5 Probability theory

5.1 Basic definitions and properties

Modern probability is mostly built upon measure theory following the work of Kolmogorov . The starting point of this field is a universe, i.e., a measurable space $(\Omega, \mathcal{F})$, where $\mathcal{F}$ is a σ-field over Ω. It is considered fixed in all this section.

We first preface this section by two basic definitions.

Definition 5.1. Any measurable function $X : \Omega \to \mathsf{E}$ where $(\mathsf{E}, \mathcal{A})$ is a measurable space is called a random variable valued in E. If $\mathsf{E} = \mathbb{R}$ (resp. $\mathbb{C}$) endowed with its Borel σ-field, X is said to be a real (resp. complex) random variable.

Definition 5.2. A universe endowed with a probability measure $\mathbb{P}$ is called a probability space, denoted in general on the form of a triplet $(\Omega, \mathcal{F}, \mathbb{P})$.

Note that for any measurable function $f : \mathsf{E} \to \mathsf{F}$, where $(\mathsf{E}, \mathcal{A})$ and $(\mathsf{F}, \mathcal{F})$ are two measurable spaces, and X a random variable valued in E, $f(X)$ is a random variable. Therefore, measurable transformation of random variables are random variables.

While the definition of a random variable do not depend on any probability measure on $(\Omega, \mathcal{F})$, this is the combination of these objects that are the central studied object of probability theory, i.e., the distribution (also called law) of X.

Proposition and Definition 5.1. *Let $X : \Omega \to \mathsf{E}$ be a random variable where $(\mathsf{E}, \mathcal{A})$ is a measurable space. The map from $\mathcal{A}$ to $[0,1]$, defined by*

$$\mathbb{P}_X(\mathsf{A}) = \mathbb{P}(X \in \mathsf{A}), \mathsf{A} \in \mathcal{A}, \tag{5.1}$$

is a probability measure on $(\mathsf{E}, \mathcal{A})$, referred to as the distribution of X.

Proof. The proof is left as an exercise. □

We now turn to the definition of the expectation of $f(X)$.

Definition 5.3. Let $f : \mathsf{E} \to \mathbb{R}$ be a real measurable function and X be a random variable values in E.

(1) If f is non-negative, we define the expectation of $f(X)$ as $\mathbb{E}[f(X)] = \int f(X(\omega))\mathbb{P}(\mathrm{d}\omega)$.

(2) For general f, such that $\mathbb{E}[|f(X)|] < +\infty$, we define

$$\mathbb{E}[f(X)] = \int f(X(\omega))\mathbb{P}(\mathrm{d}\omega) = \int f_+(X(\omega))\mathbb{P}(\mathrm{d}\omega) - \int f_-(X(\omega))\mathbb{P}(\mathrm{d}\omega)\,,$$

where we have defined $f_+ = \max(0, f)$ and $f_- = \max(0, -f)$.

Another option for the definition of $\mathsf{E}[f(X)]$ would be to change the probability measure with respect to which the integral is taken, i.e., to define $\mathsf{E}[f(X)]$ as $\int_{\mathsf{E}} f(x)\mathbb{P}_X(\mathrm{d}x)$, where $\mathbb{P}_X$ is the distribution of X. In fact, it turns out that these two quantities are equal.

Proposition 5.2 (Transfer Lemma)**.** *Let $f : \mathsf{E} \to \mathbb{R}$ be a real measurable function and X be a random variable valued in E. Assume either f is non-negative, or $\mathbb{E}[|f|(X)] < +\infty$. Then,*

$$\mathbb{E}[f(X)] = \int_\Omega f(X(\omega))\mathbb{P}(\mathrm{d}\omega) = \int_{\mathsf{E}} f(x)\mathbb{P}_X(\mathrm{d}x)\,. \tag{5.2}$$

Proof. The case f satisfying $\mathbb{E}[|f|(X)] < +\infty$ is a simple consequence of the case f non-negative. By definition (5.1), (5.2) holds for $f = \mathbb{1}_{\mathsf{A}}$. By linearity, (5.2) holds for simple functions. Then, by the approximation lemma and the monotone convergence theorem. □

Theorem 5.3. *Jensen's Inequality*

(i) Let X be a real integrable random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$ taking values in an interval I, and let ϕ be a convex function from I into $\mathbb{R}$ such that $\phi \circ X$ is $\mathbb{P}$-integrable. It holds that

$$\phi(\mathbb{E}[X]) \;=\; \phi\left(\int_\Omega X(\omega)\mathrm{d}\mathbb{P}(\omega)\right) \leq \int_\Omega \phi \circ X(\omega)\mathrm{d}\mathbb{P}(\omega) \;=\; \mathbb{E}[\phi(X)]. \tag{5.3}$$

(ii) (Equality case) Now fix $\phi : \mathbb{R} \to \mathbb{R}$ a strictly *convex function. Equality in equation (5.3) holds only for constant random variables X $\mathbb{P}$-almost surely.*

Proof. (i) We start with a short recall on convex functions.

Lemma 5.4. *Let $\phi : I \to \mathbb{R}$ be a convex function. Then for all $x \in I$, the function defined on $I \setminus \{x\}$, $p_x : y \mapsto (\phi(y) - \phi(x))/(y - x)$ is non-decreasing on $I \setminus \{x\}$.*

Proof. Left as an exercise. □

From Theorem 5.4, if ϕ is convex, it admits right and left derivatives denoted respectively by ϕ'_+ and ϕ'_- (which may be infinite at the boundary of I). Moreover these functions satisfy by definition, for all $h_1, h_2 \in \mathbb{R}$, $x \in I$, $x + h_1 \in I$, $x - h_2 \in I$,

$$(\phi(x+h_1) - \phi(x))/h_1 \geq \phi'_+(x)\,, (\phi(x) - \phi(x-h_2))/h_2 \leq \phi'_-(x)\,, \phi'_-(x) \leq \phi'_+(x)\,.$$

Thus for all $y \in I$, $\phi(y) \geq \phi'_+(x)(y-x) + \phi(x)$ which implies that for all $x \in \mathrm{int}(I)$

$$\phi(x) = \sup\left\{ax + b \,:\, a, b \in \mathbb{R} \text{ and for all } y \in I\ ay + b \leq \phi(y)\right\}\,,$$

this supremum being attained for $a = \phi'_+(x)$ and $b = \phi(x) - \phi'_+(x)x$. Moreover we use the following lemma.

Lemma 5.5. *Let $(\mathsf{X}, \mathcal{X}, \mu)$ be a measured space with μ a probability measure and $f : \mathsf{X} \to I$, where I is an interval of $\mathbb{R}$, an integrable function. Then $\int_{\mathsf{X}} f \, \mathrm{d}\mu \in I$. Moreover if $\int_{\mathsf{X}} f \, \mathrm{d}\mu \notin \mathrm{int}(I)$, then $f = \int_{\mathsf{X}} f \, \mathrm{d}\mu$, μ-a.s.*

Proof. Left as an exercise.

□

Hence $\mathbb{E}[X] \in I$. Now, we distinguish the case whether $\mathbb{E}[X] \in \mathrm{int}(I)$ or not.

Case a) $\mathbb{E}[X] \in \mathrm{int}(I)$. Then there exist $a_0, b_0 \in \mathbb{R}$ such that

$$\phi(\mathbb{E}[X]) = a_0 \mathbb{E}[X] + b_0\,, a_0 y + b_0 \leq \phi(y)\,, \text{ for all } y \in I\,.$$

Thus, by monotonicity and linearity of expectation,

$$\phi(\mathbb{E}[X]) = a_0 \mathbb{E}[X] + b_0 = \mathbb{E}[a_0 X + b_0] \leq \mathbb{E}\left[\phi(X)\right]\,.$$

Case b) $\mathbb{E}[X] \notin \mathrm{int}(I)$. Then $X = \mathbb{E}[X]$, $\mathbb{P}$-almost surely and Jensen's inequality is then clearly satisfied.

(ii) We start again with a short recall on convex functions.

Lemma 5.6. *Let $\phi : I \to \mathbb{R}$ be a strictly convex function. Then for all $x \in I$, the function defined on $I \setminus \{x\}$, $p_x : y \mapsto (\phi(y) - \phi(x))/(y - x)$ is strictly increasing on $I \setminus \{x\}$.*

Proof. Left as an exercise.

□

From Theorem 5.6, if ϕ is convex, it admits right and left derivatives denoted respectively by ϕ'_+ and ϕ'_- (which may be infinite at the boundary of I). Moreover these functions satisfy by definition, for all $h_1, h_2 \in \mathbb{R}$, $x \in I$, $x + h_1 \in I$, $x - h_2 \in I$,

$$(\phi(x+h_1) - \phi(x))/h_1 > \phi'_+(x)\,, (\phi(x) - \phi(x-h_2))/h_2 \leq \phi'_-(x)\,, \phi'_-(x) < \phi'_+(x)\,.$$

Thus for all $y \in I$, $y \neq x$, $\phi(y) > \phi'_+(x)(y-x) + \phi(x)$ which implies that for all $x \in \mathrm{int}(I)$

$$\phi(x) = \sup\left\{ax + b : a, b \in \mathbb{R} \text{ and for all } y \in I,\ y \neq x,\ ay + b < \phi(y)\right\}\,,$$

this supremum being attained for $a = \phi'_+(x)$ and $b = \phi(x) - \phi'_+(x)x$. As in the first question, we distinguish the cases starting from the fact that $\mathbb{E}[X] \in I$.

Case a) $\mathbb{E}[X] \in \mathrm{int}(I)$. Then there exist $a_0, b_0 \in \mathbb{R}$, such that

$$\phi(\mathbb{E}[X]) = a_0 \mathbb{E}[X] + b_0 \text{ and for all } y \in I,\ y \neq \mathbb{E}[X],\ a_0 y + b_0 < \phi(y)\,. \tag{5.4}$$

Also suppose that (5.3) is an equality under the assumption that ϕ is strictly convex. Then we obtain that $\mathbb{E}[\phi(X) - \phi(\mathbb{E}[X])] = 0$. But since $\mathbb{P}$-almost surely $\mathbb{E}\left[\phi(X)\right] - \phi(\mathbb{E}[X]) - a_0(X_0 - \mathbb{E}[X]) > 0$, we have that $\phi(X) = a_0\mathbb{E}[X] + b_0$ $\mathbb{P}$-a.s. Hence from (5.4), $\mathbb{P}$-a.s., $X = \mathbb{E}[X]$.

Case b) $\mathbb{E}[X] \notin \mathrm{int}(I)$, then $\mathbb{P}$-a.s., $X = \mathbb{E}[X]$.

□

5.1.1 Real random variables

Definition 5.4. Let X be a real random variable. We define the cumulative distribution function of X as

$$F_X(x) = \mathbb{P}(X \leq x)\,, \forall x \in \mathbb{R}\,.$$

Cumulative distribution functions are also simply referred to as distribution function.

Proposition 5.7. *Let X and Y be two real random variables for which distribution functions coincide on $\mathbb{Q}$. Then they have the same distribution.*

Proof. It is a simple consequence of Exercise 2.1 and the Dynkin theorem Theorem 2.7. □

Proposition 5.8. *Let F_X be the distribution function of a real random variable X. Then F_X is cadlag (continue à droite et limite à gauche in French), i.e., is right-continuous and admits left limit at any point in $\mathbb{R}$: for any $x \in \mathbb{R}$,*

$$F(x) = \lim_{y \downarrow x} F(y)\,, \quad \lim_{y \uparrow x} F(y) \text{ exists}$$

Proof. They are simple consequences of the continuity properties of $\mathbb{P}$ stated in Theorem 2.5. □

5.2 Conditional distribution

5.2.1 Elementary conditional probability

We consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For an event $\mathsf{B} \in \mathcal{F}$, $\mathbb{P}(\mathsf{B}) > 0$, we define:

Definition 5.5. We define the probability measure $\mathsf{A} \in \mathcal{F} \mapsto \mathbb{P}(\mathsf{A}|\mathsf{B})$, called the conditional probability given B, as

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

Proposition 5.9. *If* $\mathbb{P}(\mathsf{B}) > 0$*, then* $\mathbb{P}(\cdot|\mathsf{B})$ *is a probability measure on* $(\Omega, \mathcal{F})$*.*

Remark 5.1. Let $\mathsf{A}, \mathsf{B} \in \mathcal{F}$ with $\mathbb{P}(\mathsf{A}), \mathbb{P}(\mathsf{B}) > 0$. Then, it easy to notice that

$$\mathsf{A}, \mathsf{B} \text{ are independent} \quad \iff \quad \mathbb{P}(\mathsf{A}|\mathsf{B}) = \mathbb{P}(\mathsf{A}) \quad \iff \quad \mathbb{P}(\mathsf{B}|\mathsf{A}) = \mathbb{P}(\mathsf{B}) .$$

Definition 5.6. The expectation associated with $\mathbb{P}(\cdot|\mathsf{B})$ is called the conditional expectation given B and denoted for any real random variable Z, non-negative or $\mathbb{P}(\cdot|\mathsf{B})$-integrable, as $\mathbb{E}[Z|\mathsf{B}]$.

Proposition 5.10. *For any real random variable Z, non-negative or* $\mathbb{P}(\cdot|\mathsf{B})$ *integrable,* $\mathbb{E}[Z|\mathsf{B}] = (1/\mathbb{P}(\mathsf{B}))\mathbb{E}[Z\mathbb{1}_{\mathsf{B}}]$*.*

Proof. This is true for any $Z = 1_{\mathsf{A}}$ by definition and therefore we get the result by approximation Theorem 3.10. □

Theorem 5.11. *Let* I *a countable set and let* $(B_i)_{i\in\mathsf{I}}$ *be a partition of* Ω*, so* $\mathbb{P}(\sqcup_{i\in\mathsf{I}}\mathsf{B}_i) = 1$*. Then, for any* $\mathsf{A} \in \mathcal{F}$*,*

$$\mathbb{P}(\mathsf{A}) = \sum_{i\in\mathsf{I}} \mathbb{P}(\mathsf{A}|\mathsf{B}_i)\mathbb{P}(\mathsf{B}_i) .$$

Proof. Due to the σ-additivity of $\mathbb{P}$, we have

$$\mathbb{P}(\mathsf{A}) = \mathbb{P}(\sqcup_{i\in\mathsf{I}}(\mathsf{A} \cap \mathsf{B}_i)) = \sum_{i\in\mathsf{I}} \mathbb{P}(\mathsf{A} \cap \mathsf{B}_i) = \sum_{i\in\mathsf{I}} \mathbb{P}(\mathsf{A}|\mathsf{B}_i)\mathbb{P}(\mathsf{B}_i) .$$

□

Theorem 5.12. *Let* I *a countable set and let* $(B_i)_{i\in I}$ *be a partition of* Ω*, so* $\mathbb{P}(\sqcup_{i\in I} B_i) = 1$*. Then, for any* $A \in \mathcal{F}$ *with* $\mathbb{P}(A) > 0$ *and any* $k \in I$*,*

$$\mathbb{P}(B_k|A) = \frac{\mathbb{P}(A|B_k)\mathbb{P}(B_k)}{\sum_{i\in I}\mathbb{P}(A|B_i)\mathbb{P}(B_i)}$$

Proof. Note that by definition that

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

Now using Theorem 5.11 completes the proof. □

Example 5.1. In the general population, the change of having a cancer is about 2%. We consider a test which detects a cancer with probability 95%; however, with probability 10% it gives a false alarm. If the test of a patient is positive (i.e., gives an alarm), what is the probability that this patient has indeed a cancer?

We formalize the description given above. Let

$$A = \{\text{ test is positive }\}, \quad B = \{\text{ patient has cancer }\}.$$

The description above implies that

$$\mathbb{P}(B) = 0.02\,, \mathbb{P}(B^C) = 0.98\,, \mathbb{P}(A|B) = 0.95\,, \mathbb{P}(A|B^C) = 0.1\,.$$

Therefore, Bayes' formula yields

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(A|B)\mathbb{P}(B)}{\mathbb{P}(A|B)\mathbb{P}(B) + \mathbb{P}\left(A\middle|B^C\right)\mathbb{P}\left(B^C\right)} = \frac{0.95\cdot 0.02}{0.95\cdot 0.02 + 0.1\cdot 0.98} = \frac{19}{117} \approx 0.162\,.$$

On the other hand, the probability that the patient has a cancer while the test is negative is:

$$\mathbb{P}(B|A^C) = \frac{0.05\cdot 0.02}{0.05\cdot 0.02 + 0.9\cdot 0.98} = \frac{1}{883} \approx 0.00113\,.$$

Remark 5.2. Let $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$. If $A \in \mathcal{F}$, then clearly also $\mathbb{1}_A X \in L^1(\Omega,\mathcal{F},\mathbb{P})$. Some authors then use the following notation:

$$\mathbb{E}[X;A] = \mathbb{E}[\mathbb{1}_A X]\,.$$

Let I be a countable set and let $(B_i)_{i\in I}$ be pairwise disjoint events with $\sqcup_{i\in I} B_i = \Omega$. We define $\mathcal{G} = \sigma(\{B_i : i \in I\})$. For $X \in L^1(\Omega,\mathcal{F},\mathbb{P})$, we define a map $\mathbb{E}[X \mid \mathcal{G}] : \Omega \to \mathbb{R}$ by

$$\mathbb{E}[X \mid \mathcal{G}](\omega) = \sum_{i \in \mathsf{I}} \mathbb{1}_{\mathsf{B}_i}(\omega)\mathbb{E}[X \mid \mathsf{B}_i] .$$

Proposition 5.13. *Let $X \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ and $\mathcal{G} = \sigma(\{\mathsf{B}_i : i \in \mathsf{I}\})$, for $(\mathsf{B}_i)_{i\in\mathsf{I}}$ pairwise disjoint events and I a countable set. The map $\mathbb{E}[X \mid \mathcal{G}]$ has the following properties.*

(i) $\mathbb{E}[X \mid \mathcal{G}]$ *is* $\mathcal{G}$*-measurable.*

(ii) $\mathbb{E}[X \mid \mathcal{G}] \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$*, and for any* $\mathsf{A} \in \mathcal{G}$*, we have*

$$\mathbb{E}[\mathbb{1}_{\mathsf{A}}\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{1}_{\mathsf{A}} X] .$$

Proof. (i) We can suppose that $\mathsf{I} \subset \mathbb{N}$. The proof is immediate since we can write $\mathbb{E}[X \mid \mathcal{G}] = g \circ f$ with $f : \Omega \to \mathsf{I}$, $g : \mathsf{I} \to \mathbb{R}$,

$$f = \sum_{i\in\mathsf{I}} \mathsf{I}\mathbb{1}_{\mathsf{B}_i} \quad , \quad g(i) = \mathbb{E}[X \mid \mathsf{B}_i] ,$$

and both functions are measurable and in particular by definition f is $\mathcal{G}/\wp(\mathsf{I})$-measurable.

(ii) It is left as an exercise. Hint: First show that for any $\mathsf{A} \in \mathcal{G}$, there exists $\mathsf{J} \subset \mathsf{I}$ with $\mathsf{A} = \sqcup_{j\in\mathsf{J}} \mathsf{B}_j$.

□

Exercise 5.1. Let $X > 0$ be a strictly positive random variable. Show that X is exponentially distributed if and only if

$$\mathbb{P}(X > t + s \mid X > s) = \mathbb{P}(X > t) \quad \text{for all } s, t \geq 0 .$$

In particular, $X \sim \mathbf{Exp}(\theta)$ for $\theta > 0$ if and only if $\mathbb{P}(X > t + s \mid X > s) = \mathrm{e}^{-\theta t}$ for any $s, t \geq 0$.

5.2.2 Conditional Expectations

There is multiple interpretation of the condition expectation given a sub σ-field. A first is simply that it corresponds to an orthogonal projection in $\mathrm{L}^2(\Omega, \mathbb{P}, \mathcal{F})$; see Theorem 5.19. More precisely, let (X, Y) be two independent random variables valued in $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ respectively, such that $f(X, Y)$ is in L^2 for $f : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}$. Then, denoting by $h(Y) = \mathbb{E}[f(X, Y)|Y]$ as the best function of Y approximating in $\mathrm{L}^2(\Omega, \mathsf{F}, \mathbb{P})$, $f(X, Y)$.

A second interpretation is that it corresponds to the integration of a random variable given possible events not belonging to $\mathcal{G}$. A nice and useful result of this fact is that if (X, Y) are two independent random variables valued in $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ respectively, such that $f(X, Y)$ is integrable, then $\mathbb{E}[f(X, Y)|Y] = h(Y)$ where $h(y) = \mathbb{E}[f(X, y)]$. To interpret this statement, Y being independent of X all events relative to this random variable is not relevant for X. Therefore, given the second interpretation of the conditional expectation, we can treat Y as constant in the computation of $\mathbb{E}[f(X,Y)|Y]$ which leads to the result. In the generic case, not assuming X and Y are independent, $\mathbb{E}[f(X,Y)|Y]$ still corresponds to a random variable of the form $h(Y)$ for some measurable function $h : \mathsf{Y} \to \mathbb{R}$. In this setting $h(Y)$ still corresponds in some sense to the integration of $f(X,Y)$ with Y being fixed. This will naturally lead to define conditional distribution which is the topic of Chapter 7.

In the following, $\mathcal{G} \subset \mathcal{F}$ will be a sub-σ-field and $X \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. In analogy with Theorem 5.13, we make the following definition.

Definition 5.7. A real random variable Y is called a conditional expectation of real random variable (non-negative or integrable) X given $\mathcal{G}$ if:

(i) Y is $\mathcal{G}$-measurable.

(ii) For any $\mathsf{A} \in \mathcal{G}$, we have $\mathbb{E}[X\mathbb{1}_\mathsf{A}] = \mathbb{E}[Y\mathbb{1}_\mathsf{A}]$.

Theorem 5.14. *The conditional expectation of* $X \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ *given* $\mathcal{G}$ *exists and is unique (up to equality almost surely).*

Remark 5.3. (1) Since conditional expectations are defined only up to equality a.s., all equalities with conditional expectations are understood as equalities a.s., even if we do not say so explicitly.

(2) We can also show Theorem 5.14 and the results above assuming that X is a non-negative random variable.

(3) By uniqueness, we can talk about the conditional expectation of X given $\mathcal{G}$ and we denote it by $\mathbb{E}[X|\mathcal{G}]$ and $\mathbb{E}[X|Y]$ if $\mathcal{G} = \sigma(Y)$.

Proof. **Uniqueness.** Let Y and Y' be random variables that fulfill (i) and (ii). Let $\mathsf{A} = \{Y > Y'\} \in \mathcal{G}$. Then, by (ii),

$$0 = \mathbb{E}[Y\mathbb{1}_\mathsf{A}] - \mathbb{E}\left[Y'\mathbb{1}_\mathsf{A}\right] = \mathbb{E}\left[\left(Y - Y'\right)\mathbb{1}_\mathsf{A}\right] .$$

Since $\left(Y - Y'\right)\mathbb{1}_\mathsf{A} \geq 0$, we have $\mathbb{P}(\mathsf{A}) = 0$; hence $Y \leq Y'$ almost surely. Similarly, we get $Y \geq Y'$ almost surely.

Existence. Let $X^+ = X \vee 0$ and $X^- = X^+ - X$. The maps

$$\mathbb{Q}^{\pm}(A) = \mathbb{E}\left[X^{\pm}\mathbb{1}_\mathsf{A}\right] \text{ for all } \mathsf{A} \in \mathcal{G}$$

define two finite measures on $(\Omega, \mathcal{G})$. Clearly, $\mathbb{Q}^{\pm} \ll \mathbb{P}$; hence the Radon-Nikodym theorem Theorem 4.8 yields the existence of $\mathcal{G}$-measurable densities $Y^{\pm}$such that

$$\mathbb{Q}^{\pm}(A) = \int_\mathsf{A} Y^{\pm} \mathrm{d}\mathbb{P} = \mathbb{E}\left[Y^{\pm}\mathbb{1}_\mathsf{A}\right]$$

Now define $Y = Y^+ - Y^-$. We show easily that Y is a conditional expectation of X given $\mathcal{G}$. □

Theorem 5.15. *Let* $(\Omega,\mathcal{F},\mathbb{P})$ *and let* X,Y *be either non-negative or integrable random variables. Let* $\mathcal{A}\subset\mathcal{G}\subset\mathcal{F}$ *be* σ*-fields and let* $Y\in \mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$*. Then:*

(i) For any $\lambda\in\mathbb{R}$ *(*$\lambda\geq 0$ *only if* X *is not integrable),* $\mathbb{E}[\lambda X+Y\mid\mathcal{G}]=\lambda\mathbb{E}[X\mid\mathcal{G}]+\mathbb{E}[Y\mid\mathcal{G}]$.

(ii) If $X\geq Y$ *a.s., then* $\mathbb{E}[X\mid\mathcal{G}]\geq\mathbb{E}[Y\mid\mathcal{G}]$.

(iii) If $\mathbb{E}[|XY|]<+\infty$ *and* Y *is measurable with respect to* $\mathcal{G}$*, then*

$$\mathbb{E}[XY\mid\mathcal{G}]=Y\mathbb{E}[X\mid\mathcal{G}]\quad\textit{and}\quad\mathbb{E}[Y\mid\mathcal{G}]=\mathbb{E}[Y\mid Y]=Y\ .$$

(iv) (Tower property) $\mathbb{E}[\mathbb{E}[X\mid\mathcal{G}]\mid\mathcal{A}]=\mathbb{E}[\mathbb{E}[X\mid\mathcal{A}]\mid\mathcal{G}]=\mathbb{E}[X\mid\mathcal{A}]$.

(v) $\mathbb{E}[|X|\mid\mathcal{G}]\geq|\mathbb{E}[X\mid\mathcal{G}]|$.

(vi) (Independence) If $\sigma(X)$ *and* $\mathcal{G}$ *are independent, then* $\mathbb{E}[X\mid\mathcal{G}]=\mathbb{E}[X]$.

(vii) If $\mathbb{P}(\mathsf{A})\in\{0,1\}$ *for any* $\mathsf{A}\in\mathcal{G}$*, then* $\mathbb{E}[X\mid\mathcal{G}]=\mathbb{E}[X]$.

Proof. (i) The right-hand side is $\mathcal{G}$-measurable; hence, for $\mathsf{A}\in\mathcal{G}$,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{\mathsf{A}}(\lambda\mathbb{E}[X\mid\mathcal{G}]+\mathbb{E}[Y\mid\mathcal{G}])]&=\lambda\mathbb{E}[\mathbb{1}_{\mathsf{A}}\mathbb{E}[X\mid\mathcal{G}]]+\mathbb{E}[\mathbb{1}_{\mathsf{A}}\mathbb{E}[Y\mid\mathcal{G}]]=\lambda\mathbb{E}[\mathbb{1}_{\mathsf{A}}X]+\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y]\\&=\mathbb{E}[\mathbb{1}_{\mathsf{A}}(\lambda X+Y)]\ .\end{aligned}$$

(ii) Let $\mathsf{A}=\{\mathbb{E}[X\mid\mathcal{G}]<\mathbb{E}[Y\mid\mathcal{G}]\}$. By definition, $\mathsf{A}\in\mathcal{G}$ and since we have $X\geq Y$, we get $\mathbb{E}[\mathbb{1}_{\mathsf{A}}(\mathbb{E}[X|\mathcal{G}]-\mathbb{E}[Y|\mathcal{G}])]=\mathbb{E}[\mathbb{1}_{\mathsf{A}}(X-Y)]\geq 0$ and thus $\mathbb{P}(\mathsf{A})=0$.

(iii) First assume $X\geq 0$ and $Y\geq 0$. For $n\in\mathbb{N}$, define $Y_n=2^{-n}\lfloor 2^nY\rfloor$. Then $Y_n\uparrow Y$ and $Y_n\mathbb{E}[X\mid\mathcal{G}]\uparrow Y\mathbb{E}[X\mid\mathcal{G}]$ (since $\mathbb{E}[X\mid\mathcal{G}]\geq 0$ by (ii)). By the monotone convergence theorem,

$$\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y_n\mathbb{E}[X\mid\mathcal{G}]]\xrightarrow{n\to+\infty}\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y\mathbb{E}[X\mid\mathcal{G}]]$$

On the other hand,

$$\begin{aligned}\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y_n\mathbb{E}[X\mid\mathcal{G}]]&=\sum_{k=1}^{+\infty}\mathbb{E}\left[\mathbb{1}_{\mathsf{A}}\mathbb{1}_{\{Y_n=k2^{-n}\}}k2^{-n}\mathbb{E}[X\mid\mathcal{G}]\right]\\&=\sum_{k=1}^{+\infty}\mathbb{E}\left[\mathbb{1}_{\mathsf{A}}\mathbb{1}_{\{Y_n=k2^{-n}\}}k2^{-n}X\right]=\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y_nX]\xrightarrow{n\to+\infty}\mathbb{E}[\mathbb{1}_{\mathsf{A}}YX]\end{aligned}$$

Hence $\mathbb{E}[\mathbb{1}_{\mathsf{A}}Y\mathbb{E}[X\mid\mathcal{G}]]=\mathbb{E}[\mathbb{1}_{\mathsf{A}}YX]$. In the general case, write $X=X^+-X^-$and $Y=Y^+-Y^-$and exploit the linearity of the conditional expectation.

(iv) The second equality follows from (iii) with $Y=\mathbb{E}[X\mid\mathcal{A}]$ and $X=1$. Now let $\mathsf{A}\in\mathcal{A}$. Then, in particular, $\mathsf{A}\in\mathcal{G}$; hence

$$\mathbb{E}[\mathbb{1}_A \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{A}]] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A \mathbb{E}[X \mid \mathcal{A}]]$$

(v) Exercise

(vi) Exercise

(vii) Exercise

□

Remark 5.4. Theorem 5.14 implies that if $(W_n)_{n\in\mathbb{N}}$ is non-increasing, $(\mathbb{E}[W_n \mid \mathcal{G}])_{n\in\mathbb{N}}$ is also non-increasing.

Example 5.2. Let $X, Y \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$ be independent. Then

$$\mathbb{E}[X + Y \mid Y] = \mathbb{E}[X \mid Y] + \mathbb{E}[Y \mid Y] = \mathbb{E}[X] + Y\,.$$

Example 5.3. Let $X_1, \ldots, X_N$ be independent with $\mathbb{E}[X_i] = 0, i = 1, \ldots, N$. For $n = 1, \ldots, N$, define $\mathcal{G}_n = \sigma(X_1, \ldots, X_n)$ and $S_n = X_1 + \ldots + X_n$. Then, for $n \geq m$,

$$\begin{aligned}\mathbb{E}[S_n \mid \mathcal{G}_m] &= \mathbb{E}[X_1 \mid \mathcal{G}_m] + \ldots + \mathbb{E}[X_n \mid \mathcal{G}_m] \\ &= X_1 + \ldots + X_m + \mathbb{E}[X_{m+1}] + \ldots + \mathbb{E}[X_n] = S_m\,.\end{aligned}$$

By properties of the conditional expectation, since $\sigma(S_m) \subset \mathcal{G}_m$, we have

$$\mathbb{E}[S_n \mid S_m] = \mathbb{E}[\mathbb{E}[S_n \mid \mathcal{G}_m] \mid S_m] = \mathbb{E}[S_m \mid S_m] = S_m\,.$$

Theorem 5.16. *Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be two measurable spaces. Let X be a random variable, valued in $(\mathsf{X}, \mathcal{X})$ and $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-field. If X is independent of $\mathcal{G}$ and Y is $\mathcal{G}$-measurable, it holds for any random variable Y, valued in Y and measurable function $f : \mathsf{X} \times \mathsf{Y} \to \mathbb{R}$ such that $\mathbb{E}[f(X,Y)|\mathcal{G}]$ exists,*

$$\mathbb{E}[f(X,Y)|\mathcal{G}] = h_f(Y)\,, \quad h_f(y) = \mathbb{E}[f(X,y)] \quad \text{for } \mathbb{P}_Y\text{-a.e. } y\,, \tag{5.5}$$

denoting by $\mathbb{P}_Y$ the distribution of Y.

Proof. We only prove the case where $f(X,Y)$ is integrable and positive and leave the other cases as exercises. First, the statement easily holds for $f(x,y) = \mathbb{1}_A(x)\mathbb{1}_B(y)$ with $A \in \mathcal{X}$ and $B \in \mathcal{Y}$ using Theorem 5.15. Then, we can extend it to $f(x,y) = \mathbb{1}_A(x,y)$, for any measurable set $A \in \mathcal{X} \otimes \mathcal{Y}$ considering $\{A \in \mathcal{X} \otimes \mathcal{Y} :$ (5.5) holds for $f = \mathbb{1}_A\}$, showing it is a λ-system, and conclude using the $\pi - \lambda$ theorem. The proof is complete using pointwise and L^1-approximation of positive integrable functions by simple functions Theorem 3.10. □

Theorem 5.17. *(i) (Conditional monotone convergence theorem) Let $(X_n)_{n\in\mathbb{N}}$ is a sequence of non-negative random variables almost surely non-decreasing, converging to X almost surely. Then,*

$$\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}] \ .$$

(ii) (Conditional Fatou's lemma) Let $(X_n)_{n\in\mathbb{N}}$ is a sequence of non-negative random variables. Then,

$$\mathbb{E}\left[\liminf_{n\to+\infty} X_n \middle| \mathcal{G}\right] \le \liminf_{n\to+\infty} \mathbb{E}[X_n|\mathcal{G}] \ .$$

(iii) (Conditional Lebesgue dominated convergence theorem) Assume $Z \in \mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P}), Z \ge 0$ and $(X_n)_{n\in\mathbb{N}}$ is a sequence of random variables with $|X_n| \le Z$ for $n \in \mathbb{N}$ and such that $X_n \xrightarrow{n\to+\infty} X$ as. Then

$$\lim_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{G}] = \mathbb{E}[X \mid \mathcal{G}] \quad \textit{a.s. and in } \mathrm{L}^1(\mathbb{P}) \ .$$

Proof.

Exercise

Exercise

Exercise Define $W_n = \sup_{k\ge n} |X_k - X|$. Then $0 \le W_n \le 2Y$ and $W_n \xrightarrow{\text{a.s.}} 0$. By the Lebesgue's dominated convergence theorem, we have $\mathbb{E}[W_n] \xrightarrow{n\to+\infty} 0$; hence, by the triangle inequality,

$$\mathbb{E}[|\mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]|] \le \mathbb{E}[\mathbb{E}[|X_n - X| \mid \mathcal{G}]] = \mathbb{E}[|X_n - X|] \le \mathbb{E}[W_n] \xrightarrow{n\to+\infty} 0$$

As $(W_n)_{n\in\mathbb{N}}$ is non-increasing, by Theorem 5.14, $(\mathbb{E}[W_n \mid \mathcal{G}])_{n\in\mathbb{N}}$ non-increases to some limit, say, W. By Fatou's lemma,

$$\mathbb{E}[W] \le \lim_{n\to+\infty} \mathbb{E}[\mathbb{E}[W_n \mid \mathcal{G}]] = \lim_{n\to+\infty} \mathbb{E}[W_n] = 0$$

Hence $W = 0$ and thus $\mathbb{E}[W_n \mid \mathcal{G}] \xrightarrow{n\to+\infty} 0$ almost surely. However, by Theorem 5.15,

$$|\mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]| \le \mathbb{E}[W_n \mid \mathcal{G}]$$

□

Theorem 5.18. *If Y is a random variable valued in $(\mathsf{Y},\mathcal{Y})$ and $X \in \mathrm{L}^1(\Omega,\mathcal{F},\mathbb{P})$ (or X non-negative a.s.). Then, there exists a measurable function $h : \mathsf{Y} \to \mathbb{R}\cup\{+\infty\}$ such that $\mathbb{E}[X|Y] = h(Y)$.*

Proof. This is a consequence of Theorem 3.8. □

Remark 5.5. Let $X : \Omega \to \mathbb{R}$ be a random variable such that $X^- \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. We can define the conditional expectation as the monotone limit

$$\mathbb{E}[X \mid \mathcal{G}] = \lim_{n\to+\infty} \mathbb{E}[X_n \mid \mathcal{G}] \,,$$

where $-X^- \leq X_1$ and $X_n \uparrow X$. Due to the monotonicity of the conditional expectation it is easy to show that the limit does not depend on the choice of the sequence $(X_n)_{n\in\mathbb{N}}$ and that it fulfills the conditions of the definition. Analogously, we can define the conditional expectation $X^+ \in \mathrm{L}^1(\Omega, \mathcal{F}, \mathbb{P})$. For this generalization of the conditional expectation, we still have $\mathbb{E}[X \mid \mathcal{G}] \leq \mathbb{E}[Y \mid \mathcal{G}]$ a.s. if $Y \geq X$ a.s.

Corollary 5.19. *Let $\mathcal{G} \subset \mathcal{F}$ be a σ-field and let X be a random variable with $\mathbb{E}\left[X^2\right] < +\infty$. Then $\mathbb{E}[X \mid \mathcal{G}]$ is the orthogonal projection of X on $\mathrm{L}^2(\Omega, \mathcal{G}, \mathbb{P})$. That is, for any $\mathcal{G}$-measurable Y with $\mathbb{E}\left[Y^2\right] < +\infty$,*

$$\mathbb{E}\left[(X-Y)^2\right] \geq \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2\right] \,,$$

with equality if and only if $Y = \mathbb{E}[X \mid \mathcal{G}]$.

Proof. We first show that $\mathbb{E}\left[X^2\right] < +\infty$. This is an immediate consequence of the conditional Jensen inequality Theorem 5.20 but we provide an elementary proof here. For $N \in \mathbb{N}$, define the truncated random variables $|X| \wedge N$. Clearly, we have $\mathbb{E}\left[\mathbb{E}[|X| \wedge N \mid \mathcal{G}]^2\right] \leq N^2$ and using the Tower property, $\mathbb{E}\left[((|X| \wedge N) - \mathbb{E}[|X| \wedge N \mid \mathcal{G}])^2\right] \leq \mathbb{E}[(X \wedge N)^2 | \mathcal{G}]$[(1)]. Using the elementary inequality $a^2 \leq 2(a-b)^2 + 2b^2, a, b \in \mathbb{R}$, we infer

$$\begin{aligned}\mathbb{E}\left[\mathbb{E}[|X| \wedge N \mid \mathcal{G}]^2\right] &\leq 2\mathbb{E}\left[((|X| \wedge N) - \mathbb{E}[|X| \wedge N \mid \mathcal{G}])^2\right] + 2\mathbb{E}\left[(|X| \wedge N)^2\right] \\ &\leq 4\mathbb{E}\left[(|X| \wedge N)^2\right] \leq 4\mathbb{E}\left[X^2\right] \,.\end{aligned}$$

By Theorem 5.17, we get $\mathbb{E}[|X| \wedge N \mid \mathcal{G}] \uparrow \mathbb{E}[|X| \mid \mathcal{G}]$ for $N \to +\infty$. By the triangle inequality and the monotone convergence theorem, we conclude

$$\mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]^2\right] \leq \mathbb{E}\left[\mathbb{E}[|X| \mid \mathcal{G}]^2\right] = \lim_{N\to+\infty} \mathbb{E}\left[\mathbb{E}[|X| \wedge N \mid \mathcal{G}]^2\right] \leq 4\mathbb{E}\left[X^2\right] < +\infty \,.$$

Now let Y be $\mathcal{G}$-measurable and assume $\mathbb{E}\left[Y^2\right] < +\infty$. Then, by the Cauchy-Schwarz inequality, we have $\mathbb{E}[|XY|] < +\infty$. Thus, using the tower property, we infer $\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]Y]$ and $\mathbb{E}[X\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{E}[X\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{G}]] = \mathbb{E}\left[\mathbb{E}[X \mid \mathcal{G}]^2\right]$. Summing up, we have

$$\begin{aligned}&\mathbb{E}\left[(X-Y)^2\right] - \mathbb{E}\left[(X - \mathbb{E}[X \mid \mathcal{G}])^2\right] \\ &\quad= \mathbb{E}\left[X^2 - 2XY + Y^2 - X^2 + 2X\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]^2\right] \\ &\quad= \mathbb{E}\left[Y^2 - 2Y\mathbb{E}[X \mid \mathcal{G}] + \mathbb{E}[X \mid \mathcal{G}]^2\right] \\ &\quad= \mathbb{E}\left[(Y - \mathbb{E}[X \mid \mathcal{G}])^2\right] \geq 0\end{aligned}$$

[(1)]the proof is similar to show that $\mathrm{Var}(X) \leq \mathbb{E}[X^2]$

This completes the proof. □

Next we show Jensen's inequality for conditional expectations.

Theorem 5.20. *Let* $\mathsf{I} \subset \mathbb{R}$ *be an interval, let* $\varphi : I \to \mathbb{R}$ *be convex and let* X *be an* I*-valued random variable on* $(\Omega, \mathcal{F}, \mathbb{P})$*. Further, let* $\mathbb{E}[|X|] < +\infty$ *and let* $\mathcal{G} \subset \mathcal{F}$ *be a* σ*-field. Then*

$$+\infty \geq \mathbb{E}[\varphi(X) \mid \mathcal{G}] \geq \varphi(\mathbb{E}[X \mid \mathcal{G}])\,.$$

Proof. Since by Jensen inequality and $t \mapsto (t)_-$ is convex, we get that $\varphi(X)_-$ is integrable and therefore Remark 5.5 ensures the existence of $\mathbb{E}[\varphi(X) \mid \mathcal{G}]$ with values in $(-\infty, +\infty]$. In addition, since we have $\mathbb{E}[X \mid \mathcal{G}] \in \mathsf{I}$ a.s., hence $\varphi(\mathbb{E}[X \mid \mathcal{G}])$ is well-defined a.s.. Note that $X = \mathbb{E}[X \mid \mathcal{G}]$ almost surely on the event $\{\mathbb{E}[X \mid \mathcal{G}]$ is a boundary point of $\mathsf{I}\}$; hence here the claim is trivial. Indeed, without loss of generality, assume 0 is the left boundary of I and $\mathsf{A} = \{\mathbb{E}[X \mid \mathcal{G}] = 0\}$. As X assumes values in $\mathsf{I} \subset [0, +\infty)$, we have $0 \leq \mathbb{E}[X \mathbb{1}_{\mathsf{A}}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mathbb{1}_{\mathsf{A}}] = 0$; hence $X \mathbb{1}_{\mathsf{A}} = 0$. The case of a right boundary point is similar.

Hence we only need to consider the event $\mathsf{B} = \{\mathbb{E}[X \mid \mathcal{G}]$ is an interior point of $\mathsf{I}\}$. This case can be treated as Theorem 5.3 and is left to the reader.

□

Corollary 5.21. *Let* $p \in [1, +\infty]$ *and let* $\mathcal{G} \subset \mathcal{F}$ *be a sub-*σ*-field. Then the map*

$$\mathrm{L}^p(\Omega, \mathcal{F}, \mathbb{P}) \to \mathrm{L}^p(\Omega, \mathcal{G}, \mathbb{P}), \quad X \mapsto \mathbb{E}[X \mid \mathcal{G}]$$

is a contraction (that is, $\|\mathbb{E}[X \mid \mathcal{G}]\|_p \leq \|X\|_p$*) and thus continuous. Hence, for any sequence* $(X_n)_{n \in \mathbb{N}^*}$ *in* $\mathrm{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ *with* $\lim_{n \to +\infty} \|X_n - X\|_p = 0$*,* $\lim_{n \to +\infty} \|\mathbb{E}[X_n \mid \mathcal{G}] - \mathbb{E}[X \mid \mathcal{G}]\|_p = 0$.

Proof. This is an application of Jensen's inequality with $\varphi(x) = |x|^p$ for $p < +\infty$. For $p = +\infty$, note that $|\mathbb{E}[X \mid \mathcal{G}]| \leq \mathbb{E}[\|X\|_\infty \mid \mathcal{G}] = \|X\|_\infty$. □