Advanced Probability and Applications (Part II)

Olivier Lévêque, IC–LTHI, EPFL
(with special thanks to Simon Guilloud for the figures)

July 31, 2018

Contents

1 Conditional expectation **Week 9**

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.

$$\begin{aligned} \mathbb{P}(A|Y) = \varphi(Y), &\qquad \text{where } \varphi(y) = \mathbb{P}(A|\{Y=y\}), \quad y \in C \\ \mathbb{E}(X|Y) = \psi(Y), &\qquad \text{where } \psi(y) = \mathbb{E}(X|\{Y=y\}), \quad y \in C \end{aligned}$$

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

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

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

$$\begin{aligned} \psi(y) &= \mathbb{E}(X_1 + X_2|\{X_2 = y\}) = \frac{\mathbb{E}((X_1 + X_2)\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \\ &= \frac{\mathbb{E}(X_1\, 1_{\{X_2=y\}}) + \mathbb{E}(X_2\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \overset{(a)}{=} \frac{\mathbb{E}(X_1)\, \mathbb{E}(1_{\{X_2=y\}}) + \mathbb{E}(y\, 1_{\{X_2=y\}})}{\mathbb{P}(\{X_2 = y\})} \\ &= \frac{\mathbb{E}(X_1)\, \mathbb{P}(\{X_2 = y\}) + y\, \mathbb{P}(\{X_2 = y\})}{\mathbb{P}(\{X_2 = y\})} = \mathbb{E}(X_1) + y \end{aligned}$$

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 σ-fields.

1.4 Conditioning with respect to a sub-σ-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-σ-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 $\mathcal{G}$-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. $\mathbb{E}(c\,X + Y|\mathcal{G}) = c\,\mathbb{E}(X|\mathcal{G}) + \mathbb{E}(Y|\mathcal{G})$ 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-σ-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 σ-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 $\Omega = \{1, \dots, 6\}$, $\mathcal{F} = \mathcal{P}(\Omega)$ and $\mathbb{P}(\{\omega\}) = \frac{1}{6}$ for $\omega = 1, \dots, 6$ (the probability space of the die roll). Let also $X(\omega) = \omega$ be the outcome of the die roll and consider the two sub-σ-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-σ-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

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

This proposition has the following consequence: when computing the expectation of a function φ 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-σ-field of $\mathcal{F}$ and $\psi : \mathbb{R} \to \mathbb{R}$ be Borel-measurable, convex and such that $\mathbb{E}(|\psi(X)|) < +\infty$. Then

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

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 σ-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 ψ 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

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

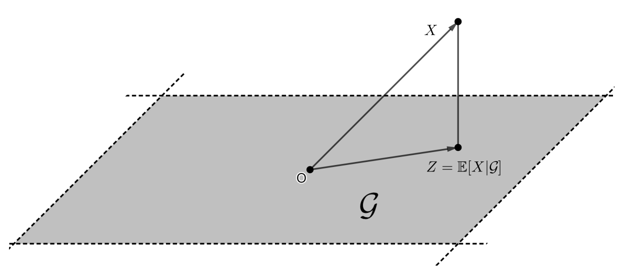
□

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

$$Z = \underset{Y \in G}{\text{argmin}} \; \mathbb{E}((X - Y)^2)$$

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 $\langle X, Y \rangle = \mathbb{E}(XY)$), as illustrated bleow:



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

Week 10

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-σ-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 ω, for $n \geq 1$. Let also $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. Then $\mathcal{F}_n \subset \mathcal{F}_{n+1}$, $\forall n \in \mathbb{N}$.

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:

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

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

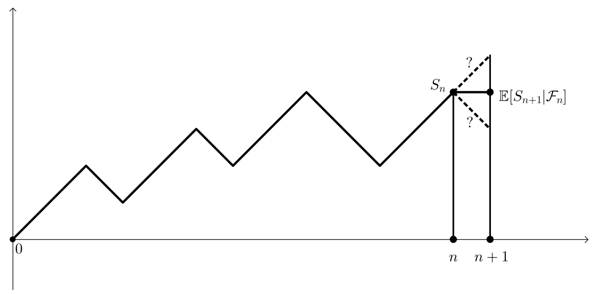
(ii) $S_n = X_1 + \ldots + X_n$ is a function of $(X_1, \ldots, X_n)$, i.e., is $\sigma(X_1, \ldots, X_n) = \mathcal{F}_n$-measurable.

(iii) We have

$$\begin{aligned} \mathbb{E}(S_{n+1}|\mathcal{F}_n) &= \mathbb{E}(S_n + X_{n+1}|\mathcal{F}_n) = \mathbb{E}(S_n|\mathcal{F}_n) + \mathbb{E}(X_{n+1}|\mathcal{F}_n) \\ &= S_n + \mathbb{E}(X_{n+1}) = S_n + 0 = S_n \quad \text{a.s.} \end{aligned}$$

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 $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$. □

Here is an additional illustration of the martingale property of the simple symmetric random walk:



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

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

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 φ 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:

$$\begin{aligned} \{T_a = n\} &= \{|X_k| < a,\ \forall 0 \leq k \leq n-1 \text{ and } |X_n| \geq a\} \\ &= \bigcap_{k=0}^{n-1} \underbrace{\{|X_k| < a\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \cap \{|X_n| \geq a\} \in \mathcal{F}_n, \quad \forall n \in \mathbb{N} \end{aligned}$$

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 σ-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:

$$A \cap \{T_2 = n\} = A \cap (\cup_{k=1}^n \{T_1 = k\}) \cap \{T_2 = n\} = \Big(\cup_{k=1}^n \underbrace{A \cap \{T_1 = k\}}_{\in \mathcal{F}_k \subset \mathcal{F}_n} \Big) \cap \{T_2 = n\} \in \mathcal{F}_n$$

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 **Week 11**

Let $(M_n,\, n \in \mathbb{N})$ be a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$, $N \in NN$ 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

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

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:

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

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

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

This concludes the proof of the theorem. □

2.4 The reflection principle

Let $(S_n, \, n \in \mathbb{N})$ be the simple symmetric random walk and

$$T = \inf\{n \geq 1 \; : \; S_n = +1 \quad \text{or} \quad n = N\}$$

As S is a martingale and T is a bounded stopping time (indeed, $T(\omega) \leq N$ for every $\omega \in \Omega$), the optional stopping theorem applies here, so it holds that $\mathbb{E}(S_T) = \mathbb{E}(S_0) = 0$. But what is the distribution of the random variable S_T? Intuitively, for N large, S_T will be $+1$ with high probability, but in case it does not this value, what is the average loss we should expect? More precisely, we are asking here for the value of

$$\mathbb{E}\left(S_T \,\middle|\, \left\{\max_{0 \leq n \leq N} S_n \leq 0\right\}\right) = \mathbb{E}\left(S_N \,\middle|\, \left\{\max_{0 \leq n \leq N} S_n \leq 0\right\}\right) = \frac{\mathbb{E}\left(S_N \, 1_{\{\max_{0 \leq n \leq N} S_n \leq 0\}}\right)}{\mathbb{P}\left(\{\max_{0 \leq n \leq N} S_n \leq 0\}\right)} \tag{2}$$

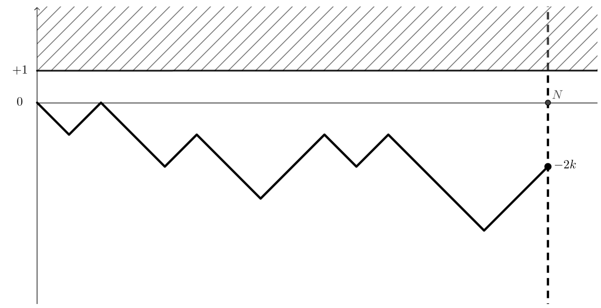
Let us first compute the denominator in (2), assuming that N is even to simplify notations:

$$\mathbb{P}\left(\left\{\max_{0 \leq n \leq N} S_n \leq 0\right\}\right) = \mathbb{P}(\{S_n \leq 0, \forall 0 \leq n \leq N\}) = \sum_{k \geq 0} \mathbb{P}(\{S_N = -2k, S_n \leq 0, \forall 0 \leq n \leq N-1\})$$

noticing that S_N can only take even values (because N itself is even) and that we are asking here that $S_N \leq 0$. Let us now consider a *fixed* value of $k \geq 0$. In order to compute the probability

$$\mathbb{P}(\{S_N = -2k, S_n \leq 0, \forall 0 \leq n \leq N-1\})$$

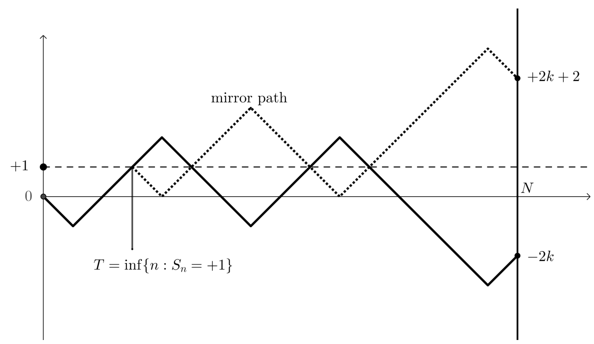
we should enurmate all paths of the following form: (please turn the page)



but this is rather complicated combinatorics. In order to avoid such a computation, first observe that

$$\mathbb{P}(\{S_N=-2k, S_n\le 0, \forall 0\le n\le N-1,\})=\mathbb{P}(\{S_N=-2k\})-\mathbb{P}(\{S_N=-2k,\ \exists 1\le n\le N_1 \text{ with } S_n=+1\})$$

A second important observation, which is at the heart of the reflection principle, is that to each path going from 0 (at time 0) to $-2k$ (at time N) "via" $+1$ corresponds a mirror path that goes from 0 to $2k+2$, also "via" $+1$, as illustrated below:



so that in total:

$$\mathbb{P}(\{S_N=-2k,\ \exists 1\le n\le N_1 \text{ with } S_n=+1\})=\mathbb{P}(\{S_N=2k+2,\ \exists 1\le n\le N_1 \text{ with } S_n=+1\})$$

A third observation is that for any $k\ge 0$, there is no way to go from 0 to $2k+2$ without crossing the $+1$ line, so that

$$\mathbb{P}(\{S_N=2k+2,\ \exists 1\le n\le N_1 \text{ with } S_n=+1\})=\mathbb{P}(\{S_N=2k+2\})$$

Finally, we obtain

$$\mathbb{P}\left(\left\{\max_{0\le n\le N} S_n\le 0\right\}\right)=\sum_{k\ge 0}(\mathbb{P}(\{S_N=-2k\})-\mathbb{P}(\{S_N=2k+2\}))=\sum_{k\ge 0}(\mathbb{P}(\{S_N=2k\})-\mathbb{P}(\{S_N=2k+2\}))$$

by symmetry. But this is a telescopic sum, and we know that for finite N, it ends before $k=+\infty$. At the end, we therefore obtain:

$$\mathbb{P}\left(\left\{\max_{0\le n\le N} S_n\le 0\right\}\right)=\mathbb{P}(\{S_N=0\})$$

which can be computed via simple combinatorics (writing here $N=2M$):

$$\mathbb{P}(\{S_{2M}=0\})=\frac{1}{2^{2M}}\binom{2M}{M}=\frac{1}{2^{2M}}\,\frac{(2M)!}{(M!)^2}$$

which gives for large M, using Stirling's formula: $M!\simeq M^M e^{-M}\sqrt{2\pi M}$:

$$\mathbb{P}(\{S_{2M}=0\}\simeq\frac{1}{2^{2M}}\,\frac{(2M)^{2M}\,e^{-2M}\,\sqrt{4\pi M}}{(M\,e^{-M}\sqrt{2\pi M})^2}=\frac{1}{\sqrt{\pi M}}$$

This leads to the approximation for large N:

$$\mathbb{P}\left(\left\{\max_{0\le n\le N} S_n \le 0\right\}\right) \simeq \sqrt{\frac{2}{\pi N}}$$

Finally, the optional stopping theorem spares us the direct computation of the numerator in (2), since

$$0 = \mathbb{E}(S_T) = 1\cdot \mathbb{P}\left(\left\{\max_{0\le n\le N} S_n \ge +1\right\}\right) + \mathbb{E}\left(S_N\, 1_{\{\max_{0\le n\le N} S_n \le 0\}}\right)$$

so

$$\mathbb{E}\left(S_N\, 1_{\{\max_{0\le n\le N} S_n \le 0\}}\right) = -1 + \mathbb{P}\left(\left\{\max_{0\le n\le N} S_n \le 0\right\}\right) \simeq -1 + \sqrt{\frac{2}{\pi N}}$$

for large N, and finally

$$\mathbb{E}\left(S_T \,\middle|\, \left\{\max_{0\le n\le N} S_n \le 0\right\}\right) \simeq \frac{-1+\sqrt{\frac{2}{\pi N}}}{\sqrt{\frac{2}{\pi N}}} = 1 - \sqrt{\frac{\pi N}{2}}$$

for large N. In conclusion, in case S does not reach the value $+1$ during the time interval $\{0, \ldots N\}$, we should expect a loss of order $-\sqrt{N}$.

2.5 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 $\forall n \ge 1$.

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

$$G_0 = 0, \quad G_n = (H\cdot M)_n = \sum_{i=1}^n H_i(M_i - M_{i-1}), \quad n \ge 1$$

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 ($\mathcal{F}_{i-1}$-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., $|H_n(\omega)| \le K_n \;\forall \omega \in \Omega$), 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) $\mathbb{E}(|G_n|) \le \sum_{i=1}^n \mathbb{E}(|H_i|\,|M_i - M_{i-1}|) \le \sum_{i=1}^n K_i\,(\mathbb{E}(|M_i|) + \mathbb{E}(|M_{i-1}|)) < +\infty$.

(ii) G_n is $\mathcal{F}_n$-measurable by construction.

(iii) $\mathbb{E}(G_{n+1}|\mathcal{F}_n) = \mathbb{E}(G_n + H_{n+1}\,(M_{n+1} - M_n)|\mathcal{F}_n) = G_n + H_{n+1}\,\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = G_n + 0 = G_n$. □

Example: "the" martingale.

Let $(M_n,\, n \in \mathbb{N})$ be the simple symmetric random walk ($M_n = X_1 + \ldots + X_n$) and consider the following strategy:

$$H_0 = 0,\; H_1 = 1,\; H_{n+1} = \begin{cases} 2H_n, & \text{if } \xi_1 = \ldots = \xi_n = -1 \\ 0, & \text{otherwise} \end{cases}$$

Notice that all the H_n are bounded random variables. Then by the above proposition, the process G defined as

$$G_0 = 0, \quad G_n = \sum_{i=1}^n H_i\,(M_i - M_{i-1}) = \sum_{i=1}^n H_i\,X_i, \quad n \geq 1$$

is a martingale. So $\mathbb{E}(G_n) = \mathbb{E}(G_0) = 0$, $\forall n \in \mathbb{N}$. Let now

$$T = \inf\{n \geq 1 \,:\, X_n = +1\}$$

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 $T(\omega) \leq N$, $\forall \omega \in \Omega$.

2.6 Doob's decomposition theorem

Week 12

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 $\mathbb{E}(M_{n+1} - M_n|\mathcal{F}_n) = 0$. □

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

$$M_n = \mathbb{E}(X|\mathcal{F}_n), \quad n \in \mathbb{N}$$

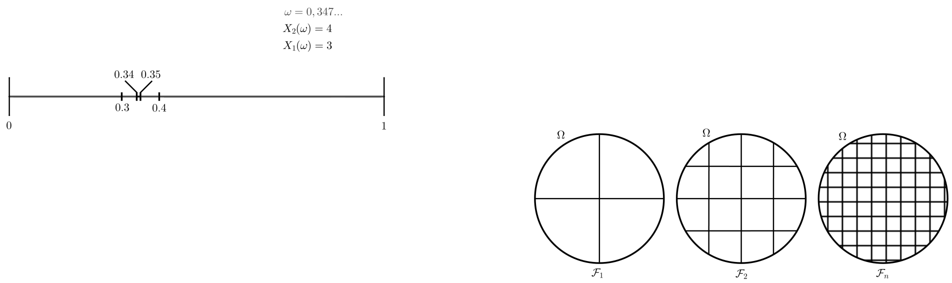
is a martingale with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$.

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 $n \in \mathbb{N}$. □

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, as illustrated on the left-hand side of the figure below: (please turn the page)



- On the right-hand side of the figure is another illustration of a Doob martingale: as time goes by, one gets more and more information about where to locate oneself in the space Ω.

- Are Doob's martingales a very particular type of martingales? No! As the following paragraph shows, there are quite many such martingales!

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

$$\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^2) < +\infty \tag{3}$$

there exists a limiting random variable M_∞ such that

(i) $M_n \underset{n\to\infty}{\to} M_\infty$ almost surely.

(ii) $\lim_{n\to\infty} \mathbb{E}\left((M_n - M_\infty)^2\right) = 0$ (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 (3) 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 (3) 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 (3) 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 (3) (leading therefore to the same conclusion) is the following:

$$\sup_{n\in\mathbb{N}} \sup_{\omega\in\Omega} |M_n(\omega)| < +\infty. \tag{4}$$

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_∞ such that the three conclusions of the theorem hold. In addition, it can be shown by contradiction that M_∞ 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 (3) 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

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

Proof. Simply replace N by ∞ in the proof of the first version and use the fact that M is a closed martingale by the convergence theorem. □

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 $-a - c \leq M_{T\wedge n}(\omega) \leq b + c$ for all $\omega \in \Omega$ and $n \in \mathbb{N}$. In particular,

$$\sup_{n\in\mathbb{N}} \mathbb{E}(M_{T\wedge n}^2) < +\infty$$

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).

3.4 Proof of the theorem

Week 13

A key ingredient for the proof: the maximal inequality. The following inequality, apart from being useful for the proof of the martingale convergence theorem, is interesting in itself. Let $(M_n,\, n \in \mathbb{N})$ be a square-integrable martingale. Then for every $N \in \mathbb{N}$ and $x > 0$,

$$\mathbb{P}\left(\left\{\max_{0 \leq n \leq N} |M_n| \geq x\right\}\right) \leq \frac{\mathbb{E}(M_N^2)}{x^2}$$

Remark. This inequality resembles Chebychev's inequality, but it is actually much stronger. In particular, note the remarkable fact that deviation probability of the maximum value of the martingale *over the whole time interval* $\{0, \ldots, N\}$ is controlled by the second moment of the martingale at the final instant N alone.

Proof. - First, let $x > 0$ and let $T_x = \inf\{n \in \mathbb{N} \,:\, |M_n| \geq x\}$: T_x is a stopping time and note that

$$\{T_x \leq N\} = \left\{\max_{0 \leq n \leq N} |M_n| \geq x\right\}$$

So what we need actually to prove is that $\mathbb{P}(\{T_x \leq N\}) \leq \frac{\mathbb{E}(M_N^2)}{x^2}$.

- Second, observe that as M is a martingale, M^2 is a submartingale. So by the optional stopping theorem, we obtain

$$\begin{aligned}
\mathbb{E}(M_N^2) &= \mathbb{E}\left(\mathbb{E}\left(M_N^2 | \mathcal{F}_{T_x \wedge N}\right)\right) \geq \mathbb{E}\left(M_{T_x \wedge N}^2\right) \geq \mathbb{E}\left(M_{T_x \wedge N}^2\, 1_{\{T_x \leq N\}}\right) \\
&= \mathbb{E}\left(M_{T_x}^2\, 1_{\{T_x \leq N\}}\right) \geq \mathbb{E}\left(x^2\, 1_{\{T_x \leq N\}}\right) = x^2\, \mathbb{P}(\{T_x \leq N\})
\end{aligned}$$

where the last inequality comes from the fact that $|M_{T_x}| \geq x$, by definition of T_x. This proves the claim. □

Proof of Theorem 3.2. - We first prove conclusion (i), namely that the sequence $(M_n,\, n \in \mathbb{N})$ converges almost surely to some limit. This proof is divided in two parts.

Part 1. We first show that for every $\varepsilon > 0$,

$$\mathbb{P}\left(\left\{\sup_{n \in \mathbb{N}} |M_{n+m} - M_m| \geq \varepsilon\right\}\right) \underset{m \to \infty}{\to} 0 \tag{5}$$

This is saying that for every $\varepsilon > 0$, the probability that the martingale M deviates by more than ε after a given time m can be made arbitrarily small by taking m large enough. This essentially says that the fluctuations of the martingale decay with time, i.e. that the martingale ultimately converges! Of course, this is just an intuition and needs a formal proof, which will be done in the second part of the proof. For now, let us focus on proving (5).

a) Let $m \in \mathbb{N}$ be fixed and define the process $(Y_n,\, n \in \mathbb{N})$ by $Y_n = M_{n+m} - M_m$, for $n \in \mathbb{N}$. Y is a square-integrable martingale, so by the maximal inequality, we have for every $N \in \mathbb{N}$ and every $\varepsilon > 0$:

$$\mathbb{P}\left(\left\{\max_{0 \leq n \leq N} |Y_n| \geq \varepsilon\right\}\right) \leq \frac{\mathbb{E}(Y_N^2)}{\varepsilon^2}$$

b) Let us now prove that

$$\mathbb{E}(Y_N^2) = \mathbb{E}(M_{m+N}^2) - \mathbb{E}(M_m^2).$$

This equality follows from the orthogonality of the increments of M. Here is a detailed proof:

$$\begin{aligned}
\mathbb{E}(Y_N^2) &= \mathbb{E}((M_{m+N} - M_m)^2) = \mathbb{E}(M_{m+N}^2) - 2\,\mathbb{E}(M_{m+N}\,M_m) + \mathbb{E}(M_m^2) \\
&= \mathbb{E}(M_{m+N}^2) - 2\,\mathbb{E}(\mathbb{E}(M_{m+N}\,M_m|\mathcal{F}_m)) + \mathbb{E}(M_m^2) \\
&= \mathbb{E}(M_{m+N}^2) - 2\,\mathbb{E}(\mathbb{E}(M_{m+N}|\mathcal{F}_m)\,M_m) + \mathbb{E}(M_m^2) \\
&= \mathbb{E}(M_{m+N}^2) - 2\,\mathbb{E}(M_m^2) + \mathbb{E}(M_m^2) = \mathbb{E}(M_{m+N}^2) - \mathbb{E}(M_m^2)
\end{aligned}$$

Gathering a) and b) together, we obtain for every $m, N \in \mathbb{N}$ and every $\varepsilon > 0$:

$$\mathbb{P}\left(\left\{\max_{0\le n\le N} |M_{m+n} - M_m| \ge \varepsilon\right\}\right) \le \frac{\mathbb{E}(M_{m+N}^2) - \mathbb{E}(M_m^2)}{\varepsilon^2}.$$

c) Assumption (3) states that $\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^2) < +\infty$. As the sequence $(\mathbb{E}(M_n^2),\ n \in \mathbb{N})$ is increasing (since M^2 is a submartingale), this also says that the sequence has a limit: $\lim_{n\to\infty} \mathbb{E}(M_n^2) = K < +\infty$. Therefore, for every $m \in \mathbb{N}$ and $\varepsilon > 0$, we obtain

$$\begin{aligned}
\mathbb{P}\left(\left\{\sup_{n\in\mathbb{N}} |M_{m+n} - M_m| \ge \varepsilon\right\}\right) &= \lim_{N\to\infty} \mathbb{P}\left(\left\{\max_{0\le n\le N} |M_{m+n} - M_m| \ge \varepsilon\right\}\right) \\
&\le \lim_{N\to\infty} \frac{\mathbb{E}(M_{m+N}^2) - \mathbb{E}(M_m^2)}{\varepsilon^2} = \frac{K - \mathbb{E}(M_m^2)}{\varepsilon^2}
\end{aligned}$$

Taking now m to infinity, we further obtain

$$\mathbb{P}\left(\left\{\sup_{n\in\mathbb{N}} |M_{m+n} - M_m| \ge \varepsilon\right\}\right) \le \frac{K - \mathbb{E}(M_m^2)}{\varepsilon^2} \underset{m\to\infty}{\to} \frac{K-K}{\varepsilon^2} = 0$$

for every $\varepsilon > 0$. This proves (5) and concludes therefore the first part of the proof.

Part 2. Let $C = \{\omega \in \Omega\ :\ \lim_{n\to\infty} M_n(\omega) \text{ exists}\}$. In this second part, we prove that $\mathbb{P}(C) = 1$, which is conclusion (i).

Here is what we have proven so far. For $m \in \mathbb{N}$ and $\varepsilon > 0$, define $A_m(\varepsilon) = \{\sup_{n\in\mathbb{N}} |M_{m+n} - M_m| \ge \varepsilon\}$. Then (5) says that for every fixed $\varepsilon > 0$, $\lim_{m\to\infty} \mathbb{P}(A_m(\varepsilon)) = 0$. We then have the following (long!) series of equivalent statements:

$$\begin{aligned}
&\forall\varepsilon > 0, \quad \lim_{m\to\infty} \mathbb{P}(A_m(\varepsilon)) = 0 \\
&\iff \forall\varepsilon > 0, \quad \mathbb{P}(\cap_{m\in\mathbb{N}} A_m(\varepsilon)) = 0 \iff \forall M \ge 1, \quad \mathbb{P}(\cap_{m\in\mathbb{N}} A_m(\tfrac{1}{M})) = 0 \\
&\iff \mathbb{P}(\cup_{M\ge1} \cap_{m\in\mathbb{N}} A_m(\tfrac{1}{M})) = 0 \iff \mathbb{P}(\cup_{\varepsilon>0} \cap_{m\in\mathbb{N}} A_m(\varepsilon)) = 0 \\
&\iff \mathbb{P}(\{\exists\varepsilon > 0 \text{ s.t. } \forall m \in \mathbb{N},\ \sup_{n\in\mathbb{N}} |M_{m+n} - M_m| \ge \varepsilon\}) = 0 \\
&\iff \mathbb{P}(\{\forall\varepsilon > 0,\ \exists m \in \mathbb{N} \text{ s.t. } \sup_{n\in\mathbb{N}} |M_{m+n} - M_m| < \varepsilon\}) = 1 \\
&\iff \mathbb{P}(\{\forall\varepsilon > 0,\ \exists m \in \mathbb{N} \text{ s.t. } |M_{m+n} - M_m| < \varepsilon,\ \forall n \in \mathbb{N}\}) = 1 \\
&\iff \mathbb{P}(\{\forall\varepsilon > 0,\ \exists m \in \mathbb{N} \text{ s.t. } |M_{m+n} - M_{m+p}| < \varepsilon,\ \forall n, p \in \mathbb{N}\}) = 1 \\
&\iff \mathbb{P}(\{\text{the sequence } (M_n,\ n \in \mathbb{N}) \text{ is a Cauchy sequence}\}) = 1 \iff \mathbb{P}(C) = 1
\end{aligned}$$

as every Cauchy sequence in $\mathbb{R}$ converges. This completes the proof of conclusion (i) in the theorem.

- In order to prove conclusion (ii) (quadratic convergence), let us recall that from what was shown above

$$\mathbb{E}((M_n - M_m)^2) = \mathbb{E}(M_n^2) - \mathbb{E}(M_m^2), \quad \forall n \ge m \ge 0$$

This, together with the fact that $\lim_{n\to\infty} \mathbb{E}(M_n^2) = K$, implies that M_n is a Cauchy sequence in L^2: it therefore converges to some limit, as the space of square-integrable random variables is complete. Let us call this limit $\widetilde{M}_\infty$. But does it hold that $\widetilde{M}_\infty = M_\infty$, the a.s. limit of part (i)? Yes, as both quadratic convergence and a.s. convergence imply convergence in probability, and we have seen in part I (Theorem 5.3) that if a sequence of random variables converges in probability to two possible limits, then these two limits are equal almost surely.

- Conclusion (iii) then follows from the following reasoning. We need to prove that $M_n = \mathbb{E}(M_\infty|\mathcal{F}_n)$ for every (fixed) $n \in \mathbb{N}$ (where M_∞ is the limit found in parts (i) and (ii)). To this end, let us go back to the very definition of conditional expectation and simply check that

(i) M_n is $\mathcal{F}_n$-measurable: this is by definition.

(ii) $\mathbb{E}(M_\infty U) = \mathbb{E}(M_n U)$ for every random variable U $\mathcal{F}_n$-measurable and bounded. This follows from the following observation:

$$\mathbb{E}(M_n U) = \mathbb{E}(M_N U), \quad \forall N \geq n$$

This equality together with the Cauchy-Schwarz inequality imply that for every $N \geq n$:

$$|\mathbb{E}(M_\infty U) - \mathbb{E}(M_n U)| = |\mathbb{E}(M_\infty U) - \mathbb{E}(M_N U)| = |\mathbb{E}((M_\infty - M_N)\, U)| \leq \sqrt{\mathbb{E}((M_\infty - M_N)^2)}\,\sqrt{\mathbb{E}(U^2)} \underset{N\to\infty}{\to} 0$$

by quadratic convergence (conclusion (ii)). So we obtain that necessarily, $\mathbb{E}(M_\infty U) = \mathbb{E}(M_n U)$ (remember that n is fixed here). This completes the proof of Theorem 3.2. □

3.5 The martingale convergence theorem: second version

Theorem 3.4. Let $(M_n,\, n \in \mathbb{N})$ be a martingale such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}(|M_n|) < +\infty \tag{6}$$

Then there exists a limiting random variable M_∞ such that $M_n \underset{n\to\infty}{\to} M_\infty$ almost surely.

We shall not go through the proof of this second version of the martingale convergence theorem[1], whose order of difficulty resembles that of the first one. Let us just make a few remarks and also exhibit an interesting example below.

Remarks. - Contrary to what one could perhaps expect, it does not necessarily hold in this case that $\lim_{n\to\infty} \mathbb{E}(|M_n - M_\infty|) = 0$, nor that $\mathbb{E}(M_\infty|\mathcal{F}_n) = M_n$ for every $n \in \mathbb{N}$.

- By the Cauchy-Schwarz inequality, we see that condition (6) is weaker than condition (3).

- On the other hand, condition (6) is of course stronger than just asking $\mathbb{E}(|M_n|) < +\infty$ for all $n \in \mathbb{N}$ (this last condition is by the way satisfied by every martingale, by definition). It is also stronger than asking $\sup_{n\in\mathbb{N}} \mathbb{E}(M_n) < +\infty$. Why? Simply because for every martingale, $\mathbb{E}(M_n) = \mathbb{E}(M_0)$ for every $n \in \mathbb{N}$, so the supremum is always finite! The same does not hold when one adds absolute values: the process $(|M_n|,\, n \in \mathbb{N})$ is a submartingale, so the sequence $(\mathbb{E}(|M_n|),\, n \in \mathbb{N})$ is non-decreasing, possibly growing to infinity.

- If M is a non-negative martingale, then $|M_n| = M_n$ for every $n \in \mathbb{N}$ and by what was just said above, condition (6) is satisfied! So non-negative martingales *always* converge to a limit almost surely! But they might not be closed at infinity.

A puzzling example. Let $(S_n,\, n \in \mathbb{N})$ be the simple symmetric random walk and $(M_n,\, n \in \mathbb{N})$ be the process defined as

$$M_n = \exp(S_n - cn), \quad n \in \mathbb{N}$$

where $c = \log\left(\frac{e+e^{-1}}{2}\right) > 0$ is such that M is a martingale, with $\mathbb{E}(M_n) = \mathbb{E}(M_0) = 1$ for every $n \in \mathbb{N}$. On top of that, M is a positive martingale, so by the previous remark, there exists a random variable

[1] It is sometimes called the first version in the literature!

M_∞ such that $M_n \underset{n\to\infty}{\to} M_\infty$ almost surely. So far so good. Let us now consider some more puzzling facts:

- A simple computation shows that $\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^2) = \sup_{n\in\mathbb{N}} \mathbb{E}(\exp(2S_n - 2cn)) = +\infty$, so we cannot conclude that (ii) and (iii) in Theorem 3.2 hold. Actually, these conclusions do not hold, as we will see below.

- What can the random variable M_∞ be? It can be shown that $S_n - cn \underset{n\to\infty}{\to} -\infty$ almost surely, from which we deduce that $M_n = \exp(S_n - cn) \underset{n\to\infty}{\to} 0$ almost surely, i.e. $M_\infty = 0$!

- It is therefore impossible that $\mathbb{E}(M_\infty|\mathcal{F}_n) = M_n$, as the left-hand side is 0, while the right-hand side is not. Likewise, quadratic convergence to 0 does not hold (this would mean that $\lim_{n\to\infty} \mathbb{E}(M_n^2) = 0$, which does not hold).

- On the contrary, we just said above that $\mathrm{Var}(M_n) = \mathbb{E}(M_n^2) - (\mathbb{E}(M_n))^2 = \mathbb{E}(M_n^2) - 1$ grows to infinity as n goes to infinity. Still, M_n converges to 0 almost surely. If this sounds puzzling to you, be reassured that you are not alone!

3.6 Generalization to sub- and supermartingales

We state below the generalization of the two convergence theorems to sub- and supermartingales.

Theorem 3.5. (Generalization of Theorem 3.2)
Let $(M_n,\, n \in \mathbb{N})$ be a square-integrable submartingale (resp., supermartingale) with respect to a filtration $(\mathcal{F}_n,\, n \in \mathbb{N})$. Under the additional assumption that

$$\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^2) < +\infty \tag{7}$$

there exists a limiting random variable M_∞ such that

(i) $M_n \underset{n\to\infty}{\to} M_\infty$ almost surely.

(ii) $\lim_{n\to\infty} \mathbb{E}\left((M_n - M_\infty)^2\right) = 0$ (quadratic convergence).

(iii) $M_n \le \mathbb{E}(M_\infty|\mathcal{F}_n)$ (resp., $M_n \ge \mathbb{E}(M_\infty|\mathcal{F}_n)$), for all $n \in \mathbb{N}$.

Theorem 3.6. (Generalization of Theorem 3.4)
Let $(M_n,\, n \in \mathbb{N})$ be a submartingale (resp., supermartingale) such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}(M_n^+) < +\infty \quad \text{(resp., } \sup_{n\in\mathbb{N}} \mathbb{E}(M_n^-) < +\infty) \tag{8}$$

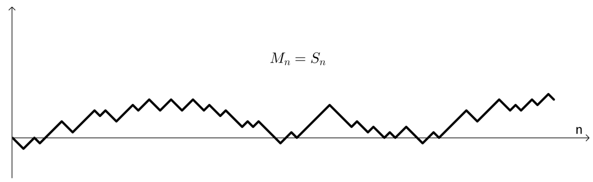
where recall here that $M_n+ = \max(M_n, 0)$ and $M_n^- = \max(-M_n, 0)$. Then there exists a limiting random variable M_∞ such that $M_n \underset{n\to\infty}{\to} M_\infty$ almost surely.

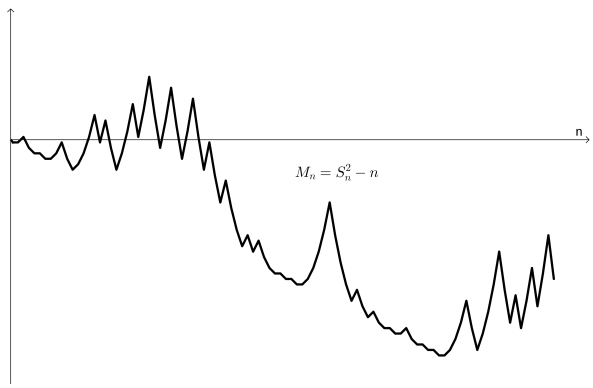
As one can see, not much changes in the assumptions and conclusions of both theorems! Let us mention some interesting consequences.

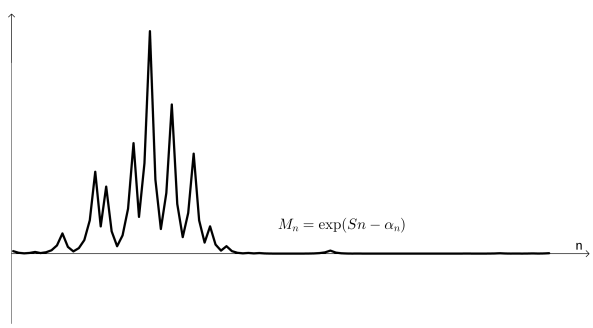
- From Theorem 3.5, it holds that if M is a sub- or a supermartingale satisfying condition (7), then M_n converges both almost surely and quadratically to some limit M_∞. In case where M is a (non-trivial) martingale, we saw previously that the limit M_∞ cannot be equal to 0, as this would lead to a contradiction, because of the third part of the conclusion stating that $M_n = \mathbb{E}(M_\infty|\mathcal{F}_n) = 0$ for all n. In the case of a sub- or supermartingale, this third part only says that $M_n \le \mathbb{E}(M_\infty|\mathcal{F}_n) = 0$ or $M_n \ge \mathbb{E}(M_\infty|\mathcal{F}_n) = 0$, which is not necessarily a contradiction.

- From Theorem 3.6, one deduces that any positive supermartingale admits an almost sure limit at infinity. But the same conclusion cannot be drawn for a positive submartingale (think simply of $M_n = n$: this very particular positive submartingale does not converge). From the same theorem, one deduces also that any *negative* submartingale admits an almost sure limit at infinity.

Finally, for illustration purposes, here are again below 3 well known martingales: $(S_n,\, n \in \mathbb{N}0)$, $(S_n^2 - n,\, n \in \mathbb{N})$ and $(M_n = \exp(S_n - cn,\, n \in \mathbb{N})$ juste seen above:







We see again here that even though theses 3 processes are all contant mean processes, they do exhibit very different behaviours!

3.7 Azuma's and McDiarmid's inequalities

Theorem 3.7. (Azuma's inequality)
Let $(M_n,\, n \in \mathbb{N})$ be a martingale such that $|M_n(\omega) - M_{n-1}(\omega)| \leq 1$ for every $n \geq 1$ and $\omega \in \Omega$. Such a martingale is said to have *bounded differences*. Assume also that M_0 is constant. Then for every $n \geq 1$ and $t > 0$, we have

$$\mathbb{P}(\{|M_n - M_0| \geq nt\}) \leq 2 \exp\left(-\frac{nt^2}{2}\right)$$

Remark. This statement resembles that of Hoeffding's inequality! The difference here is that a martingale is not necessarily a sum of i.i.d. random variables.

Proof. Let $X_n = M_n - M_{n-1}$ for $n \geq 1$. Then, by the assumptions made, $M_n - M_0 = \sum_{j=1}^n X_j$, with $|X_j(\omega)| \leq 1$ for every $j \geq 1$ and $\omega \in \Omega$, but as mentioned above, the X_j's are not necessarily i.i.d.: we only know that $\mathbb{E}(X_j|\mathcal{F}_{j-1}) = 0$ for every $j \geq 1$. We need to bound

$$\mathbb{P}\left(\left\{\left|\sum_{j=1}^n X_j\right| \geq nt\right\}\right) = \mathbb{P}\left(\left\{\sum_{j=1}^n X_j \geq nt\right\}\right) + \mathbb{P}\left(\left\{\sum_{j=1}^n X_j \leq -nt\right\}\right)$$

By Chebyshev's inequality with $\varphi(x) = e^{sx}$ and $s > 0$, we obtain

$$\begin{aligned}\mathbb{P}\left(\left\{\sum_{j=1}^n X_j \geq nt\right\}\right) \leq \frac{\mathbb{E}\left(\exp\left(s\sum_{j=1}^n X_j\right)\right)}{\exp(snt)} &= e^{-snt}\,\mathbb{E}\left(\mathbb{E}\left(\exp\left(s\sum_{j=1}^n X_j\right)\big|\mathcal{F}_{n-1}\right)\right) \\ &= e^{-snt}\,\mathbb{E}\left(\exp\left(s\sum_{j=1}^{n-1} X_j\right)\mathbb{E}\left(\exp\left(sX_n\right)\big|\mathcal{F}_{n-1}\right)\right)\end{aligned}$$

As $\mathbb{E}(X_n|\mathcal{F}_{n-1}) = 0$ and $|X_n(\omega)| \leq 1$ for every $\omega \in \Omega$, we can apply the same lemma as in the proof of Hoeffding's inequality to conclude that

$$\mathbb{E}(\exp(sX_n)|\mathcal{F}_{n-1}) \leq e^{s^2/2}$$

So

$$\mathbb{P}\left(\left\{\sum_{j=1}^n X_j \geq nt\right\}\right) \leq e^{-snt}\,\mathbb{E}\left(\exp\left(s\sum_{j=1}^{n-1} X_j\right)\right)\, e^{s^2/2}$$

and working backwards, we finally obtain the upper bound

$$\mathbb{P}\left(\left\{\sum_{j=1}^n X_j \geq nt\right\}\right) \leq e^{-snt+ns^2/2}$$

which is again minimum for $s^* = t$ and equal then to $\exp(-nt^2/2)$. By symmetry, the same bound is obtained for the other term:

$$\mathbb{P}\left(\left\{\sum_{j=1}^n X_j \leq -nt\right\}\right) \leq \exp(-nt^2/2)$$

which completes the proof. □

Generalization. Exactly like Hoeffding's inequality, Azuma's inequality can be generalized as follows. Let M be a martingale such that $M_n(\omega) - M_{n-1}(\Omega) \in [a_n, b_n]$ for every $n \geq 1$ and every $\omega \in \Omega$. Then

$$\mathbb{P}(\{|M_n - M_0| \geq nt\}) \leq 2\,\exp\left(-\frac{2n^2t^2}{\sum_{j=1}^n (b_j - a_j)^2}\right)$$

Application 1. Consider the martingale transform of Section 2.5 defined as follows. Let $(X_n,\, n \geq 1)$ be a sequence of i.i.d. random variables such that $\mathbb{P}(\{X_1 = +1\}) = \mathbb{P}(\{X_1 = -1\}) = \frac{1}{2}$. Let $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_n = \sigma(X_1, \ldots, X_n)$ for $n \geq 1$. Let $(H_n,\, n \in \mathbb{N})$ be a predictable process with respect to $(\mathcal{F}_n,\, n \in \mathbb{N})$ such that $|H_n(\omega)| \leq K_n$ for every $n \in \mathbb{N}$ and $\omega \in \Omega$. Let finally $G_0 = 0$, $G_n = \sum_{j=1}^n H_j X_j$, $n \geq 1$. Then

$$\mathbb{P}(\{|G_n - G_0| \geq nt\}) \leq 2\,\exp\left(-\frac{n^2t^2}{2\sum_{j=1}^n K_j^2}\right)$$

In the case where $K_n = K$ for every $n \in \mathbb{N}$, this says that

$$\mathbb{P}(\{|G_n - G_0| \geq nt\}) \leq 2\,\exp\left(-\frac{nt^2}{2K^2}\right)$$

We had obtained the same conclusion earlier for the random walk, but here, the increments of G are in general far from being independent.

Application 2: McDiarmid's inequality. Let $n \geq 1$ be fixed, let $X_1, \ldots, X_n$ be i.i.d. random variables and let $F : \mathbb{R}^n \to \mathbb{R}$ be a Borel-measurable function such that

$$|F(x_1, \ldots, x_j, \ldots, x_n) - F(x_1, \ldots, x_j', \ldots, x_n)| \leq K_j, \quad \forall x_1, \ldots, x_j, x_j', \ldots, x_n \in \mathbb{R},\ 1 \leq j \leq n$$

Then

$$\mathbb{P}(\{|F(X_1, \ldots, X_n) - \mathbb{E}(F(X_1, \ldots, X_n))| \geq nt\}) \leq 2 \exp\left(-\frac{n^2 t^2}{2 \sum_{j=1}^n K_j^2}\right)$$

Proof. Define $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_m = \sigma(X_1, \ldots, X_j)$ and $M_j = \mathbb{E}(F(X_1, \ldots, X_n)|\mathcal{F}_j)$ for $j \in \{0, \ldots, n\}$. By definition, M is a martingale and observe that

$$M_n = F(X_1, \ldots, X_n) \quad \text{and} \quad M_0 = \mathbb{E}(F(X_1, \ldots, X_n))$$

Moreover,

$$|M_j - M_{j-1}| = |\mathbb{E}(F(X_1, \ldots, X_n)|\mathcal{F}_j) - \mathbb{E}(F(X_1, \ldots, X_n)|\mathcal{F}_{j-1})| = |g(X_1, \ldots, X_j) - h(X_1, \ldots, X_{j-1})|$$

where $g(x_1, \ldots, x_j) = \mathbb{E}(f(x_1, \ldots, x_j, X_{j+1}, \ldots, X_n))$ and $h(x_1, \ldots, x_{j-1}) = \mathbb{E}(f(x_1, \ldots, x_{j-1}, X_j, \ldots, X_n))$. By the assumption made, we find that for every $x_1, \ldots, x_j \in \mathbb{R}$,

$$|g(x_1, \ldots, x_j) - h(x_1, \ldots, x_{j-1})| \leq \mathbb{E}(|f(x_1, \ldots, x_j, X_{j+1}, \ldots, X_n) - f(x_1, \ldots, x_{j-1}, X_j, \ldots, X_n)|) \leq K_j$$

so Azuma's inequality applies. This completes the proof. □