Université Paris–Dauphine 2025-2026
M1 MIDO

Discrete Processes

Teachers

Julien Claisse, office F412, e-mail : `claisse@ceremade.dauphine.fr`
Eleanor Archer, office F439, e-mail : `archer@ceremade.dauphine.fr`
Clément Cosco, office P212, e-mail : `cosco@ceremade.dauphine.fr`
Melissa Gonzalez Garcia, office C608, e-mail : `melissa.gonzalez-garcia@dauphine.eu`
François Simenhaus, office F422, e-mail : `simenhaus@ceremade.dauphine.fr`

Evaluation

There are a mid-term exam (P) and a terminal exam (E). The final mark is given by the formula $\max\{0.4P + 0.6E, E\}$.

Outline

This course is made up of three parts:

1. Conditional Expectation
2. Martingales
3. Markov Chains

These notes are largely inspired by those of F. Simenhaus and F. Huveneers. A first version of these notes was written together with P. Cardaliaguet.

Please feel free to point out any errors in these notes!

Contents

1 Conditional Expectation **3**
1.1 Definition of Conditional Expectation . . . 3
1.2 Discrete Random Variables . . . 4
1.3 Random Variables with a Density . . . 6
1.4 Random Variables in $L^2(\Omega)$ . . . 7
1.5 Random variables in $L^1(\Omega)$ . . . 8
1.6 Basic Properties . . . 9
1.7 Gaussian Random Variables . . . 11

2 Martingales **12**
2.1 Definition and Examples . . . 12
2.2 Stopping Times . . . 14
2.3 Martingale Transform and Stopped Process . . . 15
2.4 Optional Stopping Theorem . . . 17
2.5 Maximal Inequalities . . . 19
2.6 Convergence of Martingales . . . 21

1 Conditional Expectation

1.1 Definition of Conditional Expectation

The notion of conditional expectation is given by the following theorem, which indicates that this notion is well defined. We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$.

Theorem 1. *Let $X \in L^1$. There is a unique random variable, denoted $\mathbb{E}(X|\mathcal{G})$, with the following three properties:*

1. *$\mathbb{E}(X|\mathcal{G})$ is measurable with respect to $\mathcal{G}$.*
2. *$\mathbb{E}(X|\mathcal{G})$ is integrable.*
3. *For any random variable Z bounded and measurable with respect to $\mathcal{G}$,*

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Z).$$

Remark 1. *Note that $\mathbb{E}(X|\mathcal{G})$ is a random variable, not a number. By definition, it is characterized only up to a negligible set.*

Remark 2. *When $\mathcal{G} = \sigma(Y)$, where Y is a random variable, we write*

$$\mathbb{E}(X|Y) := \mathbb{E}(X|\sigma(Y)).$$

Note that $\mathbb{E}(X|Y)$ is a deterministic function of Y by Lemma 1 below.

Remark 3. *The third condition can be replaced by one of the following equivalent conditions:*

3'. For any event $A \in \mathcal{G}$,

$$\mathbb{E}(X\mathbf{1}_A) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})\mathbf{1}_A).$$

3". For any random variable Z measurable with respect to $\mathcal{G}$ and such that XZ is integrable,

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Z).$$

The first one is sometimes easier to verify while the second one is more general and can be useful in practice.

Remark 4. *As in the classic framework of Lebesgue integral, we can also define the conditional expectation for a nonnegative (but not necessarily integrable) random variable X. Indeed we can show existence of a unique random variable, still denoted $\mathbb{E}(X|\mathcal{G})$, taking values in $[0, +\infty]$, such that*

1. *$\mathbb{E}(X|\mathcal{G})$ is $\mathcal{G}-$measurable.*

2. For any random variable Z nonnegative and measurable with respect to $\mathcal{G}$,

$$\mathbb{E}(XZ) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Z).$$

(both expectations can be equal to $+\infty$)

Theorem 1 will be proven later, after a discussion on some specific cases where the calculation of the conditional expectation is simple and intuitive.

1.2 Discrete Random Variables

Conditioning with respect to an event: Recall that, if A and B are two events (i.e., $A, B \in \mathcal{F}$), with $\mathbb{P}(B)$ >0, then

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

This naturally leads to defining the conditional expectation of an random variable $X \in L^1$ given B (with $B \in \mathcal{F}$ and $\mathbb{P}(B) > 0$) as

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(X\mathbf{1}_B)}{\mathbb{P}(B)}.$$

This corresponds to the expectation of X under the probability measure $\mathbb{P}(\cdot|B)$.

Remark 5. *Note the difference in nature with the conditional expectation with respect to a σ-algebra defined in Theorem 1: $\mathbb{E}(X|B)$ is a number while $\mathbb{E}(X|\mathcal{G})$ (where $\mathcal{G}$ is a σ-field) is a random variable.*

The objective of this part is to explain the link between the two notions when $\mathcal{G} = \sigma(Y)$, with Y a discrete random variable. Recall that a random variable Y on Ω is said to be *discrete* if it takes only a finite or countable number of values a.s., *i.e.*, $E = \{y \in \mathbb{R}; \mathbb{P}(Y = y) > 0\}$ is finite or countable and $\sum_{y \in E} \mathbb{P}(Y = y) = 1$. In other words, it satisfies

$$Y = \sum_{y \in E} y\mathbf{1}_{Y=y} \quad \text{a.s.},$$

where we denote by $\mathbf{1}_A$ the indicator function of a set A.

For instance, if Y follows a Bernoulli distribution with parameter $p \in (0, 1)$,

$$Y = 0 \times \mathbf{1}_{Y=0} + 1 \times \mathbf{1}_{Y=1}, \qquad \text{with } \mathbb{P}(Y = 1) = p.$$

If Y follows a Poisson distribution with parameter $\lambda > 0$, then

$$Y = \sum_{n=0}^{\infty} n\mathbf{1}_{Y=n}, \qquad \text{with } \mathbb{P}(Y = n) = \frac{\lambda^n}{n!}e^{-\lambda} > 0.$$

Let Y be a discrete random variable and X an integrable random variable. Then the expression of $\mathbb{E}(X|Y)$ is explicit, given by the proposition below.

Proposition 1. *Let Y be a discrete random variable and $X \in L^1$. Consider the function*

$$u : y \in \mathbb{R} \mapsto u(y) = \begin{cases} \mathbb{E}(X|Y=y) & \text{if } \mathbb{P}(Y=y) > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\mathbb{E}(X|Y) = u(Y) = \sum_{y \in E} \frac{\mathbb{E}(X\,\mathbf{1}_{Y=y})}{\mathbb{P}(Y=y)} \mathbf{1}_{Y=y}.$$

Remark 6. *The value of u on $N = \{y \in \mathbb{R};\ \mathbb{P}(Y=y) = 0\}$ is irrelevant as* $\mathbb{P}(Y \in N) = 0$.

Remark 7. *If X is also discrete, then we have another meaningful formulation for the conditional expectation by observing that*

$$\mathbb{E}(X|Y=y) = \frac{\mathbb{E}(X\,\mathbf{1}_{Y=y})}{\mathbb{P}(Y=y)} = \sum_x x\,\mathbb{P}\left(X = x|Y=y\right),$$

which follows by writing $X = \sum_x x\mathbf{1}_{X=x}$ and permuting $\sum_x$ and $\mathbb{E}$.

Before the proof of Proposition 1, we recall a well-known result from measure theory which will prove very useful in the sequel.

Lemma 1. *A random variable Z is measurable with respect to $\sigma(Y)$ if and only if there exists a Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $Z = f(Y)$.*

Proof of Proposition 1. As u is a Borel function, $u(Y)$ is measurable with respect to $\sigma(Y)$. Moreover $u(Y)$ is integrable since

$$\mathbb{E}(|u(Y)|) = \sum_{y \in E} \frac{|\mathbb{E}(X\,\mathbf{1}_{Y=y})|}{\mathbb{P}(Y=y)} \mathbb{E}(\mathbf{1}_{Y=y}) \leq \sum_{y \in E} \mathbb{E}(|X|\,\mathbf{1}_{Y=y}) = \mathbb{E}(|X|) < \infty.$$

Finally let Z be a bounded random variable which is measurable with respect to $\sigma(Y)$. According to Lemma 1, there exists a bounded Borel function $f : \mathbb{R} \to \mathbb{R}$ such that $Z = f(Y) = \sum_{y \in E} f(y)\mathbf{1}_{Y=y}$. Then

$$\begin{aligned}
\mathbb{E}(XZ) = \mathbb{E}(\sum_{y \in E} f(y) X \mathbf{1}_{Y=y}) &= \sum_{y \in E} f(y)\mathbb{E}(X\mathbf{1}_{Y=y}) \\
&= \sum_{y \in E} f(y)\mathbb{E}(X|Y=y)\mathbb{P}(Y=y) \\
&= \sum_{y \in E} f(y)u(y)\mathbb{P}(Y=y) \\
&= \mathbb{E}(u(Y)f(Y)) = \mathbb{E}(u(Y)Z).
\end{aligned}$$

Note that, if E is countable, we can use the dominated convergence theorem to permute sum and expectation in the second equality. As $\mathbb{E}(XZ) = \mathbb{E}(u(Y)Z)$ for any random variable Z which is bounded and measurable with respect to $\sigma(Y)$, we conclude that $u(Y) = \mathbb{E}(X|Y)$ by the definition of conditional expectation. $\square$

1.3 Random Variables with a Density

Another case where the calculation of the conditional expectation $\mathbb{E}(X|Y)$ is particularly simple is the case where the pair (X, Y) has a density. Consider a pair of random variables (X, Y) with joint density $p_{X,Y}(x, y)$, and marginal densities $p_X(x)$ and $p_Y(y)$. We assume that $X \in L^1$. For all $y \in \mathbb{R}$ such that $p_Y(y) > 0$, we define the conditional density of X given $Y = y$ by

$$p_{X|Y=y}(x) \;=\; \frac{p_{X,Y}(x, y)}{p_Y(y)},$$

as well as the conditional expectation of X given $Y = y$ by

$$\mathbb{E}(X|Y = y) \;=\; \int_{\mathbb{R}} x p_{X|Y=y}(x)\mathrm{d}x.$$

Note that it differs from the conditional expectation given the event $B = \{Y = y\}$ as defined in Section 1.2 since $\mathbb{P}(Y = y) = 0$.

We can now provide an explicit formula for the conditional expectation $\mathbb{E}(X|Y)$ in this setting.

Proposition 2. *It holds*

$$\mathbb{E}(X|Y) \;=\; u(Y).$$

where

$$u : y \in \mathbb{R} \mapsto \begin{cases} \mathbb{E}(X|Y = y) & \textit{if } p_Y(y) > 0, \\ 0 & \textit{otherwise.} \end{cases}$$

Remark 8. *The value of u on $N = \{y \in \mathbb{R};\ p_Y(y) = 0\}$ is irrelevant as $\mathbb{P}(Y \in N) = 0$.*

Proof. We need to check that $u(Y)$ satisfies the three points of Theorem 1. Assume first that u is a well-defined Borel map and let us check the integrability of $u(Y)$. Indeed, using Fubini–Tonelli Theorem, we have

$$\begin{aligned}\mathbb{E}(|u(Y)|) = \int_{\mathbb{R}} |u(y)|p_Y(y)\mathrm{d}x &\le \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{p_{X,Y}(x, y)}{p_Y(y)} |x|\mathrm{d}x \right) p_Y(y)\mathrm{d}y \\ &= \int_{\mathbb{R}^2} |x|p_{X,Y}(x, y)\mathrm{d}x\mathrm{d}y = \mathbb{E}[|X|] < +\infty.\end{aligned}$$

In addition, the above computation also induces that

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{p_{X,Y}(x, y)}{p_Y(y)} |x|\mathrm{d}x \right) p_Y(y)\mathrm{d}y < +\infty,$$

and thus $u : \mathbb{R} \to \mathbb{R}$ is indeed a well-defined Borel map according to Fubini Theorem. It follows that $u(Y)$ is measurable with respect to $\sigma(Y)$.

Then let Z be a bounded random variable which is measurable with respect to $\sigma(Y)$. According to Lemma 1, there exists a Borelian function f such that $Z = f(Y)$. It holds

$$\begin{aligned}
\mathbb{E}(XZ) \;=\; \mathbb{E}(Xf(Y)) \;&=\; \int_{\mathbb{R}^2} xf(y)p_{X,Y}(x,y)\mathrm{d}x\mathrm{d}y \\
&=\; \int_{\mathbb{R}} \left(\int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \mathrm{d}x \right) f(y)p_Y(y)\mathrm{d}y \\
&=\; \int_{\mathbb{R}} u(y)f(y)p_Y(y)\mathrm{d}y \\
&=\; \mathbb{E}(u(Y)f(Y)) \;=\; \mathbb{E}(u(Y)Z).
\end{aligned}$$

In this computation, we used the equality

$$p_{X,Y}(x,y) \;=\; \frac{p_{X,Y}(x,y)}{p_Y(y)} p_Y(y),$$

valid with the convention $p_{X,Y}(x,y)/p_Y(y) = 0$ if $p_Y(y) = 0$. □

1.4 Random Variables in $L^2(\Omega)$

Let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra of $\mathcal{F}$. When X is a square integrable random variable, the conditional expectation $E(X|\mathcal{G})$ has a nice interpretation: it corresponds to the orthogonal projection onto the subspace of $\mathcal{G}$–measurable random variables. Recall that L^2 equipped with the inner product $\langle X, Y \rangle := \mathbb{E}(XY)$ is an Hilbert space.

Proposition 3. *If $X \in L^2$, then $\mathbb{E}(X|\mathcal{G})$ is the orthogonal projection of X onto the space $L^2(\mathcal{G})$ of square integrable random variables measurable with respect to $\mathcal{G}$. In particular,*

$$\mathbb{E}(X|\mathcal{G}) \;=\; \underset{Y \in L^2(\mathcal{G})}{\arg\min}\, \mathbb{E}\left[(X-Y)^2\right].$$

Proof. Let $\hat{X}$ be the orthogonal projection of X onto $L^2(\mathcal{G})$. Note that it is well-defined since $L^2(\mathcal{G})$ is a closed linear subspace of L^2. Since $\hat{X} \in L^2(\mathcal{G})$ and $L^2 \subset L^1$, $\hat{X}$ is integrable and $\mathcal{G}-$measurable. Let us now observe that $X - \hat{X} \perp Z$ for any random variable $Z \in L^2(\mathcal{G})$, and so

$$\mathbb{E}(XZ) \;= E(\hat{X}Z) + \mathbb{E}((X-\hat{X})Z) \;=\; \mathbb{E}(\hat{X}Z).$$

It follows that, for any random variable Z $\mathcal{G}-$measurable and bounded (and therefore in $L^2(\mathcal{G})$), $\mathbb{E}(XZ) = \mathbb{E}(\hat{X}Z)$. Thus $\hat{X}$ satisfies the three properties of Theorem 1 and so $\hat{X} = \mathbb{E}(X|\mathcal{G})$ by uniqueness. □

1.5 Random variables in $L^1(\Omega)$

We now proceed to the proof of Theorem 1 establishing existence and uniqueness of the conditional expectation $\mathbb{E}(X|\mathcal{G})$ when Y is only integrable. We recall that, as $\mathbb{P}$ is a probability measure, $L^2 \subset L^1$, and that the notion of orthogonal projection is not defined in L^1.

Let us provide first a preliminary result, establishing the monotonicity of conditional expectation.

Lemma 2. *Given X_1 and X_2 in L^1, let Y_1 and Y_2 satisfy the three points of Theorem 1 for X_1 and X_2 respectively. If $X_1 \leq X_2$ a.s., then $Y_1 \leq Y_2$ a.s..*

Proof. Consider the variable $Z = \mathbf{1}_{(Y_1 - Y_2) \geq 0}$ and observe that it is measurable with respect to $\mathcal{G}$. It holds

$$0 \;\geq\; \mathbb{E}((X_1 - X_2)Z) \;=\; \mathbb{E}((Y_1 - Y_2)Z) \;\geq\; 0,$$

where the equality follows from point (iii) of Theorem 1. Therefore $\mathbb{E}((Y_1 - Y_2)Z) = 0$ and thus $(Y_1 - Y_2)Z = 0$ a.s. since this variable is non-negative. This implies $Y_1 \leq Y_2$ a.s.. $\square$

We also recall the monotone convergence theorem: if $(X_n)_{n\geq 1}$ is an nondecreasing sequence of nonnegative random variables, then

$$\lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X),$$

where $X = \lim_{n\to+\infty} X_n$ denotes the a.s. limit (possibly being infinite).

Proof of Theorem 1. Uniqueness: Let Y_1 and Y_2 be two random variables satisfying the assertions of Theorem 1. By Lemma 2, we have $Y_1 \leq Y_2$ and $Y_2 \leq Y_1$ a.s.. Therefore $Y_1 = Y_2$ a.s..

Existence: First suppose that $X \geq 0$ a.s.. Let $(X_n)_{n\geq 1}$ be the sequence defined by

$$X_n = X \wedge n, \quad n \geq 1.$$

The variables X_n are bounded and therefore in L^2, and thus we can define $Y_n = \mathbb{E}(X_n|\mathcal{G})$ as the orthogonal projection of X_n onto $L^2(\mathcal{G})$. Since the sequence (X_n) is nonnegative and nondecreasing, it follows by Lemma 2 that the sequence (Y_n) is also nonnegative and nondecreasing. Let us denote by Y the a.s. limit of the sequence (Y_n).

Let us show that Y has the three desired properties. First, Y is measurable with respect to $\mathcal{G}$ as the a.s. limit of $\mathcal{G}$–measurable random variables. Then $Y \in L^1$ since it holds by monotone convergence theorem

$$\mathbb{E}(Y) = \lim_{n\to+\infty} \mathbb{E}(Y_n) = \lim_{n\to+\infty} \mathbb{E}(X_n) = \mathbb{E}(X) < +\infty,$$

where the second equality follows from the fact that $\mathbb{E}(Y_n) = \mathbb{E}(X_n)$ by property of the orthogonal projection. Finally, for any Z bounded and measurable with respect to $\mathcal{G}$, we have

$$\begin{aligned}\mathbb{E}(XZ) &= \mathbb{E}((X - X_n)Z) + \mathbb{E}(X_n Z) \ = \ \mathbb{E}((X - X_n)Z) + \mathbb{E}(Y_n Z) \\ &= \mathbb{E}((X - X_n)Z) + \mathbb{E}((Y_n - Y)Z) + \mathbb{E}(YZ),\end{aligned}$$

and the first two terms on the right-hand side tend to 0 by dominated convergence theorem Therefore, $\mathbb{E}(XZ) = \mathbb{E}(YZ)$.

In the general case, we decompose as usual

$$X = X_+ - X_-,$$

where $X_+ = \max(X, 0)$ and $X_- = -\min(X, 0)$. Each of these terms is in L^1 and nonnegative. We can apply the first part of the proof to obtain Y_+ and Y_- the conditional expectations of X_+ and X_- respectively. It remains to check that $Y = Y_+ - Y_-$ has the desired properties to be the conditional expectation of X. □

1.6 Basic Properties

Let us provide the basic and essential properties of conditional expectation. The proofs are left as an exercise.

We start by establishing elementary properties which are shared with classical expectation. Note that positivity and triangular inequality come from the monotonicity established in Lemma 2.

Proposition 4. *Let $X, Y \in L^1$ and $\lambda \in \mathbb{R}$.*

1. $\mathbb{E}(X + \lambda Y|\mathcal{G}) = \mathbb{E}(X|\mathcal{G}) + \lambda\mathbb{E}(Y|\mathcal{G})$. *(linearity)*
2. $\mathbb{E}(1|\mathcal{G}) = 1$.
3. *If $X \geq 0$ a.s., then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s. (positivity)*
4. $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X|\,|\,\mathcal{G})$ *a.s. (triangular inequality)*

The conditional expectation also shares the Jensen inequality with the classical expectation. The proof is (almost) identical.

Proposition 5 (Jensen Inequality)**.** *Let $X \in L^1$ and $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex map such that $\varphi(X) \in L^1$. Then it holds*

$$\varphi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\varphi(X)|\mathcal{G}).$$

In particular, $\|\mathbb{E}(X|\mathcal{G})\|_p \leq \|X\|_p$ for any $p \geq 1$.

Next we establish basic properties which are specific to the conditional expectation.

Proposition 6. *Let $X \in L^1$.*

1. *If X is measurable with respect to $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = X$. Further, if $Y \in L^1$ such that $XY \in L^1$, then $\mathbb{E}(XY|\mathcal{G}) = X\mathbb{E}(Y|\mathcal{G})$.*

2. *If X is independent of $\mathcal{G}$, then $\mathbb{E}(X|\mathcal{G}) = \mathbb{E}(X)$.*

3. *$\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(X)$.*

We now give the statement of the celebrated tower property which ensures that only the conditioning w.r.t. the smallest σ-algebra matters.

Proposition 7 (Tower Property)**.** *If $\mathcal{G}, \mathcal{G}'$ are two σ-algebra such that $\mathcal{G}' \subset \mathcal{G}$, then it holds*

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G}')|\mathcal{G}) \;=\; \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{G}') \;=\; \mathbb{E}(X|\mathcal{G}').$$

Let us turn now to the limit theorems for the conditional expectation. They are very similar to those for the classical expectation.

Theorem 2. *Let (X_n) be a sequence of random variables.*

1. *If $0 \le X_n \nearrow X$, then $\mathbb{E}(X_n|\mathcal{G}) \nearrow \mathbb{E}(X|\mathcal{G})$. (monotone convergence)*

2. *If $X_n \ge 0$, then $\mathbb{E}(\liminf_n X_n|\mathcal{G}) \le \liminf_n E(X_n|\mathcal{G})$. (Fatou)*

3. *If $X_n \to X$ a.s. and $|X_n| \le Z \in L^1$, then $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ a.s. and in L^1. (dominated convergence)*

4. *If $X_n \to X$ in L^p with $p \in [1, +\infty]$, then $\mathbb{E}(X_n|\mathcal{G}) \to \mathbb{E}(X|\mathcal{G})$ in L^p. (convergence in L^p)*

Finally, let us discuss a case where the computation of the conditional expectation boils down to the computation of an expectation.

Proposition 8. *Let X be $\mathcal{G}$-measurable and Y be independent of $\mathcal{G}$. Let also $\varphi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be bounded Borel-measurable. Then*

$$\mathbb{E}(\varphi(X,Y)|\mathcal{G}) = \Phi(X) \qquad \text{where } \Phi(x) = \mathbb{E}(\varphi(x,Y))\ \forall\, x \in \mathbb{R}.$$

Proof. Note first that $\Phi(X)$ is measurable with respect to $\mathcal{G}$ and bounded. Next, given Z $\mathcal{G}-$measurable and bounded, observe that the distribution $\mathbb{P}_{X,Y,Z}$ can be written as $\mathbb{P}_{X,Y,Z} = \mathbb{P}_Y \otimes \mathbb{P}_{X,Z}$ since Y is independent of (X,Z). It follows by Fubini that

$$\begin{aligned}\mathbb{E}(\varphi(X,Y)Z) &= \int_{\mathbb{R}^3} \varphi(x,y) z d\mathbb{P}_{X,Y,Z}(x,y,z) = \int_{\mathbb{R}^2} \Big(\int_{\mathbb{R}} \varphi(x,y) d\mathbb{P}_Y(y) \Big) z d\mathbb{P}_{X,Z}(x,z) \\ &= \int_{\mathbb{R}^2} \mathbb{E}(\varphi(x,Y)) z d\mathbb{P}_{X,Z}(x,z) = \int_{\mathbb{R}^2} \Phi(x) z d\mathbb{P}_{X,Z}(x,z) = \mathbb{E}(\Phi(X)Z).\end{aligned}$$

So $\mathbb{E}(\varphi(X,Y)|\mathcal{G}) = \Phi(X)$. □

1.7 Gaussian Random Variables

If $(X_0, X_1, \ldots, X_n)$ is a Gaussian vector, then $\mathbb{E}(X_0|X_1, \ldots, X_n)$ is a linear combination of $X_1, \ldots, X_n$.

Proposition 9. *We assume that $(X_0, X_1, \ldots, X_n)$ is a Gaussian vector. Then there exist real numbers $c_0, c_1 \ldots, c_n$ such that*

$$\mathbb{E}(X_0|X_1, \ldots, X_n) = c_0 + \sum_{i=1}^{n} c_i X_i.$$

This result is left as an exercise.

2 Martingales

2.1 Definition and Examples

First definitions. A *stochastic process* is a sequence $(X_n)_{n\geq 0}$ of random variables. A sequence $(\mathcal{F}_n)_{n\geq 0}$ of σ-algebras is called a *filtration* if it is nondecreasing with respect to the inclusion:

$$\mathcal{F}_n \subset \mathcal{F}_{n+1} \qquad \forall n \geq 0.$$

We say that the process (X_n) is *adapted* to the filtration $(\mathcal{F}_n)$ if X_n is measurable with respect to $\mathcal{F}_n$ for all $n \geq 0$.

Interpretation. Typically, the index n corresponds to the time and $\mathcal{F}_n$ to the information available at time n. The monotonicity of filtration means that the information accumulates over time. A process is adapted when its value X_n is known at time n.

Example. Given a process $(X_n)_{n\geq 0}$, the sequence $(\mathcal{F}_n)_{n\geq 0}$ defined by

$$\mathcal{F}_n \;=\; \sigma(X_0, \ldots, X_n),$$

is a filtration, where we recall that $\sigma(X_0, \ldots, X_n)$ is the smallest σ-algebra making the variables $X_0, \ldots, X_n$ measurable. It is called the *natural filtration* of the process (X_n). Observe that a process is always adapted to its natural filtration.

Definition of martingale. We say that a process (X_n) is a *martingale* with respect to a filtration $(\mathcal{F}_n)$ if, for all $n \geq 0$,

1. X_n is measurable with respect to $\mathcal{F}_n$,
2. X_n is integrable,
3. $\mathbb{E}(X_{n+1}|\mathcal{F}_n) = X_n$.

We say that (X_n) is a *submartingale* (resp. *supermartingale*) if condition 3 above is replaced by $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n$ (resp. $\mathbb{E}(X_{n+1}|\mathcal{F}_n) \leq X_n$).

The first condition also reads: (X_n) is adapted to $(\mathcal{F}_n)$.

Interpretation. A martingale describes the evolution of a player's total gain in a fair game: At each time n, a round of the game occurs and brings the player a marginal gain $\Delta X_n = X_n - X_{n-1}$, the game is fair in the sense that $\mathbb{E}(\Delta X_n|\mathcal{F}_n) = 0$.

Remark 9. *We simply say that (X_n) is a martingale if (X_n) is a martingale with respect to a certain filtration. In this case, (X_n) is always a martingale with respect to its natural filtration (see Exercises).*

Let us give two elementary properties which follow directly from the definition.

Proposition 10. *Let $(X_n)_{n\geq 0}$ be a martingale with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Then it holds for all $n \geq 0$,*

1. $\mathbb{E}(X_n|\mathcal{F}_k) = X_k$ *for* $0 \leq k \leq n$,
2. $\mathbb{E}(X_n) = \mathbb{E}(X_0)$.

Proof. For the first point, we fix $k \geq 0$ and we proceed by induction on $n \geq k$. The result is true for $n = k$ since X_k is measurable with respect to $\mathcal{F}_k$. Let us then assume it to be true for $n \geq k$ and prove it for $n+1$: by the tower property, the property of martingale and the induction hypothesis,

$$\mathbb{E}(X_{n+1}|\mathcal{F}_k) \;=\; \mathbb{E}(\mathbb{E}(X_{n+1}|\mathcal{F}_n)|\mathcal{F}_k) \;=\; \mathbb{E}(X_n|\mathcal{F}_k) \;=\; X_k.$$

For the second point, we proceed as follows:

$$\mathbb{E}(X_n) \;=\; \mathbb{E}(\mathbb{E}(X_n|\mathcal{F}_0)) \;=\; \mathbb{E}(X_0),$$

where the last equality follows from the first point with $k = 0$. □

Proposition 11. *If $(X_n)_{n\geq 0}$ is a martingale and $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function such that $\varphi(X_n) \in L^1$ for all $n \geq 0$, then $(\varphi(X_n))_{n\geq 0}$ is a submartingale.*

Proof. It is an easy consequence of Jensen's inequality. Indeed it holds

$$\mathbb{E}(\varphi(X_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(X_{n+1}|\mathcal{F}_n)) = \varphi(X_n).$$

□

Example: random walk on $\mathbb{Z}$. Let $(B_n)_{n\geq 1}$ be a sequence of i.i.d. random variable such that $\mathbb{P}(B_n = 1) = \mathbb{P}(B_n = -1) = \frac{1}{2}$. The process $(X_n)_{n\geq 0}$ defined by

$$\begin{cases} X_0 \;=\; 0, \\ X_{n+1} \;=\; X_n + B_{n+1}, \quad n \geq 0, \end{cases}$$

is a martingale. Note that

$$X_n \;=\; \sum_{k=1}^{n} B_k, \quad n \geq 1.$$

This example generalizes as follows: Given $(\Delta_n)_{n\geq 1}$ is a sequence of integrable i.i.d. random variables and $x \in \mathbb{R}$, the process $(X_n)_{n\geq 0}$ defined by $X_0 = x$ and

$$X_n \;=\; x + \sum_{k=1}^{n} \Delta_k - n\mathbb{E}(\Delta_1), \quad n \geq 1,$$

is a martingale, sometimes called a *generalized random walk*.

Example: multiplicative martingale. Let $(Y_n)_{n\geq 1}$ be a sequence of i.i.d. integrable random variables such that $\mathbb{E}(Y_n) = 1$ for all $n \geq 1$. The process $(X_n)_{n\geq 0}$ defined by

$$\begin{cases} X_0 \;=\; 1, \\ X_{n+1} \;=\; X_n Y_{n+1}, \quad n \geq 0, \end{cases}$$

is a martingale.

Example: closed martingale. Let Z be an integrable random variable and $(\mathcal{F}_n)_{n\geq 0}$ a filtration. Then the process $(X_n)_{n\geq 0}$ defined by

$$X_n = \mathbb{E}(Z|\mathcal{F}_n), \quad n \geq 0,$$

is a martingale. Indeed, for all $n \geq 0$, X_n is $\mathcal{F}_n$–measurable and integrable by definition of the conditional expectation. Further, it follows from the tower property that

$$\mathbb{E}(X_{n+1}|\mathcal{F}_n) \;=\; \mathbb{E}(\mathbb{E}(Z|\mathcal{F}_{n+1})|\mathcal{F}_n) \;=\; \mathbb{E}(Z|\mathcal{F}_n) \;=\; X_n.$$

2.2 Stopping Times

Definition of stopping time. We say that a random variable $T : \Omega \mapsto \mathbb{N} \cup \{+\infty\}$ is a *stopping time* with respect to a filtration $(\mathcal{F}_n)$ if

$$\{T = n\} \in \mathcal{F}_n, \quad \text{for all } n \in \mathbb{N}.$$

We say that T is *finite* if $\mathbb{P}(T < +\infty) = 1$.

The condition $\{T = n\} \in \mathcal{F}_n$ for all $n \geq 0$ can be replaced by the equivalent condition

$$\{T \leq n\} \in \mathcal{F}_n \quad \text{for all } n \geq 0.$$

Indeed, on the one hand, we have that

$$\{T \leq n\} = \{T = 0\} \cup \cdots \cup \{T = n\},$$

and $\{T = k\} \in \mathcal{F}_k \subset \mathcal{F}_n$ for all $0 \leq k \leq n$. On the other hand, it holds

$$\{T = n\} = \{T \leq n\} \cap \{T > n-1\} = \{T \leq n\} \cap \{T \leq n-1\}^{\mathsf{c}},$$

where $\{T \leq n\} \in \mathcal{F}_n$ and $\{T \leq n-1\}^{\mathsf{c}} \in \mathcal{F}_{n-1} \subset \mathcal{F}_n$.

Remark 10. *Do not confuse the notion of finite and bounded random variables: T is said to be bounded if there exists a number $M > 0$ such that $\mathbb{P}(T \leq M) = 1$. By definition, all random variables taking values in $\mathbb{N}$ are finite but they are not necessarily bounded. For instance, a Poisson variable is not bounded but a binomial variable is.*

Interpretation. A stopping time corresponds to a legitimate stopping strategy for the game: The decision to quit the game or not should be based only upon the information available at time n. For instance,

- The player leaves as soon as her total gain is larger or equal to 10:

$$T \ = \ \min\{n \geq 1 : X_n \geq 10\}.$$

  The strategy is legitimate and T is a stopping time since

$$\{T \leq n\} \ = \ \{X_1 \geq 10\} \cup \cdots \cup \{X_n \geq 10\}, \quad n \geq 1.$$

- The player leaves the game just before losing for the first time:

$$T \ = \ \min\{n \geq 0 : \Delta X_{n+1} < 0\}.$$

  The strategy is akin to cheating and T is not a stopping time.

Example: constant time. Let $n \in \mathbb{N}$, the constant time $T = n$ is a stopping time with respect to any filtration.

Example: hitting time of B. Let $B \subset \mathbb{R}$ be a Borel set and denote

$$T_B \ = \ \min\{n \geq 0 : X_n \in B\},$$

with the convention $\min \emptyset = +\infty$. The variable T_B is a stopping time with respect to the natural filtration of (X_n). Indeed, for all $n \geq 0$, we have

$$\{T_B \leq n\} \ = \ \{X_0 \in B\} \cup \cdots \cup \{X_n \in B\}$$

and $\{X_k \in B\} \in \mathcal{F}_k \subset \mathcal{F}_n$ for all $0 \leq k \leq n$.

2.3 Martingale Transform and Stopped Process

We introduce a transformation of a martingale which preserves the martingale property. We say that $(C_n)_{n\geq 1}$ is a *predictable* process when C_n is measurable with respect to $\mathcal{F}_{n-1}$ for all $n \geq 1$.

Proposition 12. *Let $(X_n)_{n\geq 0}$ be a martingale and $(C_n)_{n\geq 1}$ a predictable process such that $C_n \in L^\infty$ for all $n \geq 1$. Let also $Y_0 \in L^1$ be $\mathcal{F}_0$–measurable. Then the process $(Y_n)_{n\geq 0}$ defined by for all $n \geq 1$,*

$$Y_n \ = \ Y_0 + \sum_{k=1}^{n} C_k \Delta X_k, \quad \textit{with } \Delta X_k = X_k - X_{k-1},$$

is a martingale.

Proof. It is easy to check that Y_n is integrable and measurable with respect to $\mathcal{F}_n$ for all $n \geq 0$. Moreover, it holds for all $n \geq 1$

$$\mathbb{E}(Y_n|\mathcal{F}_{n-1}) \;=\; \mathbb{E}(Y_{n-1} + C_n \Delta X_n|\mathcal{F}_{n-1}) \;=\; Y_{n-1} + C_n \mathbb{E}(\Delta X_n|\mathcal{F}_{n-1}),$$

since Y_{n-1} and C_n are measurable with respect to $\mathcal{F}_{n-1}$. In addition, we have

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

and so $\mathbb{E}(Y_n|\mathcal{F}_{n-1}) = Y_{n-1}$. □

Example: modified random walk. Let $y \in \mathbb{Z}$ and $(B_n)_{n\geq 1}$ a sequence of i.i.d. random variables such that $\mathbb{P}(B_n = -1) = \mathbb{P}(B_n = 1) = 1/2$. Let also $D : \mathbb{Z} \to \mathbb{N}$ be a bounded function. We consider the process $(Y_n)_{n\geq 0}$ defined by

$$\begin{cases} Y_0 \;=\; y, \\ Y_{n+1} \;=\; Y_n + D(Y_n)B_{n+1}. \end{cases}$$

For $n \geq 1$, we have

$$Y_n \;=\; y + \sum_{k=1}^{n} D(Y_{k-1})B_k \;=\; y + \sum_{k=1}^{n} D(Y_{k-1})\Delta X_k,$$

where (X_n) is a symmetric random walk. Therefore the process (Y_n) is a martingale.

Example: stopped process. Given a process $(X_n)_{n\geq 0}$ and a random time $T : \Omega \to \mathbb{N} \cup \{+\infty\}$, we define the *stopped process* $(X_{n\wedge T})_{n\geq 0}$ by

$$X_{n\wedge T} \;=\; \begin{cases} X_n & \text{if } n < T, \\ X_T & \text{otherwise.} \end{cases}$$

For example, when T is the hitting time of a Borel set B, $X_{n\wedge T}$ is equal to X_n until the process attains B, and then it remains constant, equal to X_T the value of the process when it attains B.

If (X_n) is a martingale and T is a stopping time with respect to the same filtration $(\mathcal{F}_n)$, then $(X_{n\wedge T})$ is still a martingale. Indeed, for $n \geq 1$, it holds

$$X_{n\wedge T} \;=\; X_0 + \sum_{k=1}^{n\wedge T} \Delta X_k \;=\; X_0 + \sum_{k=1}^{n} \mathbf{1}_{\{T\geq k\}} \Delta X_k, \quad n \geq 1,$$

and $\{T \geq k\} = \{T \leq k-1\}^c \in \mathcal{F}_{k-1}$ for all $k \geq 1$.

2.4 Optional Stopping Theorem

Let $(X_n)_{n\geq 0}$ be a martingale and T a finite stopping time with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$. It follows from the previous section that the stopped process $(X_{n\wedge T})$ is a martingale and so we have that

$$\mathbb{E}(X_{n\wedge T}) \;=\; \mathbb{E}(X_{0\wedge T}) \;=\; \mathbb{E}(X_0).$$

The question is whether we can pass to the limit to prove that

$$\mathbb{E}(X_T) \;=\; \mathbb{E}(X_0)\ ?$$

The answer is no in general. For instance, consider a symmetric random walk (X_n) together with the stopping time

$$T \;=\; \min\{n \geq 0 : X_n = 1\}.$$

Then we can show that T is finite (see below for a proof). Nonetheless, since $X_T = 1$ by definition, it holds

$$1 \;=\; \mathbb{E}(X_T) \;\neq\; \mathbb{E}(X_0) = 0.$$

The celebrated *optional stopping theorem* provides conditions to guarantee that the equality holds.

Theorem 3 (Doob's optional sampling theorem)**.** *Let $(X_n)_{n\geq 0}$ be a martingale and T a stopping time with respect to a filtration $(\mathcal{F}_n)_{n\geq 0}$. Assume that one of the following conditions holds:*

1. *T is bounded*
2. *$\mathbb{E}(T) < +\infty$ and $(\Delta X_n)_{n\geq 1}$ is bounded in L^∞.*
3. *T is finite and $(X_{n\wedge T})_{n\geq 0}$ is bounded in L^∞.*

Then $X_T \in L^1$ and

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

Proof. Let us consider the three cases separately.

1. Suppose that T is bounded, *i.e.*, there exists an integer N (deterministic) such that $\mathbb{P}(T \leq N) = 1$. In this case, the result follows directly from the fact that the stopped process $(X_{n\wedge T})$ is a martingale:

$$\mathbb{E}(X_T) \;=\; \mathbb{E}(X_{N\wedge T}) \;=\; \mathbb{E}(X_{0\wedge T}) \;=\; \mathbb{E}(X_0).$$

2. Here the proof relies on the dominated convergence theorem. First we observe that, since T is integrable, it is also finite and so $X_{n\wedge T}$ converges a.s. to X_T. Then, by assumption, there exists a number $M > 0$ such that $|\Delta X_k| \leq M$ a.s. for all $k \geq 1$ and therefore

$$|X_{n\wedge T}| \;=\; \left|X_0 + \sum_{k=1}^{n\wedge T} \Delta X_k\right| \;\leq\; |X_0| + M(n\wedge T) \;\leq\; |X_0| + MT$$

By assumption, $\mathbb{E}(|X_0| + MT) < +\infty$ and therefore the dominated convergence theorem ensures that $X_T \in L^1$ and

$$\mathbb{E}(X_T) \;=\; \lim_{n\to\infty} \mathbb{E}(X_{n\wedge T}) \;=\; \mathbb{E}(X_0).$$

3. We proceed again by dominated convergence theorem, this time using the domination $|X_{n\wedge T}| \leq M$ a.s., valid for a certain number $M > 0$ and for all $n \geq 0$. □

Remark 11. *The list of conditions given in Theorem 3 is by no means exhaustive: If T is finite and we can find $Z \in L^1$ such that $|X_{n\wedge T}| \leq Z$, then the dominated convergence theorem ensures that $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

Example: random walk. This theorem allows us to derive several interesting facts about the random walk $(X_n)_{n\geq 0}$. Let $a \leq -1$ and $b \geq 1$ be two integers and define the stopping time

$$T \;=\; \min\{n \geq 0 : X_n = a \text{ or } X_n = b\},$$

which represents the exit time from the interval $]a, b[$.

Let us start by showing that T is finite. Denote $\ell = b - a - 1$ the maximum distance that the random walk must travel to exit the interval $]a, b[$. Observe that, if $T > m\ell$ for a certain integer $m \geq 1$, then $|X_{j+\ell} - X_j| < \ell$ for all $0 \leq j \leq (m-1)\ell$. So in particular, it holds for all $m \geq 1$,

$$\mathbb{P}(T > m\ell) \;\leq\; \mathbb{P}\left(|X_{(k+1)\ell} - X_{k\ell}| < \ell, 0 \leq k \leq m-1\right).$$

The right-hand side corresponds to the probability of an intersection of m independent events, each of probability $1 - p$ where

$$\mathbb{P}(|X_\ell - X_0| = \ell) \;=\; 2\left(\frac{1}{2}\right)^\ell = \frac{1}{2^{\ell-1}} =: p > 0,$$

and therefore

$$\mathbb{P}(T > m\ell) \;\leq\; (1-p)^m.$$

We deduce that T is finite since

$$\mathbb{P}(T = +\infty) = \lim_{m\to\infty} \mathbb{P}(T > m\ell) = 0.$$

Let us note that this computation also implies that $\mathbb{E}(T) < +\infty$ by using the formula $\mathbb{E}(T) = \sum_{n\geq 0} \mathbb{P}(T > n) \leq \sum_{m\geq 0} \ell\, \mathbb{P}(T > m\ell)$.

We can now compute the probability that the random walk leaves this interval at a or at b, *i.e.*, $\mathbb{P}(X_T = a)$ and $\mathbb{P}(X_T = b)$. As the stopped process $(X_{n\wedge T})_{n\geq 0}$ is bounded in L^∞, we can apply the optional stopping theorem and conclude that $\mathbb{E}(X_T) = \mathbb{E}(X_0) = 0$. We obtain

$$\begin{cases} \mathbb{P}(X_T = a)a + \mathbb{P}(X_T = b)b \;=\; 0 \\ \mathbb{P}(X_T = a) + \mathbb{P}(X_T = b) \;=\; 1 \end{cases}$$

and thus

$$\mathbb{P}(X_T = a) \ = \ \frac{b}{b-a}, \qquad \mathbb{P}(X_T = b) \ = \ \frac{-a}{b-a}.$$

We can also compute $\mathbb{E}(T)$. We start by noticing that the process $(Y_n)_{n\geq 0}$, defined by $Y_n = X_n^2 - n$, is a martingale. Even if none of the three conditions of the optional stopping theorem applies, we can still prove that $\mathbb{E}(Y_T) = \mathbb{E}(Y_0)$. Indeed, we observe that

$$|Y_{n\wedge T}| \ \leq \ (\max(|a|, b))^2 + n \wedge T \ \leq \ (\max(|a|, b))^2 + T.$$

Since $\mathbb{E}(T) < +\infty$ and $Y_{n\wedge T} \to Y_T$ a.s., we can apply the dominated convergence theorem to obtain

$$0 = \mathbb{E}(Y_{n\wedge T}) \to \mathbb{E}(Y_T) = \mathbb{E}(X_T^2 - T).$$

We conclude that

$$\mathbb{E}(T) \ = \ \mathbb{P}(X_T = a)a^2 + \mathbb{P}(X_T = b)b^2 \ = \ |ab|.$$

Finally, let us consider the stopping time

$$T_1 \ = \ \min\{n \geq 0 : X_n = 1\},$$

and show that T_1 is finite as claimed at the beginning of this section. Taking $b = 1$ and $a \leq -1$ in the definition of T above, we have that

$$\mathbb{P}(T_1 < +\infty) \ \geq \ \mathbb{P}(X_T = 1) = \frac{-a}{1-a}, \qquad \text{for all } a \leq -1.$$

Letting $a \to -\infty$, we deduce that $\mathbb{P}(T_1 < +\infty) = 1$. We can also conclude that $\mathbb{E}(T_1) = +\infty$ because otherwise the optional stopping theorem would imply that $1 = \mathbb{E}(X_{T_1}) = E(X_0) = 0$, which is a contradiction.

2.5 Maximal Inequalities

Let (X_n) be a symmetric random walk on $\mathbb{Z}$. We can explicitly compute the distribution of X_n at a fixed time n, and conclude that X_n is "of order $\sqrt{n}$": for instance,

$$\mathbb{E}(X_n^2) = n \quad \text{and} \quad \frac{X_n}{\sqrt{n}} \xrightarrow[n\to+\infty]{\text{law}} \mathcal{N}(0,1).$$

Here we provide results which allow us to estimate $\max_{0\leq k\leq n} |X_k|$, the radius of the smallest interval centered at 0 in which the random walk can be enclosed until time n. We will see that $\max_{0\leq k\leq n} |X_k|$ is also "of order $\sqrt{n}$".

The main result is an improvement of the Markov inequality for nonnegative submartingales, called *Doob's maximal inequality.*

Theorem 4. *Let $(Y_n)_{n\geq 0}$ be a nonnegative submartingale and $a > 0$. Then it holds for all $n \geq 0$,*

$$\mathbb{P}\left(\max_{0\leq k\leq n} Y_k \geq a\right) \ \leq \ \frac{\mathbb{E}(Y_n)}{a}.$$

In particular, this inequality holds for $Y_n = |X_n|$ where (X_n) is a martingale.

Proof. Denote by T the hitting time of $(a, +\infty)$, *i.e.*,

$$T \;=\; \min\{n \geq 0, X_n \geq a\}.$$

We start by using Markov inequality to obtain that

$$\mathbb{P}\left(\max_{0\leq k\leq n} Y_k \geq a\right) \;=\; \mathbb{P}(Y_{n\wedge T} \geq a) \;\leq\; \frac{\mathbb{E}(Y_{n\wedge T})}{a}.$$

It remains to show that $\mathbb{E}(Y_{n\wedge T}) \leq \mathbb{E}(Y_n)$ for all $n \geq 0$. This inequality is obvious for $n = 0$. Then we can compute for $n \geq 1$,

$$\begin{aligned}
\mathbb{E}(Y_n - Y_{n\wedge T}) \;&=\; \mathbb{E}((Y_n - Y_{n\wedge T})\mathbf{1}_{\{T<n\}}) \\
&=\; \sum_{k=0}^{n-1} \mathbb{E}((Y_n - Y_k)\mathbf{1}_{\{T=k\}}) \\
&=\; \sum_{k=0}^{n-1} \mathbb{E}(\mathbb{E}(Y_n - Y_k|\mathcal{F}_k)\mathbf{1}_{\{T=k\}}) \\
&=\; \sum_{k=0}^{n-1} \mathbb{E}((\mathbb{E}(Y_n|\mathcal{F}_k) - Y_k)\mathbf{1}_{\{T=k\}}) \;\geq\; 0,
\end{aligned}$$

since $\mathbb{E}(Y_n|\mathcal{F}_k) \geq Y_k$ for $0 \leq k \leq n-1$ as (Y_n) is a submartingale. □

Theorem 4 allows us to control the moments of the maximum of a martingale by the moments of the martingale itself.

Proposition 13. *Let $(Y_n)_{n\geq 0}$ be a nonnegative submartingale and $p > 1$. Then it holds for all $n \geq 0$,*

$$\mathbb{E}\left[\left(\max_{0\leq k\leq n} Y_k\right)^p\right] \;\leq\; \left(\frac{p}{p-1}\right)^p \mathbb{E}\left[Y_n^p\right].$$

In particular, this inequality holds for $Y_n = |X_n|$ where (X_n) is a martingale.

Proof. See Exercises. □

Example: the random walk. Let (X_n) be a symmetric random walk on $\mathbb{Z}$. Since the map $u \mapsto |u|$ is convex and nonnegative, the conditional Jensen inequality implies that $(|X_n|)$ is a nonnegative submartingale as seen in Proposition 11. Therefore, for all $n \geq 1$ and $K > 0$ (but the result is only interesting for $K > 1$), we obtain

$$\mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq K\sqrt{n}\right) \;\leq\; \frac{\mathbb{E}(|X_n|)}{K\sqrt{n}} \;\leq\; \frac{\sqrt{\mathbb{E}(X_n^2)}}{K\sqrt{n}} \;=\; \frac{1}{K}.$$

Let us go further and improve substantially this inequality when K is large. We start by observing that for any $a > 0$,

$$\mathbb{P}\left(\max_{1\leq k\leq n} |X_k| \geq a\right) \leq \mathbb{P}\left(\max_{1\leq k\leq n} X_k \geq a\right) + \mathbb{P}\left(\min_{1\leq k\leq n} X_k \leq -a\right)$$

where, by symmetry of the random walk,

$$\mathbb{P}\left(\min_{1\le k\le n} X_k \le -a\right) = \mathbb{P}\left(\max_{1\le k\le n} (-X_k) \ge a\right) = \mathbb{P}\left(\max_{1\le k\le n} X_k \ge a\right).$$

Furthermore, since the map $u \mapsto \mathrm{e}^{\lambda u}$ is convex and nonnegative for any $\lambda > 0$, $(\mathrm{e}^{\lambda X_n})$ is also a nonnegative submartingale by Proposition 11 and it follows from the maximal inequality that

$$\begin{aligned}\mathbb{P}\left(\max_{1\le k\le n} |X_k| \ge a\right) &\le 2\mathbb{P}\left(\max_{1\le k\le n} X_k \ge a\right)\\ &= 2\mathbb{P}\left(\max_{1\le k\le n} \mathrm{e}^{\lambda X_k} \ge \mathrm{e}^{\lambda a}\right) \le 2\mathrm{e}^{-\lambda a}\mathbb{E}(\mathrm{e}^{\lambda X_n}).\end{aligned}$$

In addition, we have that

$$\mathrm{e}^{-\lambda a}\mathbb{E}(\mathrm{e}^{\lambda X_n}) = \mathrm{e}^{-\lambda a}\cosh(\lambda)^n \le \mathrm{e}^{\frac{\lambda^2}{2}n-\lambda a},$$

where we used the inequality $\cosh(x) \le \mathrm{e}^{x^2/2}$ for all $x \in \mathbb{R}$. The right hand side is minimal for $\lambda = a/n$ and is then equal to $\mathrm{e}^{-a^2/2n}$. Choosing $a = K\sqrt{n}$ for $K > 0$, we conclude that

$$\mathbb{P}\left(\max_{1\le k\le n} |X_k| \ge K\sqrt{n}\right) \le 2\mathrm{e}^{-K^2/2}.$$

2.6 Convergence of Martingales

The goal is the section is to study the convergence of martingales as time increases to infinity. We start by discussing the convergence in L^2.

Theorem 5. *If $(X_n)_{n\ge 0}$ is a martingale bounded in L^2, then X_n converges in L^2.*

Proof. The idea of the proof is to show that (X_n) is a Cauchy sequence in the Hilbert space L^2. First observe that for all $m > n > 0$,

$$X_n = X_0 + \sum_{k=1}^{n} \Delta X_k \quad \text{and} \quad X_m - X_n = \sum_{k=n+1}^{m} \Delta X_k.$$

In addition it holds for all $k \ge 1$,

$$\mathbb{E}(X_0 \Delta X_k) = \mathbb{E}(X_0 \mathbb{E}(\Delta X_k|\mathcal{F}_0)) = 0,$$

and, for all $l > k \ge 1$,

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

Therefore, for all $m > n > 0$,

$$\mathbb{E}(X_n^2) = \mathbb{E}(X_0^2) + \sum_{k=1}^{n} \mathbb{E}(\Delta X_k^2) \text{ and } \mathbb{E}((X_m - X_n)^2) = \sum_{k=n+1}^{m} \mathbb{E}(\Delta X_k^2).$$

Since $\sup_{n\in\mathbb{N}} \mathbb{E}(X_n^2) < +\infty$ by assumption, the first equality implies that the series $\sum_{k\geq 1} \mathbb{E}(\Delta X_k^2)$ converges. The conclusion then follows from the second equality which implies that (X_n) is Cauchy in L^2. □

Remark 12. *More generally, we can show that, for any $p > 1$, a martingale bounded in L^p converges in L^p. The proof relies on the maximal inequality in the form of Proposition 13, see Exercises.*

Examples:

1. If $Z \in L^2$, then the closed martingale $(\mathbb{E}(Z|\mathcal{F}_n))_n$ is bounded in L^2 by $\mathbb{E}(Z^2)$, and therefore it converges in L^2.
2. If (X_n) is a symmetric random walk, then $E(X_n^2) = n$ for all $n \geq 0$, and so (X_n) is not bounded in L^2. The theorem does not apply but this not surprising as we know that (X_n) does not even converge in law.

Let us turn now to the almost sure convergence of martingales. The next theorem provides a simple sufficient condition and is often used in the form of the following corollary. The proof is omitted and we refer to the lecture notes of F. Simenhaus for more details.

Theorem 6. *If $(X_n)_{n\geq 0}$ is a (sub/super)martingale bounded in L^1, then (X_n) converges almost surely to a variable $X \in L^1$.*

Corollary 1. *If $(X_n)_{n\geq 0}$ is a nonnegative supermartingale (resp. nonpositive submartinagle), then it converges almost surely to a variable $X \in L^1$.*

Proof. Let (X_n) be a nonnegative supermartingale. Then it is bounded in L^1 as

$$\mathbb{E}[|X_n|] = \mathbb{E}[X_n] \leq \mathbb{E}[X_0], \quad n \geq 1,$$

and so we can apply Theorem 6. □

Remark 13. *If (X_n) is bounded in L^2, it is also bounded in L^1. Therefore, by combining the two previous results, we conclude that a martingale bounded in L^2 converges in L^2 and almost surely.*

Remark 14. *A martingale bounded in L^1 does not necessarily converge in L^1. See examples below.*

Example: multiplicative martingale. Consider the multiplicative martingale given by $X_0 = 1$ and

$$X_n = Y_1 \cdots Y_n, \quad n \geq 1,$$

with (Y_n) i.i.d. such that $\mathbb{P}(Y_n = 2) = \mathbb{P}(Y_n = 0) = 1/2$. It is clearly nonnegative and, as $\mathbb{E}(X_n) = \mathbb{E}(X_0) = 1$ for all $n \geq 0$, it is also bounded in L^1. Thus (X_n)

converges almost surely to a random variable $X \in L^1$. Actually we can identify $X = 0$ since

$$\mathbb{P}(X > 0) = \mathbb{P}(\forall\, n \in \mathbb{N},\ Y_n > 0) = \lim_{n\to\infty} \frac{1}{2^n} = 0.$$

However (X_n) does not converge to 0 in L^1 since $\mathbb{E}(|X_n - 0|) = \mathbb{E}(X_n) = 1$ for all $n \geq 0$.

Example: branching process. Let $(Y_{n,k})_{n\geq 0, k\geq 1}$ be a family of i.i.d. random variables valued in $\mathbb{N}$ such that

$$0 \ < \ \overline{Y} \ := \ \mathbb{E}(Y_{0,1}) \ < \ +\infty.$$

We consider a process $(X_n)_{n\geq 0}$ defined by

$$\begin{cases} X_0 \ = \ 1, \\ X_{n+1} \ = \ \sum_{k=1}^{X_n} Y_{n,k}, \quad n \geq 0, \end{cases}$$

with the convention $\sum_{k=1}^{0}(\dots) = 0$. Typically X_n represents the size of a population at time n, where each individual k gives birth to a random number $Y_{n,k}$ of children forming the population at time $n+1$.

One can check that the sequence $(Z_n)_{n\geq 0}$, defined by $Z_n = X_n/\overline{Y}^n$ for $n \geq 0$, is a martingale, and so $\mathbb{E}(X_n) = \overline{Y}^n$ for all $n \geq 0$. Therefore the average size of the population decreases exponentially for $\overline{Y} < 1$, remains constant and equal to 1 for $\overline{Y} = 1$, and grows exponentially for $\overline{Y} > 1$.

In the case $\overline{Y} < 1$, we can easily show that (X_n) converges to 0 almost surely. Indeed, we have that for any $\varepsilon > 0$,

$$\mathbb{P}(X_n > \varepsilon) \ \leq \ \frac{\mathbb{E}(X_n)}{\varepsilon} \ = \ \frac{\overline{Y}^n}{\varepsilon},$$

and thus

$$\sum_{n=0}^{\infty} \mathbb{P}(X_n > \varepsilon) \ < \ +\infty,$$

which implies that $X_n \to 0$ a.s. by Borel–Cantelli Lemma.

In the case $\overline{Y} = 1$ and $\mathbb{P}(Y_{0,1} = 1) < 1$, we can still show that (X_n) converges almost surely to 0 but it is more involved. First notice that we necessarily have $\mathbb{P}(Y_{0,1} = 0) =: p > 0$. Next, we observe that (X_n) is a nonnegative martingale and, as $\mathbb{E}(X_n) = 1$ for all $n \geq 0$, it is also bounded in L^1. It follows that X_n converges almost surely to a random variable $X \in L^1$. Then it holds for $n \geq 0$,

$$\begin{aligned} \mathbb{P}(X_{n+1} = 0) \ &= \ \sum_{k=0}^{\infty} \mathbb{P}(X_{n+1} = 0, X_n = k) \ = \ \sum_{k=0}^{\infty} \mathbb{P}(X_n = k) p^k \\ &= \ \mathbb{P}(X_n = 0) + \sum_{k=1}^{\infty} \mathbb{P}(X_n = k) p^k. \end{aligned}$$

Since $\mathbb{P}(X_n = 0)$ converges when $n \to \infty$, we necessarily have $\mathbb{P}(X_n = k) \to 0$ for all $k \geq 1$. Therefore, $\mathbb{P}(X = k) = 0$ for all $k \geq 1$ and, as the support of X is in $\mathbb{N}$, we conclude that $X = 0$.

Remark 15. *The case* $\overline{Y} = 1$ *and* $\mathbb{P}(Y_{0,1} = 1) < 1$ *provides an additional example in which* (X_n) *converges almost surely to* $X \in L^1$*, but does not converge in* L^1*. It also gives a counterexample to the optional sampling theorem by considering the stopping time* $T = \min\{n \geq 0 : X_n = 0\}$*. Indeed,* T *is finite as* $X_n \to 0$ *a.s., but* $0 = \mathbb{E}(X_T) \neq \mathbb{E}(X_0) = 1$.