# Contents

1 Conditional expectation  
1.1 Conditioning with respect to an event $B \in \mathcal{F}$  
1.2 Conditioning with respect to a discrete random variable $Y$  
1.3 Conditioning with respect to a continuous random variable $Y$  
1.4 Conditioning with respect to a sub-$\sigma$-field $\mathcal{G}$  
1.5 Conditioning with respect to a random variable $Y$  

2 Martingales  
2.1 Basic definitions  
2.2 Stopping times  
2.3 Doob's optional stopping theorem, version 1  
2.4 Martingale transforms  
2.5 Doob's decomposition theorem  

3 Martingale convergence theorems  
3.1 Preliminary: Doob's martingale  
3.2 The martingale convergence theorem: first version  
3.3 Consequences of the theorem
1 Conditional expectation

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

1.1 Conditioning with respect to an event \( B \in \mathcal{F} \)

The conditional probability of an event \( A \in \mathcal{F} \) given another event \( B \in \mathcal{F} \) is defined as

\[
\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}, \quad \text{provided that } \mathbb{P}(B) > 0
\]

Notice that if \( A \) and \( B \) are independent, then \( \mathbb{P}(A|B) = \mathbb{P}(A) \); the conditioning does not affect the probability. This fact remains true in more generality (see below).

In a similar manner, the conditional expectation of an integrable random variable \( X \) given \( B \in \mathcal{F} \) is defined as

\[
\mathbb{E}(X|B) = \frac{\mathbb{E}(X 1_B)}{\mathbb{P}(B)}, \quad \text{provided that } \mathbb{P}(B) > 0
\]

1.2 Conditioning with respect to a discrete random variable \( Y \)

Let us assume that the random variable \( Y \) (is \( \mathcal{F} \)-measurable and) takes values in a countable set \( C \).

\[
\mathbb{P}(A|Y) = \varphi(Y), \quad \text{where } \varphi(y) = \mathbb{P}(A|\{Y = y\}), \quad y \in C
\]

\[
\mathbb{E}(X|Y) = \psi(Y), \quad \text{where } \psi(y) = \mathbb{E}(X|\{Y = y\}), \quad y \in C
\]

If \( X \) is also a discrete random variable with values in \( C \), then

\[
\mathbb{E}(X|Y) = \psi(Y), \quad \text{where } \psi(y) = \frac{\mathbb{E}(X 1_{\{Y = y\}})}{\mathbb{P}(\{Y = y\})} = \sum_{x \in C} x \frac{\mathbb{E}(1_{\{X = x\}} 1_{\{Y = y\}})}{\mathbb{P}(\{Y = y\})} = \sum_{x \in C} x \mathbb{P}(\{X = x\}|\{Y = y\})
\]

**Important remark.** \( \varphi(y) \) and \( \psi(y) \) are functions, while \( \varphi(Y) = \mathbb{P}(A|Y) \) and \( \psi(Y) = \mathbb{E}(X|Y) \) are *random variables*. They both are functions of the outcome of the random variable \( Y \), that is, they are \( \sigma(Y) \)-measurable random variables.

**Example.** Let \( X_1, X_2 \) be two independent dice rolls and let us compute \( \mathbb{E}(X_1 + X_2|X_2) = \psi(X_2) \), where

\[
\psi(y) = \mathbb{E}(X_1 + X_2|\{X_2 = y\}) = \frac{\mathbb{E}((X_1 + X_2) 1_{\{X_2 = y\}})}{\mathbb{P}(\{X_2 = y\})}
\]

\[
= \frac{\mathbb{E}(X_1 1_{\{X_2 = y\}}) + \mathbb{E}(X_2 1_{\{X_2 = y\}})}{\mathbb{P}(\{X_2 = y\})} = \frac{\mathbb{E}(X_1) \mathbb{E}(1_{\{X_2 = y\}}) + \mathbb{E}(y 1_{\{X_2 = y\}})}{\mathbb{P}(\{X_2 = y\})}
\]

\[
= \frac{\mathbb{E}(X_1) \mathbb{P}(\{X_2 = y\}) + y \mathbb{P}(\{X_2 = y\})}{\mathbb{P}(\{X_2 = y\})} = \mathbb{E}(X_1) + y
\]

where the independence assumption between \( X_1 \) and \( X_2 \) has been used in equality (a). So finally (as one would expect), \( \mathbb{E}(X_1 + X_2|X_2) = \mathbb{E}(X_1) + X_2 \), which can be explained intuitively as follows: the expectation of \( X_1 \) conditioned on \( X_2 \) is nothing but the expectation of \( X_1 \), as the outcome of \( X_2 \) provides no information on the outcome of \( X_1 \) (\( X_1 \) and \( X_2 \) being independent); on the other hand, the expectation of \( X_2 \) conditioned on \( X_2 \) is exactly \( X_2 \), as the outcome of \( X_2 \) is known.
1.3 Conditioning with respect to a continuous random variable $Y$?

In this case, one faces the following problem: if $Y$ is a continuous random variable, $\mathbb{P}(\{Y = y\}) = 0$ for all $y \in \mathbb{R}$. So a direct generalization of the above formulas to the continuous case is impossible at first sight. A possible solution to this problem is to replace the event $\{Y = y\}$ by $\{y \leq Y < y + \varepsilon\}$ and to take the limit $\varepsilon \to 0$ for the definition of conditional expectation. This actually works, but also leads to a paradox in the multidimensional setting (known as Borel’s paradox). In addition, some random variables are neither discrete, nor continuous. It turns out that the cleanest way to define conditional expectation in the general case is through $\sigma$-fields.

1.4 Conditioning with respect to a sub-$\sigma$-field $\mathcal{G}$

In order to define the conditional expectation in the general case, one needs the following proposition.

**Proposition 1.1.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $X$ be an integrable random variable on $(\Omega, \mathcal{F}, \mathbb{P})$. There exists then an integrable random variable $Z$ such that

(i) $Z$ is $\mathcal{G}$-measurable,

(ii) $\mathbb{E}(ZU) = \mathbb{E}(XU)$ for any random variable $U \in \mathcal{G}$-measurable and bounded.

Moreover, if $Z_1, Z_2$ are two integrable random variables satisfying (i) and (ii), then $Z_1 = Z_2$ a.s.

**Definition 1.2.** The above random variable $Z$ is called the conditional expectation of $X$ given $\mathcal{G}$ and is denoted as $\mathbb{E}(X|\mathcal{G})$. Because of the last part of the above proposition, it is defined up to a negligible set.

**Definition 1.3.** One further defines $\mathbb{P}(A|\mathcal{G}) = \mathbb{E}(1_A|\mathcal{G})$ for $A \in \mathcal{F}$.

**Remark.** Notice that as before, both $\mathbb{P}(A|\mathcal{G})$ and $\mathbb{E}(X|\mathcal{G})$ are ($\mathcal{G}$-measurable) random variables.

**Properties.** The above definition does not give a computation rule for the conditional expectation; it is only an existence theorem. The properties listed below will therefore be of help for computing conditional expectations. The proofs of the first two are omitted, while the next five are left as (important!) exercises.

- **Linearity.** $\mathbb{E}(cX + Y|\mathcal{G}) = c\mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$ a.s.

- **Monotonicity.** If $X \geq Y$ a.s., then $\mathbb{E}(X|\mathcal{G}) \geq \mathbb{E}(Y|\mathcal{G})$ a.s. (so if $X \geq 0$ a.s., then $\mathbb{E}(X|\mathcal{G}) \geq 0$ a.s.)

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

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

- **If $X$ is $\mathcal{G}$-measurable, then $\mathbb{E}(X|\mathcal{G}) = X$ a.s.**

- **If $Y$ is $\mathcal{G}$-measurable and bounded (or if $Y$ is $\mathcal{G}$-measurable and both $X$ and $Y$ are square-integrable; what actually matters here is that the random variable $XY$ is integrable), then $\mathbb{E}(XY|\mathcal{G}) = \mathbb{E}(X|\mathcal{G})Y$ a.s.**

- **If $\mathcal{H}$ is a sub-$\sigma$-field of $\mathcal{G}$, then $\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G}) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H}) = \mathbb{E}(X|\mathcal{H})$ a.s.** (in other words, the smallest $\sigma$-field always “wins”: this property is also known as the “towering property” of conditional expectation)
Some of the above properties are illustrated below with an example.

**Example.** Let $\Omega = \{1, \ldots, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}((\omega)) = \frac{1}{6}$ for $\omega = 1, \ldots, 6$ (the probability space of the die roll). Let also $X(\omega) = \omega$ be the outcome of the die roll and consider the two sub-$\sigma$-fields:

$$\mathcal{G} = \sigma(\{1,3\}, \{2\}, \{5\}, \{4,6\}) \quad \text{and} \quad \mathcal{H} = \sigma(\{1,3,5\}, \{2,4,6\})$$

Then $\mathbb{E}(X) = 3.5$,

$$\mathbb{E}(X|\mathcal{G})(\omega) = \begin{cases} 2 & \text{if } \omega \in \{1,3\} \text{ or } \omega = 2 \\ 5 & \text{if } \omega \in \{4,6\} \text{ or } \omega = 5 \end{cases} \quad \text{and} \quad \mathbb{E}(X|\mathcal{H})(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1,3,5\} \\ 4 & \text{if } \omega \in \{2,4,6\} \end{cases}$$

So $\mathbb{E}(\mathbb{E}(X|\mathcal{G})) = \mathbb{E}(\mathbb{E}(X|\mathcal{H})) = \mathbb{E}(X)$. Moreover,

$$\mathbb{E}(\mathbb{E}(X|\mathcal{G})|\mathcal{H})(\omega) = \begin{cases} \frac{1}{3}(2 + 2 + 5) = 3 & \text{if } \omega \in \{1,3,5\} \\ \frac{1}{3}(2 + 5 + 5) = 4 & \text{if } \omega \in \{2,4,6\} \end{cases} = \mathbb{E}(X|\mathcal{H})(\omega)$$

and

$$\mathbb{E}(\mathbb{E}(X|\mathcal{H})|\mathcal{G})(\omega) = \begin{cases} 3 & \text{if } \omega \in \{1,3\} \text{ or } \omega = 5 \\ 4 & \text{if } \omega \in \{4,6\} \text{ or } \omega = 2 \end{cases} = \mathbb{E}(X|\mathcal{H})(\omega)$$

The proposition below (given here without proof) is an extension of some of the above properties.

**Proposition 1.4.** Let $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$, $X, Y$ be two random variables such that $X$ is independent of $\mathcal{G}$ and $Y$ is $\mathcal{G}$-measurable, and let $\varphi : \mathbb{R}^2 \to \mathbb{R}$ be a Borel-measurable function such that $\mathbb{E}(|\varphi(X,Y)|) < +\infty$. Then

$$\mathbb{E}(\varphi(X,Y)|\mathcal{G}) = \psi(Y) \quad \text{a.s., where } \psi(y) = \mathbb{E}(\varphi(X,y))$$

This proposition has the following consequence: when computing the expectation of a function $\varphi$ of two independent random variables $X$ and $Y$, one can always divide the computation in two steps by writing

$$\mathbb{E}(\varphi(X,Y)) = \mathbb{E}(\mathbb{E}(\varphi(X,Y)|\mathcal{G})) = \mathbb{E}(\psi(Y))$$

where $\psi(y) = \mathbb{E}(\varphi(X,y))$ (this is actually nothing but Fubini’s theorem).

Finally, the proposition below (given again without proof) shows that Jensen’s inequality also holds for conditional expectation.

**Proposition 1.5.** Let $X$ be a random variable, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be Borel-measurable, convex and such that $\mathbb{E}(|\psi(X)|) < +\infty$. Then

$$\psi(\mathbb{E}(X|\mathcal{G})) \leq \mathbb{E}(\psi(X)|\mathcal{G}) \quad \text{a.s.}$$

In particular, $|\mathbb{E}(X|\mathcal{G})| \leq \mathbb{E}(|X||\mathcal{G})$ a.s.

### 1.5 Conditioning with respect to a random variable $Y$

Once the definition of conditional expectation with respect to a $\sigma$-field is set, it is natural to define it for a generic random variable $Y$:

$$\mathbb{E}(X|Y) = \mathbb{E}(X|\sigma(Y)) \quad \text{and} \quad \mathbb{P}(A|Y) = \mathbb{P}(A|\sigma(Y))$$

**Remark.** Since any $\sigma(Y)$-measurable random variable may be written as $g(Y)$, where $g$ is a Borel-measurable function, the definition of $\mathbb{E}(X|Y)$ may be rephrased as follows.

**Definition 1.6.** $E(X|Y) = \psi(Y)$, where $\psi : \mathbb{R} \to \mathbb{R}$ is the unique Borel-measurable function such that $\mathbb{E}(\psi(Y)g(Y)) = \mathbb{E}(Xg(Y))$ for any function $g : \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded.
In two particular cases, the function $\psi$ can be made explicit, which allows for concrete computations.

- If $X, Y$ are two discrete random variables with values in a countable set $C$, then

$$E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \sum_{x \in C} x \mathbb{P}(\{X = x\}|\{Y = y\}), \quad y \in C$$

which matches the formula given in Section 1.2. The proof that it also matches the theoretical definition of conditional expectation is left as an exercise.

- If $X, Y$ are two jointly continuous random variables with joint pdf $p_{X,Y}$, then

$$E(X|Y) = \psi(Y), \quad \text{where} \quad \psi(y) = \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dx, \quad y \in \mathbb{R}$$

and $p_Y$ is the marginal pdf of $Y$ given by $p_Y(y) = \int_{\mathbb{R}} p_{X,Y}(x,y) \, dx$, assumed here to be strictly positive.

Let us check that the random variable $\psi(Y)$ is indeed the conditional expectation of $X$ given $Y$ according to Definition 1.6: for any function $g: \mathbb{R} \to \mathbb{R}$ Borel-measurable and bounded, one has

$$\mathbb{E}(\psi(Y) g(Y)) = \int_{\mathbb{R}} \psi(y) g(y) p_Y(y) \, dy = \int_{\mathbb{R}} \left( \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dx \right) g(y) p_Y(y) \, dy = \int \int_{\mathbb{R}^2} x g(y) p_{X,Y}(x,y) \, dx \, dy = \mathbb{E}(X g(Y))$$

Finally, the conditional expectation satisfies the following proposition when $X$ is a square-integrable random variable.

**Proposition 1.7.** Let $X$ be a square-integrable random variable, $\mathcal{G}$ be a sub-$\sigma$-field of $\mathcal{F}$ and $G$ be the linear subspace of square-integrable and $\mathcal{G}$-measurable random variables. Then the conditional expectation $Z$ of $X$ with respect to $\mathcal{G}$ is equal a.s. to the random variable $Z$ satisfying

$$Z = \arg \min_{Y \in \mathcal{G}} \mathbb{E}((X - Y)^2)$$

In other words, this is saying that $Z$ is the orthogonal projection of $X$ onto the linear subspace $G$ of square-integrable and $\mathcal{G}$-measurable random variables (the scalar product considered here being $\langle X, Y \rangle = \mathbb{E}(XY)$). In particular,

$$\mathbb{E}((X - Z) U) = 0, \quad \text{for any } U \in G$$

which is nothing but a variant of condition (ii) in the definition of conditional expectation.

## 2 Martingales

### 2.1 Basic definitions

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

**Definition 2.1.** A *filtration* is a sequence $(\mathcal{F}_n, n \in \mathbb{N})$ of sub-$\sigma$-fields of $\mathcal{F}$ such that $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

**Example.** Let $\Omega = [0, 1]$, $\mathcal{F} = \mathcal{B}([0,1])$, $X_n(\omega) = n^{th}$ decimal of $\omega$, for $n \geq 1$. Let also $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$. 

5
Definitions 2.2. - A discrete-time process \((X_n, n \in \mathbb{N})\) is said to be adapted to the filtration \((\mathcal{F}_n, n \in \mathbb{N})\) if \(X_n\) is \(\mathcal{F}_n\)-measurable \(\forall n \in \mathbb{N}\).

- The natural filtration of a process \((X_n, n \in \mathbb{N})\) is defined as \(\mathcal{F}_n^X = \sigma(X_0, \ldots, X_n)\), \(n \in \mathbb{N}\). It represents the available amount of information about the process at time \(n\).

Remark. A process is adapted to its natural filtration, by definition.

Let now \((\mathcal{F}_n, n \in \mathbb{N})\) be a given filtration.

Definition 2.3. A discrete-time process \((M_n, n \in \mathbb{N})\) is a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) if

(i) \(\mathbb{E}(|M_n|) < +\infty, \forall n \in \mathbb{N}\).

(ii) \(M_n\) is \(\mathcal{F}_n\)-measurable, \(\forall n \in \mathbb{N}\) (i.e., \((M_n, n \in \mathbb{N})\) is adapted to \((\mathcal{F}_n, n \in \mathbb{N})\)).

(iii) \(\mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n\) a.s., \(\forall n \in \mathbb{N}\).

A martingale is therefore a fair game: the expectation of the process at time \(n + 1\) given the information at time \(n\) is equal to the value of the process at time \(n\).

Remark. Conditions (ii) and (iii) are actually redundant, as (iii) implies (ii).

Properties. If \((M_n, n \in \mathbb{N})\) is a martingale, then

\(- \mathbb{E}(M_{n+1}) = \mathbb{E}(M_n) (= \ldots = \mathbb{E}(M_0)), \forall n \in \mathbb{N}\) (by the first property of conditional expectation).

\(- \mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0\) a.s. (nearly by definition).

\(- \mathbb{E}(M_{n+m}|\mathcal{F}_n) = M_n\) a.s., \(\forall n, m \in \mathbb{N}\).

This last property is important, as it says that the martingale property propagates over time. Here is a short proof, which uses the towering property of conditional expectation:

\[ \mathbb{E}(M_{n+m}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(M_{n+m}|\mathcal{F}_{n+m-1})|\mathcal{F}_n) = \mathbb{E}(M_{n+m-1}|\mathcal{F}_n) = \ldots = \mathbb{E}(M_{n+1}|\mathcal{F}_n) = M_n \text{ a.s.} \]

Example: the simple symmetric random walk.

Let \((S_n, n \in \mathbb{N})\) be the simple symmetric random walk: \(S_0 = 0, S_n = X_1 + \ldots + X_n\), where the \(X_n\) are i.i.d. and \(\mathbb{P}(\{X_1 = +1\}) = \mathbb{P}(\{X_1 = -1\}) = 1/2\).

Let us define the following filtration: \(\mathcal{F}_0 = \{\emptyset, \Omega\}, \mathcal{F}_n = \sigma(X_1, \ldots, X_n), n \geq 1\). Then \((S_n, n \in \mathbb{N})\) is a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\). Indeed:

(i) \(\mathbb{E}(|S_1|) \leq \mathbb{E}(|X_1|) + \ldots + \mathbb{E}(|X_n|) = 1 + \ldots + 1 = n < +\infty, \forall n \in \mathbb{N}\).

(ii) \(S_n = X_1 + \ldots + X_n\) is a function of \((X_1, \ldots, X_n)\), i.e., is \(\sigma(X_1, \ldots, X_n) = \mathcal{F}_n\)-measurable.

(iii) We have

\[ \mathbb{E}(S_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n + X_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) = S_n + \mathbb{E}(X_{n+1}) = S_n + 0 = S_n \text{ a.s.} \]

The first equality on the second line follows from the fact that \(S_n\) is \(\mathcal{F}_n\)-measurable and that \(X_{n+1}\) is independent of \(\mathcal{F}_n = \sigma(X_1, \ldots, X_n)\).

\[ \square \]

Remark. Even though one uses generally the same letter “M” for both martingales and Markov process, these are a priori completely different processes! A possible way to state the Markov property is to say that

\[ \mathbb{E}(g(M_{n+1})|\mathcal{F}_n) = \mathbb{E}(g(M_{n+1})|X_n) \text{ a.s. for any } g : \mathbb{R} \to \mathbb{R} \text{ continuous and bounded} \]

which is clearly different from the above stated martingale property. Beyond the use of the same letter “M”, the confusion between the two notions comes also from the fact that the simple symmetric random walk is usually taken a paradigm example for both martingales and Markov processes.
**Generalization.** If the random variables $X_n$ are i.i.d. and such that $\mathbb{E}(|X_1|) < +\infty$ and $\mathbb{E}(X_1) = 0$, then $(S_n, n \in \mathbb{N})$ is also a martingale (in particular, $X_1 \sim \mathcal{N}(0,1)$ works).

**Definition 2.4.** Let $(\mathcal{F}_n, n \in \mathbb{N})$ be a filtration. A process $(M_n, n \in \mathbb{N})$ is a **submartingale** (resp. a **supermartingale**) with respect to $(\mathcal{F}_n, n \in \mathbb{N})$ if

(i) $\mathbb{E}(|M_n|) < +\infty$, $\forall n \in \mathbb{N}$.

(ii) $M_n$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$.

(iii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \geq M_n$ a.s., $\forall n \in \mathbb{N}$ (resp. $\mathbb{E}(M_{n+1}|\mathcal{F}_n) \leq M_n$ a.s., $\forall n \in \mathbb{N}$).

**Remarks.** - Not every process is either a sub- or a supermartingale!
- The appellations sub- and supermartingale are counter-intuitive. They are due to historical reasons.
- Condition (ii) is now necessary in itself, as (iii) does not imply it.
- If $(M_n, n \in \mathbb{N})$ is both a submartingale and a supermartingale, then it is a martingale.

**Example: the simple asymmetric random walk.**

- If $\mathbb{P}(\{X_1 = +1\}) = p = 1 - \mathbb{P}(\{X_1 = -1\})$ with $p \geq 1/2$, then $S_n = X_1 + \ldots + X_n$ is a submartingale.

- More generally, $S_n = X_1 + \ldots + X_n$ is a submartingale if $\mathbb{E}(X_1) \geq 0$.

**Proposition 2.5.** If $(M_n, n \in \mathbb{N})$ is a martingale with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$ and $\varphi : \mathbb{R} \to \mathbb{R}$ is a Borel-measurable and convex function such that $\mathbb{E}(|\varphi(M_n)|) < +\infty$, $\forall n \in \mathbb{N}$, then $(\varphi(M_n), n \in \mathbb{N})$ is a submartingale.

**Proof.** (i) $\mathbb{E}(|\varphi(M_n)|) < +\infty$ by assumption.

(ii) $\varphi(M_n)$ is $\mathcal{F}_n$-measurable as $M_n$ is (and $\varphi$ is Borel-measurable).

(iii) $\mathbb{E}(\varphi(M_{n+1})|\mathcal{F}_n) \geq \varphi(\mathbb{E}(M_{n+1}|\mathcal{F}_n)) = \varphi(M_n)$ a.s.

In (iii), the first inequality follows from Jensen’s inequality and the second follows from the fact that $M$ is a martingale.

**Example.** If $(M_n, n \in \mathbb{N})$ is a square-integrable martingale (i.e., $\mathbb{E}(M_n^2) < +\infty$, $\forall n \in \mathbb{N}$), then the process $(M_n^2, n \in \mathbb{N})$ is a submartingale (as $x \mapsto x^2$ is convex).

### 2.2 Stopping times

**Definitions 2.6.** - A **random time** is a random variable $T$ with values in $\mathbb{N} \cup \{+\infty\}$. It is said to be **finite** if $T(\omega) < +\infty$ for every $\omega \in \Omega$ and **bounded** if there exists moreover an integer $N$ such that $T(\omega) \leq N$ for every $\omega \in \Omega$ (Notice that a finite random time is not necessarily bounded).

- Let $(X_n, n \in \mathbb{N})$ be a stochastic process and assume $T$ is finite. One then defines $X_T(\omega) = X_{T(\omega)}(\omega) = \sum_{n \in \mathbb{N}} X_n(\omega) 1_{\{T=n\}}(\omega)$.

- A **stopping time** with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$ is a random time $T$ such that $\{T = n\} \in \mathcal{F}_n$, $\forall n \in \mathbb{N}$.

**Example.** Let $(X_n, n \in \mathbb{N})$ be a process adapted to $(\mathcal{F}_n, n \in \mathbb{N})$ and $a > 0$. Then $T_a = \inf\{n \in \mathbb{N} : |X_n| \geq a\}$ is a stopping time with respect to $(\mathcal{F}_n, n \in \mathbb{N})$. Indeed:

$$\{T_a = n\} = \{ |X_k| < a, \forall 0 \leq k \leq n-1 \text{ and } |X_n| \geq a \}$$

$$= \bigcap_{k=0}^{n-1} \left( \{ |X_k| < a \} \cap \{ |X_n| \geq a \} \right) \in \mathcal{F}_n, \quad \forall n \in \mathbb{N}$$
Definition 2.7. Let $T$ be a stopping time with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$. One defines the information one possesses at time $T$ as the following $\sigma$-field:

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

Facts.

- If $T(\omega) = N \forall \omega \in \Omega$, then $\mathcal{F}_T = \mathcal{F}_N$. This is obvious from the definition.
- If $T_1, T_2$ are stopping times such that $T_1(\omega) \leq T_2(\omega) \forall \omega \in \Omega$, then $\mathcal{F}_{T_1} \subset \mathcal{F}_{T_2}$. Indeed, if $T_1(\omega) \leq T_2(\omega) \forall \omega \in \Omega$ and $A \in \mathcal{F}_{T_1}$, then for all $n \in \mathbb{N}$, we have:

$$A \cap \{ T_2 = n \} = A \cap (\cup_{k=1}^{n} \{ T_1 = k \}) \cap \{ T_2 = n \} = \left( \bigcup_{k=1}^{n} A \cap \{ T_1 = k \} \right) \cap \{ T_2 = n \} \in \mathcal{F}_n$$

so $A \in \mathcal{F}_{T_2}$. By the way, here is an example of stopping times $T_1, T_2$ such that $T_1(\omega) \leq T_2(\omega) \forall \omega \in \Omega$: let $0 < a < b$ and consider $T_1 = \inf\{ n \in \mathbb{N} : |X_n| \geq a \}$ and $T_2 = \inf\{ n \in \mathbb{N} : |X_n| \geq b \}$.
- A random variable $Y$ is $\mathcal{F}_T$-measurable if and only if $Y 1_{\{ T = n \}}$ is $\mathcal{F}_n$-measurable, $\forall n \in \mathbb{N}$. As a consequence: if $(X_n, n \in \mathbb{N})$ is adapted to $(\mathcal{F}_n, n \in \mathbb{N})$, then $X_T$ is $\mathcal{F}_T$-measurable.

2.3 Doob’s optional stopping theorem, version 1

Let $(M_n, n \in \mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n, n \in \mathbb{N})$, $N \in \mathbb{N}$ be fixed and $T_1, T_2$ be two stopping times such that $0 \leq T_1(\omega) \leq T_2(\omega) \leq N < +\infty, \forall \omega \in \Omega$. Then

$$\mathbb{E}(M_{T_2} | \mathcal{F}_{T_1}) = M_{T_1} \text{ a.s.}$$

In particular, $\mathbb{E}(M_{T_2}) = \mathbb{E}(M_{T_1})$.

In particular, if $T$ is a stopping time such that $0 \leq T(\omega) \leq N < +\infty, \forall \omega \in \Omega$, then $\mathbb{E}(M_T) = \mathbb{E}(M_0)$.

Remarks. - The above theorem says that the martingale property holds even if one is given the option to stop at any (bounded) stopping time.
- The theorem also holds for sub- and supermartingales (i.e., if $M$ is a submartingale, then $\mathbb{E}(M_{T_2} | \mathcal{F}_{T_1}) \geq M_{T_1}$, a.s.).

Proof. - We first show that if $T$ is a stopping time such that $0 \leq T(\omega) \leq N, \forall \omega \in \Omega$, then

$$\mathbb{E}(M_N | \mathcal{F}_T) = M_T \quad (1)$$

Indeed, let $Z = M_T = \sum_{n=0}^{N} M_n 1_{\{ T = n \}}$. We check below that $Z$ is the conditional expectation of $M_N$ given $\mathcal{F}_T$:

(i) $Z$ is $\mathcal{F}_T$-measurable: $Z 1_{\{ T = n \}} = M_n 1_{\{ T = n \}}$ is $\mathcal{F}_n$-measurable $\forall n$, so $Z$ is $\mathcal{F}_T$-measurable.

(ii) $\mathbb{E}(ZU) = \mathbb{E}(M_N U), \forall U$ $\mathcal{F}_T$-measurable and bounded:

$$\mathbb{E}(ZU) = \sum_{n=0}^{N} \mathbb{E}(M_n 1_{\{ T = n \}} U) = \sum_{n=0}^{N} \mathbb{E}(M_n U | \mathcal{F}_n) 1_{\{ T = n \}} \mathcal{F}_n \text{-measurable} = \sum_{n=0}^{N} \mathbb{E}(M_n 1_{\{ T = n \}} U) = \mathbb{E}(M_N U)$$

- Second, let us check that $\mathbb{E}(M_{T_2} | \mathcal{F}_{T_1}) = M_{T_1}$:

$$M_{T_1} \quad (1) \text{ with } T = T_1 \quad \mathcal{F}_{T_1} \subset \mathcal{F}_{T_2} \quad \mathbb{E}(\mathbb{E}(M_N | \mathcal{F}_{T_2}) | \mathcal{F}_{T_1}) \quad (1) \text{ with } T = T_2 \quad \mathbb{E}(M_{T_2} | \mathcal{F}_{T_1})$$

This concludes the proof of the theorem.
2.4 Martingale transforms

**Definition 2.8.** A process \((H_n, n \in \mathbb{N})\) is said to be *predictable* with respect to a filtration \((\mathcal{F}_n, n \in \mathbb{N})\) if \(H_0 = 0\) and \(H_n\) is \(\mathcal{F}_{n-1}\)-measurable \(\forall n \geq 1\).

**Remark.** If a process is predictable, then it is adapted.

Let now \((\mathcal{F}_n, n \in \mathbb{N})\) be a filtration, \((H_n, n \in \mathbb{N})\) be a predictable process with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) and \((M_n, n \in \mathbb{N})\) be a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\).

**Definition 2.9.** The process \(G\) defined as
\[
G_0 = 0, \quad G_n = (H \cdot M)_n = \sum_{i=1}^{n} H_i (M_i - M_{i-1}), \quad n \geq 1
\]
is called the *martingale transform of* \(M\) through \(H\).

**Remark.** This process is the discrete version of the stochastic integral. It represents the gain obtained by applying the strategy \(H\) to the game \(M\):
- \(H_i\) = amount bet on day \(i\) \((\mathcal{F}_{i-1}\text{-measurable})\).
- \(M_i - M_{i-1}\) = increment of the process \(M\) on day \(i\).
- \(G_n\) = gain on day \(n\).

**Proposition 2.10.** If \(H_n\) is a bounded random variable for each \(n\) (i.e., \(|H_n(\omega)| \leq K_n \; \forall \omega \in \Omega\)), then the process \(G\) is a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\).

In other words, one cannot win on a martingale!

**Proof.** (i) \(\mathbb{E}(|G_n|) \leq \sum_{i=1}^{n} \mathbb{E}(|H_i| |M_i - M_{i-1}|) \leq \sum_{i=1}^{n} K_i \left(\mathbb{E}(|M_i|) + \mathbb{E}(|M_{i-1}|)\right) < +\infty\).
(ii) \(G_n\) is \(\mathcal{F}_n\)-measurable by construction.
(iii) \(\mathbb{E}(G_{n+1}|\mathcal{F}_n) = \mathbb{E}(G_n + H_{n+1} (M_{n+1} - M_n)|\mathcal{F}_n) = G_n + H_{n+1} \mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = G_n + 0 = G_n. \quad \square \)

**Example: “the” martingale.**

Let \((M_n, n \in \mathbb{N})\) be the simple symmetric random walk \((M_n = \xi_1 + \ldots + \xi_n)\) and consider the following strategy:
\[
H_0 = 0, \quad H_1 = 1, \quad H_{n+1} = \begin{cases} 2H_n, & \text{if } \xi_1 = \ldots = \xi_n = -1 \\ 0, & \text{otherwise} \end{cases}
\]
Notice that all the \(H_n\) are bounded random variables. Then by the above proposition, the process \(G\) defined as
\[
G_0 = 0, \quad G_n = \sum_{i=1}^{n} H_i (M_i - M_{i-1}) = \sum_{i=1}^{n} H_i \xi_i, \quad n \geq 1
\]
is a martingale. So \(\mathbb{E}(G_n) = \mathbb{E}(G_0) = 0, \; \forall n \in \mathbb{N}\). Let now
\[
T = \inf\{n \geq 1 : \xi_n = +1\}
\]
\(T\) is a stopping time and it is easily seen that \(G_T = +1\). But then \(\mathbb{E}(G_T) = 1 \neq 0 = \mathbb{E}(G_0)\)? Is there a contradiction? Actually no. The optional stopping theorem does not apply here, because the time \(T\) is unbounded: \(\mathbb{P}(T = n) = 2^{-n}, \; \forall n \in \mathbb{N}\), i.e., there does not exist \(N\) fixed such that \(T(\omega) \leq N, \; \forall \omega \in \Omega\).
2.5 Doob’s decomposition theorem

**Theorem 2.11.** Let \((X_n, n \in \mathbb{N})\) be a submartingale with respect to a filtration \((\mathcal{F}_n, n \in \mathbb{N})\). Then there exists a martingale \((M_n, n \in \mathbb{N})\) with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) and a process \((A_n, n \in \mathbb{N})\) predictable with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) and increasing (i.e., \(A_n \leq A_{n+1} \forall n \in \mathbb{N}\)) such that \(A_0 = 0\) and \(X_n = M_n + A_n, \forall n \in \mathbb{N}\). Moreover, this decomposition of the process \(X\) is unique.

**Proof.** (main idea)
\(\mathbb{E}(X_{n+1}|\mathcal{F}_n) \geq X_n\), so a natural candidate for the process \(A\) is to set \(A_0 = 0\) and \(A_{n+1} = A_n + \mathbb{E}(X_{n+1}|\mathcal{F}_n) - X_n \geq A_n\), which is a predictable and increasing process. Then, \(M_0 = X_0\) and \(M_{n+1} - M_n = X_{n+1} - X_n - (A_{n+1} - A_n) = X_{n+1} - \mathbb{E}(X_{n+1}|\mathcal{F}_n)\) is indeed a martingale, as \(\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0\). \(\square\)

3 Martingale convergence theorems

3.1 Preliminary: Doob’s martingale

**Proposition 3.1.** Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space, \((\mathcal{F}_n, n \in \mathbb{N})\) be a filtration and \(X : \Omega \to \mathbb{R}\) be an \(\mathcal{F}\)-measurable and integrable random variable. Then the process \((M_n, n \in \mathbb{N})\) defined as
\[M_n = \mathbb{E}(X|\mathcal{F}_n), \quad n \in \mathbb{N}\]
is a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\).

**Proof.** (i) \(\mathbb{E}(|M_n|) = \mathbb{E}(|\mathbb{E}(X|\mathcal{F}_n)|) \leq \mathbb{E}(\mathbb{E}(|X||\mathcal{F}_n)) = \mathbb{E}(|X|) < +\infty\), for all \(n \in \mathbb{N}\).
(ii) By the definition of conditional expectation, \(M_n = \mathbb{E}(X|\mathcal{F}_n)\) is \(\mathcal{F}_n\)-measurable, for all \(n \in \mathbb{N}\).
(iii) \(\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \mathbb{E}(\mathbb{E}(X|\mathcal{F}_{n+1})|\mathcal{F}_n) = \mathbb{E}(X|\mathcal{F}_n) = M_n\), for all \(n \in \mathbb{N}\). \(\square\)

**Remarks.** - This process describes the situation where one acquires more and more information about a random variable. Think e.g. at the case where \(X\) is a number drawn uniformly at random between 0 and 1, and one reads this number from left to right: while reading, one obtains more and more information about the number.
- Is this a very particular type of martingale? No! As the following paragraph shows, this “example” is actually quite general...

3.2 The martingale convergence theorem: first version

**Theorem 3.2.** Let \((M_n, n \in \mathbb{N})\) be a square-integrable martingale (i.e. a martingale such that \(\mathbb{E}(M_n^2) < +\infty\) for all \(n \in \mathbb{N}\)) with respect to a filtration \((\mathcal{F}_n, n \in \mathbb{N})\). Under the additional assumption that
\[\sup_{n \in \mathbb{N}} \mathbb{E}(M_n^2) < +\infty, (2)\]
there exists a limiting random variable \(M_\infty\) such that
(i) \(M_n \xrightarrow{n \to \infty} M_\infty\) almost surely.
(ii) \(\lim_{n \to \infty} \mathbb{E}((M_n - M_\infty)^2) = 0\) (quadratic convergence).
(iii) \(M_n = \mathbb{E}(M_\infty|\mathcal{F}_n)\), for all \(n \in \mathbb{N}\) (this last property is referred to as the martingale \(M\) being “closed at infinity”).
Remarks. - Condition (2) is of course much stronger than just asking that $\mathbb{E}(M_n^2) < +\infty$ for every $n$. Think for example at the simple symmetric random walk $S_n$: $\mathbb{E}(S_n^2) = n < +\infty$ for every $n$, but the supremum is infinite.

- By conclusion (iii) in the theorem, any square-integrable martingale satisfying condition (2) is actually a Doob martingale (take $X = M_\infty$)!

- A priori, one could think that all the conclusions of the theorem hold true if one replaces all the squares by absolute values in the above statement (such as e.g. replacing condition (2) by $\sup_{n \in \mathbb{N}} \mathbb{E}(|M_n|) < +\infty$, etc.). This is wrong, and we will see interesting counter-examples later.

- A stronger condition than (2) (leading therefore to the same conclusion) is the following:

$$\sup_{n \in \mathbb{N}} \sup_{\omega \in \Omega} |M_n(\omega)| < +\infty \quad (3)$$

Martingales satisfying this stronger condition are called bounded martingales.

Example 3.3. Let $M_0 = x$, where $x \in [0, 1]$ is a fixed number, and let us define recursively:

$$M_{n+1} = \begin{cases} M_n^2, & \text{with probability } \frac{1}{2} \\ 2M_n - M_n^2, & \text{with probability } \frac{1}{2} \end{cases}$$

The process $M$ is a bounded martingale. Indeed:

(i) By induction, if $M_n \in [0, 1]$, then $M_{n+1} \in [0, 1]$, for every $n \in \mathbb{N}$, so as $M_0 = x \in [0, 1]$, we obtain

$$\sup_{n \in \mathbb{N}} \sup_{\omega \in \Omega} |M_n(\omega)| \leq 1 < +\infty$$

(ii) $\mathbb{E}(M_{n+1}|\mathcal{F}_n) = \frac{1}{2} M_n^2 + \frac{1}{2} (2M_n - M_n^2) = M_n$, for every $n \in \mathbb{N}$.

By the theorem, there exists therefore a random variable $M_\infty$ such that the three conclusions of the theorem hold. In addition, it can be shown by contradiction that $M_\infty$ takes values in the binary set $\{0, 1\}$ only, so that

$$x = \mathbb{E}(M_0) = \mathbb{E}(M_\infty) = \mathbb{P}(M_\infty = 1)$$

3.3 Consequences of the theorem

Before diving into the proof of the above important theorem, let us first explore a few of its interesting consequences.

Optional stopping theorem, version 2. Let $(\mathcal{F}_n, n \in \mathbb{N})$ be a filtration, let $(M_n, n \in \mathbb{N})$ be a square-integrable martingale with respect to $(\mathcal{F}_n, n \in \mathbb{N})$ which satisfies condition (2) and let $0 \leq T_1 \leq T_2 \leq +\infty$ be two stopping times with respect to $(\mathcal{F}_n, n \in \mathbb{N})$. Then

$$\mathbb{E}(M_{T_2}|\mathcal{F}_{T_1}) = M_{T_1} \text{ a.s. and } \mathbb{E}(M_{T_2}) = \mathbb{E}(M_{T_1})$$

Proof. Simply replace $N$ by $\infty$ in the proof of the first version and use the fact that $M$ is a closed martingale by the convergence theorem. \qed

Stopped martingale. Let $(M_n, n \in \mathbb{N})$ be a martingale and $T$ be a stopping time with respect to a filtration $(\mathcal{F}_n, n \in \mathbb{N})$, without any further assumption. Let us also define the stopped process

$$(M_{T \wedge n}, n \in \mathbb{N})$$

where $T \wedge n = \min\{T, n\}$ by definition. Then this stopped process is also a martingale with respect to $(\mathcal{F}_n, n \in \mathbb{N})$ (we skip the proof here, which uses the first version of the optional stopping theorem).
Optional stopping theorem, version 3. Let \((M_n, n \in \mathbb{N})\) be a martingale with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) such that there exists \(c > 0\) with \(|M_{n+1}(\omega) - M_n(\omega)| \leq c\) for all \(\omega \in \Omega\) and \(n \in \mathbb{N}\) (this assumption ensures that the martingale does not make jumps of uncontrolled size: the simple symmetric random walk \(S_n\) satisfies in particular this assumption). Let also \(a, b > 0\) and

\[ T = \inf\{n \in \mathbb{N} : M_n \leq -a \text{ or } M_n \geq b\}. \]

Observe that \(T\) is a stopping time with respect to \((\mathcal{F}_n, n \in \mathbb{N})\) and that \(-a - c \leq M_T(\omega) \leq b + c\) for all \(\omega \in \Omega\) and \(n \in \mathbb{N}\). In particular,

\[ \sup_{n \in \mathbb{N}} \mathbb{E}(M^2_{T \wedge n}) < +\infty \]

so the stopped process \((M_{T \wedge n}, n \in \mathbb{N})\) satisfies the assumptions of the first version of the martingale convergence theorem. By the conclusion of this theorem, the stopped martingale \((M_{T \wedge n}, n \in \mathbb{N})\) is closed, i.e. it admits a limit \(M_{T \wedge \infty} = M_T\) and

\[ \mathbb{E}(M_T) = \mathbb{E}(M_{T \wedge \infty}) = \mathbb{E}(M_{T \wedge 0}) = \mathbb{E}(M_0) \]

**Application.** Let \((S_n, n \in \mathbb{N})\) be the simple symmetric random walk (which satisfies the above assumptions with \(c = 1\)) and \(T\) be the above stopping time (with \(a, b\) positive integers). Then \(\mathbb{E}(S_T) = \mathbb{E}(S_0) = 0\). Given that \(S_T \in \{-a, +b\}\), we obtain

\[ 0 = \mathbb{E}(S_T) = (+b) \mathbb{P}(\{S_T = +b\}) + (-a) \mathbb{P}(\{S_T = -a\}) = bp - a(1 - p), \quad \text{where } p = \mathbb{P}(\{S_T = +b\}) \]

From this, we deduce that \(\mathbb{P}(\{S_T = +b\}) = p = \frac{a}{a+b}\).

**Remark.** Note that the same reasoning does not hold if we replace the stopping time \(T\) by a stopping time of the form

\[ T' = \inf\{n \in \mathbb{N} : M_n \geq b\} \]

There is indeed no guarantee in this case that the stopped martingale \((M_{T' \wedge n}, n \in \mathbb{N})\) is bounded (from below).