mid X_n \in A\}$$

be the first hitting time of A. Let $f\colon E \to \mathbb{R}_+$. If the function f is harmonic (resp. superharmonic) on A^c, then for all $x \in A^c$, the stopped process $X^{\tau_A} = (X_{n\wedge\tau_A})_{n\geq 0}$ is a martingale (resp. a supermartingale) under $\mathbb{P}_x$.

Proof. We have

$$\mathbb{E}_x\left(f(X_{(n+1)\wedge\tau_A}) \mid \mathcal{F}_n\right) = \mathbb{E}_x\left(f(X_{n+1})\mathbb{1}_{\{\tau_A>n\}} \mid \mathcal{F}_n\right) + \mathbb{E}_x\left(f(X_{\tau_A})\mathbb{1}_{\{\tau_A\leq n\}} \mid \mathcal{F}_n\right).$$

Since τ_A is a stopping time, we have $\{\tau_A > n\} \in \mathcal{F}_n$, hence,

$$\begin{aligned}\mathbb{E}_x\left(f(X_{(n+1)\wedge\tau_A})\mathbb{1}_{\{\tau_A>n\}} \mid \mathcal{F}_n\right) &= \mathbb{1}_{\{\tau_A>n\}}\,\mathbb{E}_x\left(f(X_{n+1}) \mid \mathcal{F}_n\right)\\ &= \mathbb{1}_{\{\tau_A>n\}}Qf(X_n)\\ &= \mathbb{1}_{\{\tau_A>n\}}f(X_n),\end{aligned}$$

since f is harmonic on A^c and on $\{\tau_A > n\}$, $X_n \in A^c$. For the second term, we just remark that $f(X_{\tau_A})\mathbb{1}_{\{\tau_A\leq n\}} = f(X_{n\wedge\tau_A})\mathbb{1}_{\{\tau_A\leq n\}}$ is $\mathcal{F}_n$ measurable. Hence,

$$\mathbb{E}_x\left(f(X_{(n+1)\wedge\tau_A}) \mid \mathcal{F}_n\right) = f(X_n)1_{\{\tau_A>n\}} + f(X_{n\wedge\tau_A})\mathbb{1}_{\{\tau_A\leq n\}} = f(X_{n\wedge\tau_A}),$$

so $(X_{n\wedge\tau_A})_{n\geq 0}$ is a martingale. □

The following theorem now gives a solution of the discrete Dirichlet problem.

Theorem 8.3. *Let $A \subset E$ be a non-empty subset. Let $g\colon A \to \mathbb{R}_+$ be a bounded function. Suppose that the hitting time of A is finite a.s. Define the function h on E by:*

$$h(x) = \mathbb{E}_x(g(X_{\tau_A})), \quad x \in E.$$

Then, h is the unique bounded function on E such that h is harmonic on A^c and coincides with g on A.

Proof. If $x \in A$, then $\tau_A = 0$ $\mathbb{P}_x$-a.s., hence $h(x) = \mathbb{E}_x(g(X_0)) = g(x)$, so h and g coincide on A. If $x \in A^c$, $\tau_A = 1 + \tau_A \circ \theta_1$ a.s., where $(\theta_n)_n$ denotes the shift operator, hence

$$h(x) = \mathbb{E}_x(g(X_{\tau_A})) = \mathbb{E}_x(\mathbb{E}_x(g(X_{\tau_A}) \circ \theta_1 \mid \mathcal{F}_1)) = \mathbb{E}_x(\mathbb{E}_{X_1}(g(X_{\tau_A}))),$$

by the Markov property. Hence,

$$h(x) = \mathbb{E}_x(h(X_1)) = Qh(x),$$

so h is harmonic on A^c. It remains to prove the unicity. Let f be another bounded function which is harmonic on A^c and coincide with g on A. For $x \in A$, $f(x) = g(x) = h(x)$, so f and h coincide on A. For $x \in A^c$, define $Y_n = f(X_{n\wedge\tau_A})$, for all $n \geq 0$. By the previous proposition, Y is a martingale under $\mathbb{P}_x$. Since it is bounded, by the martingale convergence theorem Y_n converges $\mathbb{P}_x$-a.s. and in L^1 to $f(X_{\tau_A}) = g(X_{\tau_A})$, since $X_{\tau_A} \in A$. Hence,

$$f(x) = \mathbb{E}_x(Y_0) = \mathbb{E}_x(Y_n) \to \mathbb{E}_x(g(X_{\tau_A})),$$

by the L^1-convergence. Hence, $f = h$ on A^c. □