Probability II – Lecture notes

Gaultier Lambert
University of Zurich

October 14, 2020

Contents

1 Basics from Analysis 3 and Probability 1

The goal of this section is to review basic facts from Analysis 3 and Probability 1 that we will use for the rest of the course.

If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we define for $p \in [1, \infty)$,

$$L^p = \{X : \Omega \to \mathbb{R}, \mathcal{F}\text{-measurable}, \mathbb{E}[|X|^p] < \infty\}$$

Up to equivalence classes, L^p is a Banach space with norm $\|X\|_{L^p} = \mathbb{E}[|X|^p]^{1/p}$. We also define

$$\|X\|_{L^\infty} = \inf\{C > 0 : \mathbb{P}[|X| \leq C] = 1\}$$

and

$$L^\infty = \{X : \Omega \to \mathbb{R}, \mathcal{F}\text{-measurable}, \|X\|_{L^\infty} < \infty\}.$$

Up to equivalence classes, L^∞ is also a Banach space with norm $\|X\|_{L^\infty}$.

The following results will be relevant in this course. If $(\alpha_k) \in \mathbb{R}$ and $(A_k) \in \mathcal{F}$ are disjoint, then $X_n = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$ is called **a simple function**.

Theorem 1.1 (Density)**.**
1) *If $X \geq 0$, there exists a sequence of simple functions $X_n \geq 0$ such that $X_n \nearrow X$ as $n \to \infty$.*
2) *Simple functions are dense in L^p for any $p \in [1, \infty]$.*
3) *$C_c^\infty(\mathbb{R})$ is dense in $L^p(\mathbb{R})$ for any $p \in [1, \infty)$.*

Proposition 1.2.
1) *$L^q \subset L^p$ for any $p, q \in [1, \infty]$ with $p < q$.*
2) *Hölder's inequality: For any $p, q \in [1, \infty]$ with $1/p + 1/q = 1$,*

$$\mathbb{E}[XY] \leq \|X\|_p \|Y\|_q.$$

3) *Moreover, it holds* $\|X\|_{L^p} = \sup\{\mathbb{E}[|X|Y] : Y \geq 0, \|Y\|_{L^q} = 1\}$.

Remark 1.3. *If $p \in [1, \infty)$, the* sup *in 3) is attained for $Y = |X|^{p/q}$. This implies that if $q > 1$ is given by $1/p + 1/q = 1$, then L^q is dual of L^p.*

Here is another fundamental inequality in probability theory.

Proposition 1.4. *Jensen's inequality: If $\varphi : \mathbb{R} \to \mathbb{R}$ is a convex function and $X \in L^1$, then*

$$\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)].$$

Let us now review the two fundamental results on convergence of independent random variables.

Theorem 1.5 (Kolmogorov's Law of large numbers)**.** *Let (X_n) be a sequence of i.i.d. random variables in L^1 with mean μ. Then it holds almost surely and in L^1,*

$$\lim_{N\to\infty} \frac{X_1 + \cdots + X_N}{N} = \mu.$$

One particular goal of this course is to give a proof of this result and to generalize it in the case where the random variables (X_n) are not necessarily independent (see Section **??**).

The fluctuations of this sum around its mean are of order $1/\sqrt{N}$ and asymptotically Gaussian.

Theorem 1.6 (Central limit theorem)**.** *Let (X_n) be a sequence of independent random variables in L^2 with $\mathbb{E}[X_n] = 0$ and $\mathbb{E}[X_n^2] = \sigma_n^2$. Let $\Sigma_N = \sqrt{\sum_{n=1}^N \sigma_n^2}$ and let us suppose that for any $\epsilon > 0$,*

$$\lim_{N\to\infty} \frac{1}{\Sigma_N^2} \sum_{n=1}^N \mathbb{E}\left[X_n^2 \mathbf{1}_{|X_n|>\epsilon\Sigma_N}\right] = 0. \tag{1.1}$$

Then $\frac{X_1+\cdots+X_N}{\Sigma_N} \Rightarrow \mathcal{N}_{0,1}$ as $N \to \infty$.

(1.1) is known as the Lindeberg's condition.

Proof by Lindeberg's replacement strategy. Let $f \in C_c^\infty(\mathbb{R})$. By Taylor's theorem, it holds for any $\epsilon > 0$ and all $x, y \in \mathbb{R}$,

$$|f(x+y) - f(x) - f'(x)y - f''(x)y^2/2| \le R_\epsilon(y) := C\big(\epsilon\mathbf{1}_{|y|\le\epsilon} + \mathbf{1}_{|y|>\epsilon}\big)y^2, \tag{1.2}$$

where $C = \|f''\|_\infty + \|f'''\|_\infty/6$ depends only on the test function f.

Let (Z_n) be a sequence of independent Gaussian random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ which is also independent from (X_n), with $Z_n \sim \mathcal{N}_{0,\sigma_n^2}$. Let us denote for any $N \in \mathbb{N}$ and $n \le N$, $X_{N,n} := X_n/\Sigma_N$ and $Z_{N,n} := Z_n/\Sigma_N$. The strategy is to find out the distribution of $\frac{X_1+\cdots+X_N}{\Sigma_N} = X_{N,1} + \cdots + X_{N,N}$ by replacing the random variables $X_{N,k}$ by $Z_{N,k}$ successively. Let us also denote

$$S_N^{(k)} = Z_{N,1} + \cdots + Z_{N,k} + X_{N,k+1} + \cdots + X_{N,N}, \qquad \text{for} \quad k \in \{0, \ldots, N\}.$$

In particular, we have

$$S_N^{(0)} = X_{N,1} + \cdots + X_{N,N} = \tfrac{X_1+\cdots+X_N}{\Sigma_N} \quad \text{and} \quad S_N^{(N)} = Z_{N,1} + \cdots + Z_{N,N}. \tag{1.3}$$

Fix $k \in \{1, \ldots, N\}$, by applying (1.2) with $x = Z_{N,1} + \cdots + Z_{N,k-1} + X_{N,k+1} + \cdots + X_{N,N}$ and $y = X_{N,k}$, we obtain

$$|f(S_N^{(k-1)}) - f(x) - f'(x)X_{N,k} - f''(x)X_{N,k}^2/2| \le R_\epsilon(X_{N,k}).$$

By taking expectations on both sides, this implies that

$$|\mathbb{E}[f(S_N^{(k-1)})] - \mathbb{E}\left[f(x)\right] - \mathbb{E}\left[f''(x)\right]\mathbb{E}\left[X_{N,k}^2\right]/2| \leq \mathbb{E}\left[R_\epsilon(X_{N,k})\right], \tag{1.4}$$

where we used that x is independent from $y = X_{N,k}$ and that $\mathbb{E}\left[X_{N,k}\right] = 0$.
Similarly, with $x = Z_{N,1} + \cdots + Z_{N,k-1} + X_{N,k+1} + \cdots + X_{N,N}$ and $y = Z_{N,k}$,

$$|\mathbb{E}[f(S_N^{(k)})] - \mathbb{E}\left[f(x)\right] - \mathbb{E}\left[f''(x)\right]\mathbb{E}\left[Z_{N,k}^2\right]/2| \leq \mathbb{E}\left[R_\epsilon(Z_{N,k})\right]. \tag{1.5}$$

By combining the bounds (1.4)–(1.5) and using the triangle inequality, since $\mathbb{E}\left[X_{N,k}^2\right] = \mathbb{E}\left[Z_{N,k}^2\right]$ by construction, we obtain

$$|\mathbb{E}[f(S_N^{(k)})] - \mathbb{E}[f(S_N^{(k-1)})]| \leq \mathbb{E}\left[R_\epsilon(X_{N,k})\right] + \mathbb{E}\left[R_\epsilon(Z_{N,k})\right].$$

By summing the previous estimates for $k = 1, \ldots, N$ and using that the terms are telescoping, this implies that

$$|\mathbb{E}[f(S_N^{(N)}) - f(S_N^{(0)})]| \leq \sum_{k=1}^{N} |\mathbb{E}[f(S_N^{(k)}) - \mathbb{E}[f(S_N^{(k-1)})| \leq \sum_{k=1}^{N} \mathbb{E}\left[R_\epsilon(X_{N,k})\right] + \sum_{k=1}^{N} \mathbb{E}\left[R_\epsilon(Z_{N,k})\right]. \tag{1.6}$$

Note that by definitions, $\sum_{k=1}^{N} \mathbb{E}\left[X_{N,k}^2\right] = 1$ and the error can be controlled by

$$\begin{aligned} \sum_{k=1}^{N} \mathbb{E}\left[R_\epsilon(X_{N,k})\right] &= C\epsilon \sum_{k=1}^{N} \mathbb{E}\left[X_{N,k}^2 \mathbf{1}_{|X_{N,k}| \leq \epsilon}\right] + C \sum_{k=1}^{N} \mathbb{E}\left[X_{N,k}^2 \mathbf{1}_{|X_{N,k}| > \epsilon}\right] \\ &\leq C\epsilon + C \textstyle\sum_{k=1}^{N} \mathbb{E}\left[X_{N,k}^2 \mathbf{1}_{|X_{N,k}| > \epsilon}\right] \end{aligned}$$

Now, using the condition (1.1), we verify that for any $\epsilon > 0$,

$$\limsup_{N \to \infty} \textstyle\sum_{k=1}^{N} \mathbb{E}\left[R_\epsilon(X_{N,k})\right] \leq C\epsilon.$$

We want to prove a similar claim for the Gaussian random variables $Z_{N,k}$. This is the goal of the next exercice.

Exercise 1.1. 1) Show that the condition (1.1) implies that $\lim_{N \to \infty} \max_{n \leq N} \left(\frac{\sigma_n^2}{\Sigma_N^2}\right) = 0$.
2) Deduce that the sequence $(Z_n)_{n \geq 1}$ also satisfies Lindeberg's condition (1.1).
3) Show that $Z_{N,1} + \cdots + Z_{N,n} \sim \mathcal{N}_{0,1}$.

Exactly as above, 2) implies that $\limsup_{N\to\infty} \sum_{k=1}^N \mathbb{E}\left[R_\epsilon(Z_{N,k})\right] \le C\epsilon$. Hence, by (1.6), we conclude that

$$\limsup_{N\to\infty} |\mathbb{E}[f(S_N^{(N)}) - f(S_N^{(0)})]| \le 2C\epsilon.$$

Since the parameter ϵ is arbitrarily small, this shows that $\lim_{N\to\infty} \mathbb{E}[f(S_N^{(N)}) - f(S_N^{(0)})] = 0$.

To finish the proof, we use (1.3) and note that by 3), $\mathbb{E}[f(S_N^{(N)})] = \mathbb{E}[f(\mathcal{N}_{0,1})]$, so we have shown that

$$\lim_{N\to\infty} \mathbb{E}[f(\tfrac{X_1+\cdots+X_N}{\Sigma_N})] = \mathbb{E}[f(\mathcal{N}_{0,1})]. \tag{1.7}$$

Since (1.7) holds for any test function $f \in C_c^\infty(\mathbb{R})$, we conclude that $\frac{X_1+\cdots+X_N}{\Sigma_N} \Rightarrow \mathcal{N}_{0,1}$. □

2 Conditional expectation

2.1 L^2 theory

$L^2(\Omega, \mathcal{F}, \mathbb{P})$ is a **Hilbert space** with inner product $X, Y \mapsto \mathbb{E}[XY]$. The following result from analysis will be relevant for our construction of conditional expectation.

Theorem 2.1 (Orthogonal Projection)**.** *Let $\mathcal{H}$ be a Hilbert space and $\mathcal{L} \subset \mathcal{H}$ be a (closed) subspace. We define $x \in \mathcal{H} \mapsto \mathrm{d}(x, \mathcal{L}) = \inf\{\|x-y\| : y \in \mathcal{L}\}$. For any $x \in \mathcal{H}$, there exists a* **unique** *$x_{\mathcal{L}} \in \mathcal{L}$ such that*

$$\|x - x_{\mathcal{L}}\| = \mathrm{d}(x, \mathcal{L}).$$

$x_{\mathcal{L}}$ is called **the (orthogonal) projection** *of x on $\mathcal{L}$ and it is characterized by the following property: $z \in \mathcal{L}$ satisfies*

$$\langle z, y\rangle = \langle x, y\rangle, \qquad \forall y \in \mathcal{L}, \tag{2.1}$$

if and only if $z = x_{\mathcal{L}}$.

Proof. We are going to use the (parallelogram) identity: for any $y, y' \in \mathcal{H}$,

$$\|y - y'\|^2 + \|y + y'\|^2 = 2(\|y\|^2 + \|y'\|^2). \tag{2.2}$$

By definition, there exists a sequence $y_n \in \mathcal{L}$ such that $\|y_n - x\| \to \mathrm{d}(x, \mathcal{L})$ as $n \to \infty$. Let us show that (y_n) is a Cauchy sequence. By (2.2), we obtain

$$\|y_n - y_m\|^2 + \|2x - y_n - y_m\|^2 = 2\|x - y_n\|^2 + 2\|x - y_m\|^2.$$

Using that $\|2x - y_n - y_m\| = 2\|x - \frac{y_n+y_m}{2}\| \ge 2\,\mathrm{d}(x, \mathcal{L})$, we obtain

$$\|y_n - y_m\|^2 \le \epsilon_n + \epsilon_m \qquad \text{where} \quad \epsilon_n = 2\|x - y_n\|^2 - 2\,\mathrm{d}(x, \mathcal{L})^2.$$

As $\epsilon_n \to 0$ as $n \to \infty$, this shows that (y_n) is a Cauchy sequence and, as $\mathcal{H}$ is complete, $y_n \to x_{\mathcal{L}}$ in $\mathcal{H}$. Since $\mathcal{L}$ is closed, we also have $x_{\mathcal{L}} \in \mathcal{L}$ and by continuity,

$$\|x_{\mathcal{L}} - x\| = \lim_{n\to\infty} \|y_n - x\| = \mathrm{d}(x, \mathcal{L}).$$

The Cauchy property also implies that this limit $x_{\mathcal{L}} \in \mathcal{L}$ is unique.

To prove that the projection $x_{\mathcal{L}}$ satisfies (2.1), fix $y \in \mathcal{L}$ and consider the map:

$$\lambda \in \mathbb{R} \mapsto \|x - x_{\mathcal{L}} + \lambda y\|^2 = \|x - x_{\mathcal{L}}\|^2 + 2\lambda\langle x - x_{\mathcal{L}}, y\rangle + \lambda^2\|y\|^2.$$

This map is a polynomial of degree 2 and (by definition of $x_{\mathcal{L}}$) its unique minimum is attained at $\lambda = 0$. Hence, we have

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\|x - x_{\mathcal{L}} - \lambda y\|^2\bigg|_{\lambda=0} = 2\langle x - x_{\mathcal{L}}, y\rangle = 0.$$

Since it is true for any $y \in \mathcal{L}$, this shows that $(x - x_{\mathcal{L}}) \in \mathcal{L}^{\perp}$ (equivalently, $x_{\mathcal{L}}$ satisfies (2.1)). To conclude the proof, it remains to show that the equation (2.1) has a unique solution. If $z \in \mathcal{L}$ satisfies (2.1), then for any $y \in \mathcal{L}$,

$$\|x - y\|^2 = \|x - z\|^2 + \|z - y\|^2 + 2\underbrace{\langle x - z, y - z\rangle}_{=0 \text{ since } (y-z)\in\mathcal{L}} \geq \|x - z\|^2.$$

Now, taking the infimum over all $y \in \mathcal{L}$ on the LHS, this shows that

$$\mathrm{d}(x, \mathcal{L}) = \|x - z\|.$$

This computation implies that $z = x_{\mathcal{L}}$ is the (orthogonal) projection of x on $\mathcal{L}$. □

Corollary 2.2. *For any (closed) subspace $\mathcal{L} \subset \mathcal{H}$, the map $\pi_{\mathcal{L}} : x \in \mathcal{H} \mapsto x_{\mathcal{L}} \in \mathcal{L}$ is linear.*

Proof. For $\alpha \in \mathbb{R}$ and $x, x' \in \mathcal{H}$, we have to show that $\pi_{\mathcal{L}}(\alpha x + x') = \alpha x_{\mathcal{L}} + x'_{\mathcal{L}}$. By Theorem 2.1, it suffices to check that this vector satisfies (2.1). This follows directly by bi-linearity of the inner product on $\mathcal{H}$. □

Remark 2.3. *The linear map $\pi_{\mathcal{L}}$ is an orthogonal projection in the sense that $\pi_{\mathcal{L}}|_{\mathcal{L}} = \mathrm{I}$ and that $\mathrm{Ker}(\pi_{\mathcal{L}}) = 0$. Moreover, it is immediate to check that if $\mathcal{L}' \subset \mathcal{L}$ is a (closed) subspace, then $\pi_{\mathcal{L}'}\pi_{\mathcal{L}} = \pi_{\mathcal{L}'}$ (this is the* 'tower property'*).*

Let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. By definition, $\mathcal{L} = L^2(\Omega, \mathcal{A}, \mathbb{P})$ is a closed linear subspace of $\mathcal{H} = L^2(\Omega, \mathcal{F}, \mathbb{P})$ and we denote by $\mathbb{E}[\cdot|\mathcal{A}]$ the (linear) projection $\mathcal{H} \to \mathcal{L}$.

Then, for any $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, this <u>defines</u> a random variable $\mathbb{E}[X|\mathcal{A}] \in L^2(\Omega, \mathcal{A}, \mathbb{P})$. Moreover by Theorem 2.1, an $\mathcal{A}$-measurable random variable Z satisfies

$$\mathbb{E}[ZY] = \mathbb{E}[XY], \qquad \forall Y \in L^2(\Omega, \mathcal{A}, \mathbb{P}), \tag{2.3}$$

if and only if $Z = \mathbb{E}[X|\mathcal{A}]$.

The random variable $\mathbb{E}[X|\mathcal{A}]$ is called the **conditional expectation of X knowing $\mathcal{A}$**. The conditional expectation has the following basic properties.

Proposition 2.4. *The map $\mathbb{E}[\cdot|\mathcal{A}]$ is <u>linear</u> and <u>positive</u> on $L^2(\Omega, \mathcal{F}, \mathbb{P})$ and it satisfies*

$$\|\mathbb{E}[X|\mathcal{A}]\|_{L^2} \leq \|X\|_{L^2}, \qquad \forall X \in L^2(\Omega, \mathcal{F}, \mathbb{P}) \tag{2.4}$$

Proof. Linearity and (2.4) follow immediately by definition (see Corollary 2.2). Suppose that $X \geq 0$ with $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. Since the event $A = \{\mathbb{E}[X|\mathcal{A}] \leq 0\} \in \mathcal{A}$, on the one-hand by (2.3), we have

$$\mathbb{E}[\mathbb{E}[X|\mathcal{A}]\,\mathbf{1}_A] \geq 0.$$

On the other-hand since $\mathbb{E}[X|\mathcal{A}]\,\mathbf{1}_A \leq 0$, we conclude that $\mathbb{E}[X|\mathcal{A}]\,\mathbf{1}_{\{\mathbb{E}[X|\mathcal{A}]\leq 0\}} = 0$. This shows that the random variable $\mathbb{E}[X|\mathcal{A}]$ is non-negative (hence, the map $\mathbb{E}[\cdot|\mathcal{A}]$ is positive). □

The random variable $\mathbb{E}[X|\mathcal{A}]$ has the following interpretation: it is the best approximation of X in the space of $\mathcal{A}$-measurable functions for the mean-square distance in the sense that $\mathbb{E}[X|\mathcal{A}]$ is the unique minimizer of $\|Z - X\|_{L^2}$ over all $Z \in L^2(\Omega, \mathcal{A}, \mathbb{P})$.

2.2 General theory

In this section, we generalize the concept of conditional expectation in two cases: *i)* $X \geq 0$ and *ii)* $X \in L^1$. These generalizations rely on Proposition 2.4 and two preliminary Lemmas.

Lemma 2.5. *If $X, Z \geq 0$ are $\mathcal{F}$-measurable and $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A]$ for all $A \in \mathcal{F}$, then $X = Z$ a.s..*

Proof. For any $\epsilon > 0$ and $t \geq 0$, the events $A_{\epsilon,t} = \{X \geq t + \epsilon, t \geq Z\} \in \mathcal{A}$. By assumption,

$$\mathbb{E}\left[Z\mathbf{1}_{A_{\epsilon,t}}\right] \leq t\mathbb{P}[A_{\epsilon,t}] \quad \text{and} \quad \mathbb{E}\left[X\mathbf{1}_{A_{\epsilon,t}}\right] > t\mathbb{P}[A_{\epsilon,t}].$$

This implies that $\mathbb{P}[A_{\epsilon,t}] = 0$ for any $t \in \mathbb{Q}_+$. Moreover, by monotone convergence, for any $t \in \mathbb{Q}_+$,

$$\mathbb{P}[X \geq t \geq Z] = \lim_{\epsilon \to 0} \mathbb{P}[A_{\epsilon,t}] = 0.$$

Now, by density, $\mathbb{P}[X \geq Z] = \bigcup_{t \in \mathbb{Q}_+} \mathbb{P}[X \geq t \geq Z] = 0$. By symmetry, we also have $\mathbb{P}[X \leq Z] = 0$ and we conclude that $X = Z$. ☐

Lemma 2.6. *If $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathbb{E}[X\mathbf{1}_A] = 0$ for all $A \in \mathcal{F}$, then $X = 0$.*

Proof. The events $A_\pm = \{X \in \mathbb{R}_\pm\} \in \mathcal{F}$, so by assumptions: $\mathbb{E}\left[X\mathbf{1}_{A_\pm}\right] = 0$. This implies that the random variables $\pm X\mathbf{1}_{A_\pm} \geq 0$ have mean-zero, so that $X\mathbf{1}_{A_\pm} = 0$. ☐

Let us now extend the definition of $\mathbb{E}[X|\mathcal{A}]$ to any positive random variable X.

Theorem 2.7. *Suppose that $X \geq 0$ is $\mathcal{F}$-measurable and $\mathcal{A} \subset \mathcal{F}$ is a σ-algebra. There exists a unique random variable $Z \geq 0$ which is $\mathcal{A}$-measurable such that*

$$\mathbb{E}[ZW] = \mathbb{E}[XW], \qquad \forall W \geq 0, \mathcal{A}\text{-measurable}. \tag{2.5}$$

We denote $Z = \mathbb{E}[X|\mathcal{A}]$.

Proof. The random variables $X_n = X \wedge n \in L^2$, so we can define $Z_n = \mathbb{E}[X_n|\mathcal{A}] \geq 0$ (see Proposition 2.4). Moreover, since (X_n) is non-decreasing, it follows that (Z_n) is also non-decreasing. Hence, $Z_\infty = \lim_{n\to\infty} Z_n$ exists with $Z_\infty \in [0, \infty]$ and for any $W \geq 0$ ($\mathcal{F}$-measurable),

$$\lim_{n\to\infty} \mathbb{E}[Z_n W] = \mathbb{E}[Z_\infty W]. \tag{2.6}$$

Now for any $Y \in L^2(\Omega, \mathcal{A}, \mathbb{P})$, $\mathbb{E}[Z_n Y] = \mathbb{E}[X_n Y]$ by applying (2.3), so that by (2.6),

$$\mathbb{E}[XY] = \lim_{n\to\infty} \mathbb{E}[X_n Y] = \mathbb{E}[Z_\infty Y].$$

In particular, this holds for all positive simple function Y which are $\mathcal{A}$-measurable and by Theorem 1.1.1), this shows that Z_∞ satisfies (2.5). Uniqueness of this solution follows directly from Lemma 2.5. ☐

Similarly, we can also extend the definition of $\mathbb{E}[X|\mathcal{A}]$ to $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.

Theorem 2.8. *Suppose that $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$ is a σ-algebra. There exists a unique random variable $Z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ such that*

$$\mathbb{E}[ZW] = \mathbb{E}[XW], \qquad \forall W \in L^\infty(\Omega, \mathcal{A}, \mathbb{P}). \tag{2.7}$$

We denote $Z = \mathbb{E}[X|\mathcal{A}]$ and it satisfies $\|\mathbb{E}[X|\mathcal{A}]\|_{L^1} \leq \|X\|_{L^1}$.

Proof. We decompose $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ as

$$X = X_+ - X_-, \qquad \text{where } X_\pm = |X| \mathbf{1}_{X \in \mathbb{R}_\pm}. \tag{2.8}$$

We apply Theorem 2.7 and set $Z_\pm = \mathbb{E}[X_\pm|\mathcal{A}] \geq 0$. Note that since $X_\pm \in L^1$, we also have $Z_\pm \in L^1$ by (2.5) with $W = 1$. Let $Z := Z_+ - Z_-$. By construction, $Z \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and by (2.5), it satisfies

$$\mathbb{E}[ZW] = \mathbb{E}[XW], \qquad \forall W \in L^\infty(\Omega, \mathcal{A}, \mathbb{P}), W \geq 0.$$

This implies that Z satisfies (2.7) by decomposing W according to (2.8) and using linearity of $\mathbb{E}[\cdot]$. Uniqueness of this solution follows directly from Lemma 2.6. It just remains to verify that $\|Z\|_{L^1} \leq \|X\|_{L^1}$. If we take $W = \text{sgn}(Z)$ in (2.8), since $\|\text{sgn}(Z)\|_{L^\infty} \leq 1$ we deduce from Hölder's inequality that

$$\|Z\|_{L^1} = \mathbb{E}[Z\text{sgn}(Z)] = \mathbb{E}[X\text{sgn}(Z)] \leq \|X\|_{L^1}. \qquad \square$$

We are now ready to formalize our definition of conditional expectation.

Definition 2.1. Let $X \geq 0$ or $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. We define $\mathbb{E}[X|\mathcal{A}]$ to be the unique random variable Z which is $\mathcal{A}$-measurable and satisfies

$$\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[X\mathbf{1}_A], \qquad \forall A \in \mathcal{A}. \tag{2.9}$$

$\mathbb{E}[X|\mathcal{A}]$ is called the **conditional expectation of X knowing $\mathcal{A}$** and it satisfies:
i) $\mathbb{E}[X|\mathcal{A}] \geq 0$ if $X \geq 0$. *ii)* $\mathbb{E}[X|\mathcal{A}] \in L^1$ if $X \in L^1$.

This definition is consistent with both Theorem 2.7 and Theorem 2.8. In particular by a standard *density argument* (using Theorem 1.1): *i)* if $X \geq 0$, the condition (2.9) is equivalent to (2.5). *ii)* if $X \in L^1$, the condition (2.9) is equivalent to (2.7).

Remark 2.9. • The **fundamental property** (2.9) characterizes $\mathbb{E}[X|\mathcal{A}]$. It means that to verify whether an $\mathcal{A}$-measurable random variable Z is $\mathbb{E}[X|\mathcal{A}]$, it suffices to check that it satisfies (2.9). In particular, if X is $\mathcal{A}$-measurable, then $\mathbb{E}[X|\mathcal{A}] = X$.
• By the **monotone class Theorem**, it suffices to check that (2.9) is satisfies for all $A \in \Pi$ where Π is any π-system so that $\mathcal{A} = \boldsymbol{\sigma}(\Pi)$.
• Observe that $\mathbb{E}[\cdot] = \mathbb{E}[\cdot|\mathcal{A}]$ where $\mathcal{A} = \{\emptyset, \Omega\}$ is the trivial σ-algebra.

Notation 2.2. • For any $B \in \mathcal{F}$, we denote by $\mathbb{P}[B|\mathcal{A}] = \mathbb{E}[\mathbf{1}_B|\mathcal{A}]$.
• For any random variable Y on $(\Omega, \mathcal{F}, \mathbb{P})$, we denote $\mathbb{E}[\cdot|Y] = \mathbb{E}[\cdot|\boldsymbol{\sigma}(Y)]$.

2.3 Properties of conditional expectation

The goal of this section is to report on how to compute conditional expectations, that is to give the basic properties of $\mathbb{E}[\cdot|\mathcal{A}]$. This first proposition follows immediately from the construction in Section 2.2 – see also Corollary 2.4.

Proposition 2.10. *The map $\mathbb{E}[\cdot|\mathcal{A}]$ is* <u>*linear*</u> *and* <u>*positive*</u> *on $L^1(\Omega, \mathcal{F}, \mathbb{P})$.*

As a consequence of Proposition 2.10, one has similar calculus rules for $\mathbb{E}[\cdot|\mathcal{A}]$ as for the usual expectation $\mathbb{E}[\cdot]$, except that the output is a **random variable**.

Proposition 2.11. *Let (X_n) be random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra.*
1.(Monotone convergence) If there exists a constant $C \geq 0$ such that $X_n \geq -C$ and $X_n \nearrow X_\infty$ a.s., then as $n \to \infty$,

$$\mathbb{E}[X_n|\mathcal{A}] \nearrow \mathbb{E}[X_\infty|\mathcal{A}] \quad a.s..$$

2.(Fatou's lemma) If there exists a constant $C \geq 0$ such that $X_n \geq -C$, then a.s.,

$$\liminf_{n\to\infty} \mathbb{E}[X_n|\mathcal{A}] \geq \mathbb{E}\big[\liminf_{n\to\infty} X_n|\mathcal{A}\big].$$

3. If $X_n \to X_\infty$ in L^1, then $\mathbb{E}[X_n|\mathcal{A}] \to \mathbb{E}[X_\infty|\mathcal{A}]$ in L^1.

Proposition 2.11.3. is a direct consequence from the last bound in Theorem 2.8. This statement replaces the usual *dominated convergence theorem.*

Proof. 1. Since $X_n \geq -C$, its conditional expectation is well defined in the sense of Theorem 2.7. By positivity, the sequence $\mathbb{E}[X_n|\mathcal{A}]$ is non-decreasing so it has limit $Z \geq -C$ which is $\mathcal{A}$-measurable. To check that $Z = \mathbb{E}[X_\infty|\mathcal{A}]$ it suffices to verify it satisfies the condition (2.9): for any $A \in \mathcal{A}$,

$$\mathbb{E}[Z\mathbf{1}_A] = \lim_{n\to\infty} \mathbb{E}[\mathbb{E}[X_n|\mathcal{A}]\,\mathbf{1}_A] = \lim_{n\to\infty} \mathbb{E}[X_n\mathbf{1}_A] = \mathbb{E}[X_\infty\mathbf{1}_A],$$

where both limits exist by the usual monotone convergence theorem.
2. We use the usual argument, by positivity, it holds for any $n \in \mathbb{N}$,

$$\mathbb{E}\big[\inf_{k\geq n} X_k|\mathcal{A}\big] \leq \inf_{k\geq n} \mathbb{E}[X_k|\mathcal{A}].$$

Using 1., we have $\lim_{n\to\infty} \mathbb{E}\big[\inf_{k\geq n} X_k|\mathcal{A}\big] = \mathbb{E}\big[\liminf_{n\to\infty} X_n|\mathcal{A}\big]$ which yields the claim. □

We also have the following version of Jensen's inequality which is important.

Lemma 2.12 (Jensen's inequality)**.** *Let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra, $\varphi : \mathbb{R} \to \mathbb{R}$ be a convex function and $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$, then a.s.,*

$$\varphi(\mathbb{E}[X|\mathcal{A}]) \le \mathbb{E}[\varphi(X)|\mathcal{A}].$$

Proof. We rely on the fact that a function $\varphi : \mathbb{R} \to \mathbb{R}$ is convex if and only if $\varphi(x) = \sup_{n\ge1}(\alpha_n x + \beta_n)$ for two sequences $\alpha_n, \beta_n \in \mathbb{R}$. Then, by Proposition 2.10,

$$\mathbb{E}[\varphi(X)|\mathcal{A}] \ge \sup_{n\ge1} \big(\mathbb{E}[\alpha_n X + \beta_n|\mathcal{A}]\big) = \sup_{n\ge1} \big(\alpha_n \mathbb{E}[X|\mathcal{A}] + \beta_n\big).$$

Note that we do not exclude the possibility that $\mathbb{E}[\varphi(X)|\mathcal{A}] = +\infty$ here and we used that $X \in L^1$ so that the RHS is well-defined with the trivial observation that $\mathbb{E}[1|\mathcal{A}] = 1$. By definition, the RHS equals to $\varphi(\mathbb{E}[X|\mathcal{A}]) \in \mathbb{R}$. ☐

Remark 2.13. *The inequality of Lemma 2.12 also holds if we assume instead that $\varphi : \mathbb{R}_+ \to \mathbb{R}$ is convex and $X \ge 0$.*

We obtain the following inequalities as a direct consequence of Lemma 2.12.

Proposition 2.14 (Norm's estimates)**.** *Let $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. For any $p \in [1, \infty]$ and for all $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$,*

$$\|\mathbb{E}[X|\mathcal{A}]\|_{L^p} \le \|X\|_{L^p}.$$

Proof. Since the function $x \in \mathbb{R} \mapsto |x|^p$ is convex for any $p \in [1, \infty)$, by Jensen's inequality,

$$|\mathbb{E}[X|\mathcal{A}]|^p \le \mathbb{E}[|X|^p|\mathcal{A}]. \tag{2.10}$$

Taking expectation and using that $\mathbb{E}[\mathbb{E}[|X|^p|\mathcal{A}]] = \mathbb{E}[|X|^p|]$, this implies the claim.
For $p = \infty$, we use Proposition 1.2.3) and write

$$\|\mathbb{E}[X|\mathcal{A}]\|_{L^\infty} = \sup\{\mathbb{E}[|\mathbb{E}[X|\mathcal{A}]|Y] : Y \ge 0, Y \text{ is } \mathcal{A}\text{-measurable}, \mathbb{E}[Y] = 1\}.$$

Note that since $\mathbb{E}[X|\mathcal{A}]$ is $\mathcal{A}$-measurable, we can take the supremum only over function Y which are also $\mathcal{A}$-measurable. Now, using (2.10) with $p = 1$, it holds for any $Y \ge 0$, $\mathcal{A}$-measurable,

$$\mathbb{E}[|\mathbb{E}[X|\mathcal{A}]|Y] \le \mathbb{E}[\mathbb{E}[|X||\mathcal{A}]Y] = \mathbb{E}[|X|Y] \le \|X\|_\infty \mathbb{E}[Y].$$

This inequality implies that $\|\mathbb{E}[X|\mathcal{A}]\|_{L^\infty} \le \|X\|_\infty$. ☐

Let us now give the three crucial properties which are specific to conditional expectation.

Proposition 2.15 (Important properties)**.** *Let $X \geq 0$ or $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$.*
1.(*Tower property*) *If $\mathcal{B} \subset \mathcal{A} \subset \mathcal{F}$ are σ-algebra, then*

$$\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{A}\right]|\mathcal{B}\right] = \mathbb{E}\left[X|\mathcal{B}\right].$$

2.(*Taking out what is known*) *If Y is $\mathcal{A}$-measurable and $Y \geq 0$ or $XY \in L^1$, then*

$$\mathbb{E}\left[XY|\mathcal{A}\right] = Y\mathbb{E}\left[X|\mathcal{A}\right].$$

3. *If $\boldsymbol{\sigma}(X)$ is independent from $\mathcal{A}$, then $\mathbb{E}[X|\mathcal{A}] = \mathbb{E}[X]$.*

Proof. For 1., 2. and 3., it suffices to check that the fundamental property (2.9) is satisfied. □

Our last property is a generalization of Proposition 2.15.3.

Proposition 2.16. *Let X, Y be two random variables such that Y is $\mathcal{A}$-measurable and $\boldsymbol{\sigma}(X)$ is independent from $\mathcal{A}$. Let h be a real-valued function such that $h \geq 0$ or $h(X, Y) \in L^1$. Then*

$$\mathbb{E}\left[h(X,Y)|\mathcal{A}\right] = g(Y) \qquad \text{where} \quad g(y) = \int h(x,y)\mathbf{P}_X(\mathrm{d}x).$$

Proof. In class, we give an easier proof in case $\mathcal{A} = \boldsymbol{\sigma}(Y)$. We now give a general argument. Let $Z \in L^\infty(\Omega, \mathcal{A}, \mathbb{P})$ with $Z \geq 0$. Using that $\boldsymbol{\sigma}(X)$ and $\mathcal{A}$, the law of the vector (X, Y, Z) factorizes: $\mathbf{P}_{(X,Y,Z)} = \mathbf{P}_X \otimes \mathbf{P}_{(Y,Z)}$. By Fubini's theorem and the definition of g, this implies that

$$\mathbb{E}\left[h(X,Y)Z\right] = \iint h(x,y)z\mathbf{P}_X(\mathrm{d}x)\mathbf{P}_{(Y,Z)}(\mathrm{d}y,\mathrm{d}z) = \int g(y)z\mathbf{P}_{(Y,Z)}(\mathrm{d}y,\mathrm{d}z) = \mathbb{E}\left[g(Y)Z\right].$$

Note that g is measurable and applying Fubini's theorem is justified in both case $h \geq 0$ or $h \in L^1(\mathbf{P}_X \otimes \mathbf{P}_Y)$. By (2.9), this computation shows that $\mathbb{E}[h(X,Y)|\mathcal{A}] = g(Y)$. □

Proposition 2.17. *Let $X \in L^1$ (or $X \geq 0$) and $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$ be two σ-algebras so that $\mathcal{B}$ is independent of $\boldsymbol{\sigma}(X) \vee \mathcal{A}$. Then $\mathbb{E}[X|\mathcal{A}]$ is also independent from $\mathcal{B}$ and*

$$\mathbb{E}\left[X|\mathcal{A} \vee \mathcal{B}\right] = \mathbb{E}\left[X|\mathcal{A}\right].$$

Proof. See Exercise □

2.4 Explicit cases

2.4.1 Discrete setting

Let us suppose that $\mathcal{A} = \boldsymbol{\sigma}(\Pi)$ is generated by a (countable) partition $\Pi = \{A_n\}$ of Ω (that is the events A_n are disjoint with $\mathbb{P}[A_n] > 0$ and $\Omega = \bigcup_n A_n$). In this case, it holds for any random variable $X \geq 0$ or $X \in L^1(\Omega, \mathcal{F}, \mathbb{P})$,

$$\mathbb{E}[X|\mathcal{A}] = \textstyle\sum_n \mathbf{1}_{A_n} \mathbb{E}[X|A_n], \qquad \mathbb{E}[X|A_n] = \frac{\mathbb{E}[X\mathbf{1}_{A_n}]}{\mathbb{P}[A_n]}. \tag{2.11}$$

Note that since the events A_n are disjoint, for every $\omega \in \Omega$, the sum on the RHS of (2.11) contains at most one non-zero term. This explicit representation is particularly useful when $\mathcal{A} = \boldsymbol{\sigma}(N)$ for a discrete random variable N.

2.4.2 Random variables with a density

Here, we assume that (X, Y) is a random vector with a density. It means that its law has the form $\mathbf{P}_{(X,Y)}(\mathrm{d}x, \mathrm{d}y) = f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y$ where $f_{(X,Y)} \in L^1$ and $f_{(X,Y)} \geq 0$.
Then Y also has a density which is given by $f_Y(y) = \displaystyle\int f_{(X,Y)}(x, y)\mathrm{d}x$.
Let us define a new function $q \geq 0$ by

$$q(x, y) = \mathbf{1}_{\{f_Y(y)>0\}} \frac{f_{(X,Y)}(x, y)}{f_Y(y)}.$$

Proposition 2.18. *Under the above assumptions, for any function $h \geq 0$ or $h \in L^1$, one has*

$$\mathbb{E}[h(X)|Y] = \int h(x)q(x, Y)\mathrm{d}x.$$

Proof. Let $S = \{f_Y(y) > 0\}$ and note that the event $\{Y \in S\}$ has probability 1. Indeed, we have $\mathbb{P}[Y \notin S] = \displaystyle\int_{\{f_Y(y)=0\}} f_Y(y)\mathrm{d}y = 0$. Moreover, by definition of q, it holds for any Borel set B,

$$\begin{aligned}
\mathbb{E}\left[\mathbf{1}_{Y\in B} \int h(x)q(x, Y)\mathrm{d}x\right] &= \iint h(x)\mathbf{1}_{y\in B} q(x, y) f_Y(y)\mathrm{d}y \\
&= \iint h(x)\mathbf{1}_{y\in B, y\in S} f_{(X,Y)}(x, y)\mathrm{d}x\mathrm{d}y \\
&= \mathbb{E}[h(X)\mathbf{1}_{Y\in B, Y\in S}].
\end{aligned}$$

By (2.9), since $\boldsymbol{\sigma}(Y) = \{Y \in B : B \text{ is a Borel set}\}$ and $\mathbf{1}_{Y\in S} = 1$ a.s., this proves the claim. □

2.4.3 Gaussian vectors

Definition 2.3. $X = (X_1, \ldots, X_n)$ is a **Gaussian vector** on $L^2(\Omega, \mathcal{F}, \mathbb{P})$ if for any $a \in \mathbb{R}^n$, $\langle X, a\rangle$ is a real-valued Gaussian variable. We define the mean and **covariance matrix** of X by

$$\mu_i = \mathbb{E}\left[X_i\right], \quad \Sigma_{ij} = \mathbb{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right], \quad \text{for } i, j \in \{1, \ldots, n\}.$$

We denote the law of the vector X by $\mathbf{P}_X = \mathcal{N}_{(\mu,\Sigma)}$. Then, for any $a \in \mathbb{R}^n$,

$$\mathbf{E}_X[e^{\mathbf{i}\langle X,a\rangle}] = e^{\mathbf{i}\langle \mu,a\rangle - \langle a,\Sigma a\rangle/2}.$$

Lemma 2.19. *Let $X = (X_1, \ldots, X_n)$ be a Gaussian vector with law $\mathcal{N}_{(0,\Sigma)}$.*
1) The random variables $(X_1, \ldots, X_n)$ are independent if and only if the matrix Σ is diagonal.
2) There exists a (fixed) lower-triangular matrix L and a Gaussian vector $Z \sim \mathcal{N}_{(0,\mathrm{I})}$ so that $X = \mathrm{L}Z$.

Proof. See exercises □

Proposition 2.20. *Let $(Y_1, \ldots, Y_n, X)$ be a Gaussian vector with law $\mathcal{N}_{(0,\Sigma)}$ and let $\mathcal{A} = \boldsymbol{\sigma}(Y_1, \ldots, Y_n)$. There exists a (fixed) $\lambda \in \mathbb{R}^n$ so that $Z = \mathbb{E}\left[X|\mathcal{A}\right] = \langle \lambda, Y\rangle$. Moreover, for any function $f : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}_+$,*

$$\mathbb{E}\left[f(X,Y)|\mathcal{A}\right] = \int f(Z+w,Y)\mathbf{P}_{\mathcal{N}(0,\sigma^2)}(\mathrm{d}w) = \int f(Z+w,Y)\frac{e^{-w^2/2\sigma^2}}{\sqrt{2\pi}\sigma}\mathrm{d}w,$$

where $\sigma = \|X - Z\|_{L^2} \in \mathbb{R}_+$.

Proof. We use a similar idea as in the proof of Lemma 2.19. Let $Z = \langle \lambda, Y\rangle$ be the L^2-projection of X on $\mathrm{span}(Y)$. By Theorem 2.1, we have

$$\mathbb{E}\left[(X-Z)Y_k\right] = 0 \qquad \text{for all } k \in \{1, \ldots n\}.$$

Since $(X - Z, Y_1, \ldots, Y_n)$ is a Gaussian vector, this implies that $\boldsymbol{\sigma}(Z - X)$ is independent from $\mathcal{A}$. Consequently, we can apply Proposition 2.16 by writing $f(X,Y) = h(X - Z, Y)$. Then as $(X - Z)$ is a centered Gaussian, we obtain $\mathbb{E}\left[f(X,Y)|\mathcal{A}\right] = g(Y)$ where

$$g(y) = \int h(w,y)\mathbf{P}_{(X-Z)}(\mathrm{d}w) = \int f(Z+w,y)\mathbf{P}_{\mathcal{N}(0,\sigma^2)}(\mathrm{d}w),$$

where $\sigma = \|X - Z\|_{L^2} \in \mathbb{R}_+$ is the appropriate standard deviation. Moreover, if we apply this result with $f(x,y) = x$ (this function is in $L^1(\mathbf{P}_{\mathcal{N}(0,\sigma^2)})$), we get $\mathbb{E}\left[X|\mathcal{A}\right] = Z = \langle \lambda, Y\rangle$. □

2.5 Alternative construction: L^1 theory

This construction relies on the Radon–Nikodym Theorem that we now review.

Definition 2.4. Let μ, η be two measures on $(\Omega, \mathcal{F})$. We say that μ is **absolutely continuous** with respect to η if

$$\mu(A) = 0 \qquad \forall A \in \{O \in \mathcal{F} : \eta(O) = 0\}.$$

Then, we write $\mu \ll \eta$.

Theorem 2.21 (Radon–Nikodym)**.** *Let μ, η be two σ-finite measures on $(\Omega, \mathcal{F})$. Then, $\mu \ll \eta$ if and only if there exists a (unique) function*[1] *$f : \Omega \to \mathbb{R}_+$ so that*

$$\mu(A) = \int_A f \mathrm{d}\eta.$$

f is called the **density** *of μ with respect to η and it is usually denoted by $f = \frac{\mathrm{d}\mu}{\mathrm{d}\eta}$.*

Proof. We admit this result for now. □

Let $0 \leq X < \infty$ be a $\mathcal{F}$-measurable random variable and $\mathcal{A} \subset \mathcal{F}$ be a σ-algebra. It is straightforward to check that the set function $A \in \mathcal{A} \mapsto \mathbb{E}[X\mathbf{1}_A]$ is a σ-finite measure which is absolutely continuous with respect to $\mathbb{P}$ on $(\Omega, \mathcal{A})$. Therefore, by Theorem 2.21, it admits a density that we denote by $\mathbb{E}[X|\mathcal{A}]$.
By definition, $\mathbb{E}[X|\mathcal{A}] \geq 0$ and it is the unique $\mathcal{A}$-measurable random variable which satisfies

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{A}]\,\mathbf{1}_A], \qquad \forall\, A \in \mathcal{A}.$$

This gives an alternative proof of Theorem 2.7.

[1] f is uniquely defined as an equivalent class of measurable maps on $(\Omega, \mathcal{F}, \eta)$.

3 Discrete time martingale

We consider the probability space $(\Omega, \mathcal{F}_\infty, \mathbb{P})$. The interest of martingale in probability theory lie in that the martingale property is generally easy to verify and there are stopping theorems and general convergence results;

3.1 Definitions

Definition 3.1. • A **filtration** is a sequence of σ-algebras $\mathcal{F} = (\mathcal{F}_n)_{n=0}^\infty$ such that

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_\infty.$$

• A sequence of random variables $X = (X_n)_{n=0}^\infty$ is called **adapted to $\mathcal{F}$** if

$$X_n \text{ is } \mathcal{F}_n\text{-measurable for any } n \in \mathbb{N}_0.$$

• A sequence of random variables $X = (X_n)_{n=1}^\infty$ is called **predictable (with respect to $\mathcal{F}$)** if

$$X_n \text{ is } \mathcal{F}_{n-1}\text{-measurable for any } n \in \mathbb{N}.$$

• X is called a **random process** if $X_n \in L^1(\Omega, \mathcal{F}_n, \mathbb{P})$ for any $n \in \mathbb{N}$.

Definition 3.2. If $X = (X_n)_{n=0}^\infty$ is a sequence of random variables, we define its filtration

$$\mathcal{F}^X = \left(\mathcal{F}_n^X = \boldsymbol{\sigma}(X_0, \ldots, X_n)\right)_{n=0}^\infty.$$

$\mathcal{F}^X$ is called **the canonical filtration of X**.

Definition 3.3. A **random process** X is a **martingale (with respect to $\mathcal{F}$)** if for any $m \le n$,

$$\mathbb{E}\left[X_n | \mathcal{F}_m\right] = X_m.$$

Definition 3.4. A random sequence $X = (X_n)_{n=0}^\infty$ is called **a positive martingale (with respect to $\mathcal{F}$)** if X is **adapted**, $X_n \ge 0$ and for any $m \le n$,

$$\mathbb{E}\left[X_n | \mathcal{F}_m\right] = X_m.$$

Remark 3.1. *If X is a martingale with respect to $\mathcal{F}$, it need not be a martingale with respect to its own filtration $\mathcal{F}^X \prec \mathcal{F}$.*

Definition 3.5. • A **random process** X is a **submartingale (with respect to $\mathcal{F}$)** if for any $m \le n$,

$$\mathbb{E}\left[X_n | \mathcal{F}_m\right] \ge X_m.$$

• We say that X is a **supermartingale** if $-X$ is a submartingale.

Remark 3.2. *A martingale is a random process with constant expectation:* $\mathbb{E}[X_n] = \mathbb{E}[X_0]$. *Submartingales have non-decreasing expectations:* $\mathbb{E}[X_n] \geq \mathbb{E}[X_m]$ *for any* $m \leq n$.

The most basic example of a random process X is a random walk:

$$X_n = Z_1 + \cdots + Z_n \qquad \text{for } n \geq 1,$$

where Z_k are i.i.d. rv in L^1. With respect to its canonical filtration $\mathcal{F}^X$,

- X is a martingale if $\mathbb{E}[Z_k] = 0$.
- X is a supermartingale if $\mathbb{E}[Z_k] \leq 0$.
- X is a submartingale if $\mathbb{E}[Z_k] \geq 0$.

Proposition 3.3 (Convex transformations)**.** *Let* $\varphi : \mathbb{R} \to \mathbb{R}$ *be a convex function so that* $\mathbb{E}[\varphi(X_n)] < \infty$ *for any* $n \in \mathbb{N}_0$. *We denote* $\varphi(X) = (\varphi(X_n))_{n=0}^{\infty}$.
- *If* X *is a martingale, then* $\varphi(X)$ *is a submartingale.*
- *If* φ *is non-decreasing and* X *is a submartingale, then* $\varphi(X)$ *is also a submartingale.*

Proof. These claims follow directly from Jensen's inequality; see Lemma 2.12. □