Chapter II

Conditional expectation

II.1 Introduction

Let X be a square integrable real-valued random variable. The constant c which minimizes $\mathbb{E}[(X-c)^2]$ is the expectation of X. Indeed, we have, with $\mu = \mathbb{E}[X]$:

$$\mathbb{E}[(X-c)^2] = \mathbb{E}[(X-\mu)^2 + (\mu-c)^2 + 2(X-\mu)(\mu-c)] = \mathrm{Var}(X) + (\mu-c)^2.$$

In some sense, the expectation of X is the best approximation of X by a constant (with a quadratic criterion).

More generally, the conditional expectation of X given another random variable Y will de defined as the best approximation of X by a function of Y. In order to give a precise definition, we shall prove in Section II.2 that the space of square integrable real-valued random variables is an Hilbert space. In Section II.3, we define the conditional expectation of a square integrable random variable as a projection. We then extend the conditional expectation to integrable and non-negative random variables. We provide explicit formulas for discrete and continuous random variables in Section II.4.

II.2 The L^p space

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. For $p \in [1, +\infty)$, let $\mathcal{L}^p(\mathbb{P})$ denote the set of real valued random variables X defined on $(\Omega, \mathcal{F})$ such that $\mathbb{E}[|X|^p] < +\infty$. When there is no ambiguity on the underlying measure, we shall simply write $\mathcal{L}^p$.

Minkowski inequality and the linearity of the expectation yield that $\mathcal{L}^p$ is a vector space and the map $\|\cdot\|_p$ from $\mathcal{L}^p$ to $[0, +\infty)$ defined by $\|X\|_p = \mathbb{E}[|X|^p]^{1/p}$ is a semi-norm. Notice that $\|X\|_p = 0$ implies that a.s. $X = 0$ thanks to Lemma I.44. We deduce that the space $(L^p, \|\cdot\|_p)$, where L^p is the space $\mathcal{L}^p$ quotiented by the equivalence relation "a.s. equal to", is a normed vector space. We shall use the same notation for an element of $\mathcal{L}^p$ and for its equivalent class in L^p. The next proposition asserts that the space L^p is complete and thus it is a Banach space.

A sequence $(X_n, n \in \mathbb{N})$ of elements of L^p is said to converge in L^p to a limit, say X, if $X \in L^p$, and $\lim_{n\to+\infty} \|X_n - X\|_p = 0$.

Proposition II.1. *Let $p \in [1, +\infty)$. The vector space L^p with the norm $\|\cdot\|_p$ is complete: every Cauchy sequence of L^p converges in L^p. That is, for every sequence $(X_n, n \in \mathbb{N})$ of real-valued random variables such that $X_n \in L^p$ for all $n \in \mathbb{N}$ and $\lim_{\min(n,m)\to\infty} \|X_n - X_m\|_p = 0$, there exists a real-valued random variable $X \in L^p$ such that $\lim_{n\to+\infty} \|X_n - X\|_p = 0$.*

Proof. Let $(X_n, n \in \mathbb{N})$ be a Cauchy sequence of L^p, that is $X_n \in L^p$ for all $n \in \mathbb{N}$ and $\lim_{\min(n,m)\to+\infty} \|X_n - X_m\|_p = 0$. Consider the sub-sequence $(n_k, k \in \mathbb{N})$ defined by $n_0 = 0$ and for $k \geq 1$, $n_k = \inf\{m > n_{k-1}; \|X_i - X_j\|_p \leq 2^{-k}$ for all $i \geq m, j \geq m\}$. In particular, we have $\left\|X_{n_{k+1}} - X_{n_k}\right\|_p \leq 2^{-k}$ for all $k \geq 1$. Minkowski inequality and the monotone convergence imply that $\left\|\sum_{k\in\mathbb{N}} |X_{n_{k+1}} - X_{n_k}|\right\|_p < +\infty$ and thus $\sum_{k\in\mathbb{N}} |X_{n_{k+1}} - X_{n_k}|$ is a.s. finite. The series with general term $(X_{n_{k+1}} - X_{n_k})$ is a.s. absolutely converging. By considering the convergence of the partial sums, we get that the sequence $(X_{n_k}, k \in \mathbb{N})$ converges a.s. towards a limit, say X. This limit is a real-valued random variable, thanks to Corollary I.47. We deduce from Fatou lemma that:

$$\|X_m - X\|_p \leq \liminf_{k\to+\infty} \|X_m - X_{n_k}\|_p .$$

This implies that $\lim_{m\to+\infty} \|X_m - X\|_p = 0$, and Minkowski inequality gives that $X \in L^p$. □

We give an elementary criterion for the L^p convergence for a.s. converging sequences.

Lemma II.2. *Let $p \in [1, +\infty)$. Let $(X_n, n \in \mathbb{N})$ be a sequence or real-valued random variables belonging to L^p which converges a.s. towards X. The convergence holds in L^p (i.e. $\lim_{n\to+\infty} \|X - X_n\|_p = 0$) if and only if $\lim_{n\to+\infty} \|X_n\|_p = \|X\|_p$.*

Proof. Assume $\lim_{n\to+\infty} \|X - X_n\|_p = 0$. Using Minkowski inequality, we deduce that $\left|\|X\|_p - \|X_n\|_p\right| \leq \|X - X_n\|_p$. This proves that $\lim_{n\to+\infty} \|X_n\|_p = \|X\|_p$.

On the other hand, assume that $\lim_{n\to+\infty} \|X_n\|_p = \|X\|_p$. Set $Y_n = 2^{p-1}(|X_n|^p + |X|^p)$ and $Y = 2^p|X|^p$. Since the function $x \mapsto |x|^p$ is convex, we get $|X_n - X|^p \leq Y_n$ for all $n \in \mathbb{N}$. We also have $\lim_{n\to+\infty} Y_n = Y$ a.s. and $\lim_{n\to+\infty} \mathbb{E}[Y_n] = \mathbb{E}[Y] < +\infty$. The dominated convergence theorem I.51 gives then that $\lim_{n\to+\infty} \mathbb{E}[|X_n - X|^p] = \mathbb{E}[\lim_{n\to+\infty} |X_n - X|^p] = 0$. This ends the proof. □

Consider the case $p = 2$. The bilinear form $\langle\cdot,\cdot\rangle_{L^2}$ on L^2 defined by $\langle X, Y\rangle_{L^2} = \mathbb{E}[XY]$ is the scalar product corresponding to the norm $\|\cdot\|_2$. The space L^2 with the product scalar $\langle\cdot,\cdot\rangle_{L^2}$ is thus an Hilbert space. Notice that square-integrable real-valued random variables which are independent are orthogonal for the scalar product $\langle\cdot,\cdot\rangle_{L^2}$.

We shall consider the following results on projection.

Theorem II.3. *Let H be a closed vector sub-space of L^2 and $X \in L^2$.*

1. *(Existence.) There exists a real-valued random variable $X_H \in H$, called the orthogonal projection of X on H, such that $\mathbb{E}[(X - X_H)^2] = \inf\{\mathbb{E}[(X - Y)^2]; Y \in H\}$. And, for all $Y \in H$, we have $\mathbb{E}[XY] = \mathbb{E}[X_H Y]$.*

2. *(Uniqueness.) Let $Z \in H$ such that $\mathbb{E}[(X-Z)^2] = \inf\{\mathbb{E}[(X-Y)^2];\, Y \in H\}$ or such that $\mathbb{E}[ZY] = \mathbb{E}[XY]$ for all $Y \in H$. Then, we have that a.s. $Z = X_H$.*

Proof. We set $a = \inf\{\mathbb{E}[(X-Y)^2]; Y \in H\}$. Using the median formula, it is easy to get that for all $Y', Z' \in L^2$:

$$\mathbb{E}[(Z'-Y')^2] + \mathbb{E}[(Z'+Y')^2] = 2\mathbb{E}[Z'^2] + 2\mathbb{E}[Y'^2].$$

Let $(X_n, n \in \mathbb{N})$ be a sequence of H such that $\lim_{n\to+\infty} \mathbb{E}[(X-X_n)^2] = a$. Using the median formula with $Z' = X_n - X$ and $Y' = X_m - X$, we get:

$$\mathbb{E}[(X_n - X_m)^2] = 2\mathbb{E}[(X-X_n)^2] + 2\mathbb{E}[(X-X_m)^2] - 4\mathbb{E}[(X-I)^2],$$

with $I = (X_n + X_m)/2 \in H$. As $\mathbb{E}[(X-I)^2] \geq a$, we deduce that the sequence $(X_n, n \in \mathbb{N})$ is a Cauchy sequence in L^2 and thus converge, say towards X_H. Since H is closed, we get the limit X_H belongs to H.

Let $Y \in H$ and $Z \in H$ be such that $\mathbb{E}[(X-Z)^2] = a$. The function $t \mapsto \mathbb{E}[(X-Z+tY)^2] = a + t\mathbb{E}[(X-Z)Y] + t^2\mathbb{E}[Y^2]$ is minimal for $t = 0$. This implies that its derivative at $t = 0$ is zero, that is $\mathbb{E}[(X-Z)Y] = 0$. In particular, we have $\mathbb{E}[(X-X_H)Y] = 0$. This ends the proof of property 1.

On the one hand, let $Z \in H$ be such that $\mathbb{E}[(X-Z)^2] = a$. We deduce from the previous arguments that for all $Y \in H$:

$$\mathbb{E}[(X_H - Z)Y] = \mathbb{E}[(X-Z)Y] - \mathbb{E}[(X-X_H)Y] = 0.$$

Taking $Y = (X_H - Z)$, gives that $\mathbb{E}[(X_H - Z)^2] = 0$ and thus a.s. $Z = X_H$, see Lemma I.44.

On the other hand, if there exists $Z \in H$ such that $\mathbb{E}[ZY] = \mathbb{E}[XY]$ for all $Y \in H$, arguing as above, we deduce that a.s. $Z = X_H$. □

According to the remarks at the beginning of paragraph II.1, we see that if X is a real-valued square-integrable random variable, then $\mathbb{E}[X]$ can be seen as the orthogonal projection of X on the vector space of the constant random variables.

II.3 Conditional expectation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let $\mathcal{H} \subset \mathcal{F}$ be a σ-field. We recall that a random variable X (which is by definition $\mathcal{F}$-measurable) is $\mathcal{H}$-measurable if $\sigma(X)$, the σ-field generated by X, is a subset of $\mathcal{H}$, see Corollary I.26. The expectation of X conditionally on $\mathcal{H}$ corresponds to the best "approximation" of X by an $\mathcal{H}$-mesurable random variable.

Notice that if X is a real-valued random variable such that $\mathbb{E}[X]$ is well defined, then $\mathbb{E}[X\mathbf{1}_A]$ is also well defined for any $A \in \mathcal{F}$.

Definition II.4. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. We say that a real-valued $\mathcal{H}$-measurable random variable Z such that $\mathbb{E}[Z]$ is well defined is the expectation of X conditionally on $\mathcal{H}$ if:*

$$\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[Z\mathbf{1}_A] \quad \text{for all } A \in \mathcal{H}. \tag{II.1}$$

The next lemma asserts that, if the expectation of X conditionally on $\mathcal{H}$ exists then it is unique up to an a.s. equality. It will be denoted by $\mathbb{E}[X|\mathcal{H}]$.

Lemma II.5 (Uniqueness of the conditional expectation)**.** *Let Z and Z' be real-valued random variables $\mathcal{H}$-measurable such that $\mathbb{E}[Z]$ and $\mathbb{E}[Z']$ are well defined and such that $\mathbb{E}[Z\mathbf{1}_A] = \mathbb{E}[Z'\mathbf{1}_A]$ for all $A \in \mathcal{H}$. Then, we get that a.s. $Z = Z'$.*

Proof. Let $n \in \mathbb{N}$ and consider $A = \{n \geq Z > Z' \geq -n\}$ which belongs to $\mathcal{H}$. By linearity, we deduce from the hypothesis that $\mathbb{E}[(Z - Z')\mathbf{1}_{\{n \geq Z > Z' \geq -n\}}] = 0$. Lemma I.44 implies that $\mathbb{P}(n \geq Z > Z' \geq -n) = 0$ and thus $\mathbb{P}(+\infty > Z > Z' > -\infty) = 0$ by monotone convergence. Considering $A = \{Z = +\infty, n \geq Z'\}$, $A = \{Z \geq n, Z' = -\infty\}$ and $A = \{Z = +\infty, Z' = -\infty\}$ leads eventually to $\mathbb{P}(Z > Z', Z = +\infty \text{ or } Z' = -\infty)$. So we get $\mathbb{P}(Z > Z') = 0$. By symmetry, we deduce that a.s. $Z = Z'$. □

We use the orthogonal projection theorem on Hilbert spaces, to define the conditional expectation for square-integrable real-valued random variables.

Proposition II.6. *If $X \in L^2$, then $\mathbb{E}[X|\mathcal{H}]$ is the orthogonal projection defined in Proposition II.3, of X on the vector space H of all square-integrable $\mathcal{H}$-measurable random variables.*

Proof. The set H corresponds to the space L^2 associated to the probability space $(\Omega, \mathcal{H}, \mathbb{P})$. Thus it is closed thanks to Proposition II.1. The set H is thus a closed vector subspace of L^2. Property 1 from Theorem II.3 implies then that the orthogonal projection of $X \in L^2$ on H is the expectation of X conditionally on $\mathcal{H}$. □

We have the following properties.

Proposition II.7. *Let X and Y be real-valued square-integrable random variables.*

1. *Positivity. If a.s. $X \geq 0$ then we have that a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$.*

2. *Linearity. For $a, b \in \mathbb{R}$, we have that a.s. $\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}]$.*

3. *Monotone convergence. Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued square integrable random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n \leq X_{n+1}$. Then, we have that a.s.:*

$$\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to+\infty} X_n \middle| \mathcal{H}\right].$$

Proof. Let X be square-integrable and a.s. non-negative. We set $A = \{\mathbb{E}[X|\mathcal{H}] < 0\}$. We have:

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

where we used that $A \in \mathcal{H}$ and (II.1) for the equality. We deduce that $\mathbb{E}[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A] = 0$ and thus a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$ according to Lemma I.44.

The linearity property is a consequence of the linearity property of the expectation, (II.1) and Lemma II.5.

Let $(X_n, n \in \mathbb{N})$ be a sequence of real-valued square-integrable random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n \leq X_{n+1}$. We deduce from the linearity and positivity properties of the conditional expectation that for all $n \in \mathbb{N}$ a.s. $0 \leq \mathbb{E}[X_n|\mathcal{H}] \leq \mathbb{E}[X_{n+1}|\mathcal{H}]$. The random-variable $Z = \lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}]$ is $\mathcal{H}$-measurable according to Corollairy I.47 and a.s. non-negative. The monotone convergence theorem implies that for all $A \in \mathcal{H}$:

$$\mathbb{E}\left[Z\mathbf{1}_A\right] = \lim_{n\to+\infty} \mathbb{E}\left[\mathbb{E}[X_n|\mathcal{H}]\mathbf{1}_A\right] = \lim_{n\to+\infty} \mathbb{E}\left[X_n\mathbf{1}_A\right] = \mathbb{E}\left[\lim_{n\to+\infty} X_n\mathbf{1}_A\right].$$

We deduce from (II.1) and Lemma II.5 that $Z = \mathbb{E}[\lim_{n\to+\infty} X_n|\mathcal{H}]$, which ends the proof. □

We extend the definition of conditional expectations to random variables whose expectation is well defined.

Proposition II.8. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. Then its expectation conditionally on $\mathcal{H}$, $\mathbb{E}[X|\mathcal{H}]$, exists and is unique up to an a.s. equality. Furthermore the expectation of $\mathbb{E}[X|\mathcal{H}]$ is well defined and is equal to $\mathbb{E}[X]$:*

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\right] = \mathbb{E}[X]. \tag{II.2}$$

If X is non-negative a.s. (resp. integrable), so is $\mathbb{E}[X|\mathcal{H}]$.

Proof. Consider first the case where X is a.s. non-negative. The random variable X is the a.s. limit of a sequence of positive square-integrable random variables. Property 3 from Proposition II.7 implies that $\mathbb{E}[X|\mathcal{H}]$ exists. It is unique thanks to Lemma II.5. It is a.s. non-negative as limit of non-negative random variables. Taking $A = \Omega$ in (II.1), we get (II.2).

In the general case, recall that $X^+ = \max(X, 0)$ and $X^- = \max(-X, 0)$. From the previous argument the expectations of $\mathbb{E}[X^+|\mathcal{H}]$ and $\mathbb{E}[X^-|\mathcal{H}]$ are well defined and respectively equal to $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$. Since one of those two expectation is finite, we deduce that a.s. $\mathbb{E}[X^+|\mathcal{H}]$ if finite or a.s. $\mathbb{E}[X^-|\mathcal{H}]$ is finite. Then use (II.1) and Lemma II.5 to deduce that $\mathbb{E}[X^+|\mathcal{H}] - \mathbb{E}[X^-|\mathcal{H}]$ is equal to $\mathbb{E}[X|\mathcal{H}]$, the expectation, of X conditionally on $\mathcal{H}$. Since $\mathbb{E}[X|\mathcal{H}]$ is the difference of two non-negative random variable, one of them being integrable, we deduce that the expectation of $\mathbb{E}[X|\mathcal{H}]$ is well defined and use (II.1) with $A = \Omega$ to get (II.2). Eventually, if X is integrable, so are $\mathbb{E}[X^+|\mathcal{H}]$ and $\mathbb{E}[X^-|\mathcal{H}]$ thanks to (II.2) for non-negative random variables. This implies that $\mathbb{E}[X|\mathcal{H}]$ is integrable. □

We summarize in the next proposition the properties of the conditional expectation directly inherited from the properties of the expectation.

Proposition II.9. *We have the following properties.*

1. *Positivity. If X is a real-valued random variable such that a.s. $X \geq 0$, then a.s. $\mathbb{E}[X|\mathcal{H}] \geq 0$.*

2. *Linearity. For $a, b \in \mathbb{R}$, X, Y real-valued integrable random-variables, we have $\mathbb{E}[aX + bY|\mathcal{H}] = a\mathbb{E}[X|\mathcal{H}] + b\mathbb{E}[Y|\mathcal{H}]$.*

3. *Monotony. For X, Y real-valued integrable random variables such that a.s. $X \leq Y$, we have $\mathbb{E}[X|\mathcal{H}] \leq \mathbb{E}[Y|\mathcal{H}]$.*

4. *Monotone convergence. Let $(X_n, n \in \mathbb{N})$ be real-valued random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n \leq X_{n+1}$. Then we have that a.s.:*

$$\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to+\infty} X_n \middle| \mathcal{H}\right].$$

5. *Fatou Lemma. Let $(X_n, n \in \mathbb{N})$ be real-valued random variables such that for all $n \in \mathbb{N}$ a.s. $0 \leq X_n$. Then we have that a.s.:*

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

6. *Dominated convergence. Let X, Y, $(X_n, n \in \mathbb{N})$ and $(Y_n, n \in \mathbb{N})$ be real-valued random variables such that for all $n \in \mathbb{N}$ a.s. $|X_n| \leq Y_n$, a.s. $\lim_{n\to+\infty} X_n = X$ and a.s. $\lim_{n\to+\infty} Y_n = Y$. We assume that a.s. $\lim_{n\to+\infty} \mathbb{E}[Y_n|\mathcal{H}] = \mathbb{E}[Y|\mathcal{H}]$ and a.s. $\mathbb{E}[Y|\mathcal{H}] < +\infty$. Then we have that a.s.:*

$$\lim_{n\to+\infty} \mathbb{E}[X_n|\mathcal{H}] = \mathbb{E}\left[\lim_{n\to+\infty} X_n \middle| \mathcal{H}\right].$$

7. *The Tchebychev, Hölder, Cauchy-Schwarz, Minkowski and Jensen inequalities from Propositions I.54 and I.58 holds with the expectation replaced by the conditional expectation.*

For example, we state Jensen inequality from property 7 above. Let φ be a real-valued convex function defined on $\mathbb{R}^d$. Let X be an integrable $\mathbb{R}^d$-valued random variable. Then, we have by Jensen inequality that $\mathbb{E}[\varphi(X)]$ is well defined and a.s.:

$$\mathbb{E}\left[\varphi(X)|\,\mathcal{H}\right] \geq \varphi(\mathbb{E}\left[X|\,\mathcal{H}\right]). \tag{II.3}$$

Furthermore, if φ is strictly convex, the inequality in (II.3) is an equality if and only if X is $\mathcal{H}$-measurable.

Proof. The positivity property comes from Proposition II.8. The linearity property comes from the linearity of the expectation, (II.1) and Lemma II.5. The monotony property is a consequence of the positivity and linearity properties. The proof of the monotone convergence theorem is based on the same arguments as in the proof of Proposition II.7. Fatou Lemma and the dominated convergence theorem are consequences of the monotone convergence theorem, see proof of Lemma I.50 and of Theorem I.51. The proof of the inequalities is similar to the proof of Propositions I.54 and I.58. □

Using the monotone or dominated convergence theorems, it is easy to prove the following Corollary which generalizes (II.1).

Corollary II.10. *Let X and Y be two real-valued random variables. We assume that Y is $\mathcal{H}$-measurable. We have:*

$$\mathbb{E}[XY] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]Y\right], \tag{II.4}$$

if either X and Y are square integrable, or if X is integrable and Y a.s. bounded, or if X and Y are a.s. non-negative.

We say that a random variable X is independent from a σ-field $\mathcal{H}$ if the two σ-fields $\sigma(X)$ and $\mathcal{H}$ are independent. Equivalently X is independent of $\mathcal{H}$ if it is independent of $\mathbf{1}_A$ for all $A \in \mathcal{H}$. We complete the properties of the conditional expectation.

Proposition II.11. *Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined.*

1. *If X is $\mathcal{H}$-measurable, then we have that a.s. $\mathbb{E}[X|\mathcal{H}] = X$.*
2. *If X is independent of $\mathcal{H}$, then we have that a.s. $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[X]$.*
3. *If Y is a bounded real-valued $\mathcal{H}$-measurable random variable and X is integrable, then we have that a.s. $\mathbb{E}[YX|\mathcal{H}] = Y\mathbb{E}[X|\mathcal{H}]$.*
4. *If $\mathcal{G} \subset \mathcal{H}$ is a σ-field, then we have that a.s. $\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}\right] = \mathbb{E}[X|\mathcal{G}]$.*
5. *If $\mathcal{G} \subset \mathcal{F}$ is a σ-field independent of $\mathcal{H} \vee \sigma(X)$, then we have that a.s. $\mathbb{E}[X|\mathcal{G} \vee \mathcal{H}] = \mathbb{E}[X|\mathcal{H}]$.*

Proof. If X is $\mathcal{H}$-measurable, then we can choose $Z = X$ in (II.1) and use Lemma II.5 to conclude. If X is independent of $\mathcal{H}$, then for all $A \in \mathcal{H}$, we have $\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[X]\mathbb{E}[\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X]\mathbf{1}_A]$, and we can choose $Z = \mathbb{E}[X]$ in (II.1) and use Lemma II.5 to conclude. If Y a bounded real-valued $\mathcal{H}$-measurable random variable, then according to (II.4), we have for $A \in \mathcal{H}$, $\mathbb{E}[XY\mathbf{1}_A] = \mathbb{E}[\mathbb{E}[X|\mathcal{H}]Y\mathbf{1}_A]$, and we can choose $Z = Y\mathbb{E}[X|\mathcal{H}]$ in (II.1) and use Lemma II.5 to conclude.

We prove property 4. Let $A \in \mathcal{G} \subset \mathcal{H}$. We have:

$$\mathbb{E}\left[\mathbb{E}[X|\mathcal{G}]\mathbf{1}_A\right] = \mathbb{E}[X\mathbf{1}_A] = \mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]\mathbf{1}_A\right] = \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}[X|\mathcal{H}]|\mathcal{G}\right]\mathbf{1}_A\right].$$

where we used (II.1) with $\mathcal{H}$ replaced by $\mathcal{G}$ for the first equality, (II.1) for the second and (II.1) with $\mathcal{H}$ replaced by $\mathcal{G}$ and X by $\mathbb{E}[X|\mathcal{H}]$ for the last. Then we deduce property 4 from Definition II.4 and Lemma II.5.

We prove property 5 first when X is integrable. For $A \in \mathcal{G}$ and $B \in \mathcal{H}$, we have:

$$\mathbb{E}\left[\mathbf{1}_{A\cap B}X\right] = \mathbb{E}\left[\mathbf{1}_A\mathbf{1}_B X\right] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B X] = \mathbb{E}[\mathbf{1}_A]\mathbb{E}[\mathbf{1}_B\mathbb{E}[X|\mathcal{H}]] = \mathbb{E}[\mathbf{1}_A\mathbf{1}_B\mathbb{E}[X|\mathcal{H}]],$$

where we used that $\mathbf{1}_A$ is independent of $\mathcal{H}\vee\sigma(X)$ in the second equality and independent of $\mathcal{H}$ in the last. Using the dominated convergence theorem, we get that $\mathcal{A} = \{A \in \mathcal{F}, \mathbb{E}\left[\mathbf{1}_A X\right] = \mathbb{E}[\mathbf{1}_A\mathbb{E}[X|\mathcal{H}]]\}$ is a monotone class. It contains $\mathcal{C} = \{A \cap B;\ A \in \mathcal{G}, B \in \mathcal{H}\}$ which is stable by finite intersection. The monotone class theorem implies that $\mathcal{A}$ contains $\sigma(\mathcal{C})$ and thus $\mathcal{G} \vee \mathcal{H}$. Then we deduce property 5 from Definition II.4 and Lemma II.5. Use the monotone convergence theorem to extend the result to non-negative random variable and use that $\mathbb{E}[X|\mathcal{H}'] = \mathbb{E}[X^+|\mathcal{H}'] - \mathbb{E}[X^-|\mathcal{H}']$ for any σ-field $\mathcal{H}' \subset \mathcal{F}$ to extend the result to any real random variable X such that $\mathbb{E}[X]$ is well defined. □

We extend the definition of conditional expectation to $\mathbb{R}^d$-valued random variables.

Definition II.12. *Let $d \in \mathbb{N}^*$. Let $X = (X_1, \ldots, X_d)$ be an $\mathbb{R}^d$-valued integrable random variable. The conditional expectation of X conditionally on $\mathcal{H}$, denoted by $\mathbb{E}[X|\mathcal{H}]$, is given by $(\mathbb{E}[X_1|\mathcal{H}], \ldots, \mathbb{E}[X_d|\mathcal{H}])$.*

II.4 Conditional expectation with respect to a random variable

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let V be a random variable taking values in a measurable space $(E, \mathcal{S})$. Recall that $\sigma(V)$ denote the σ-field generated by V. Let X be a real-valued random variable. We write $\mathbb{E}[X|V]$ for $\mathbb{E}[X|\sigma(V)]$. As $\mathbb{E}[X|V]$ is a $\sigma(V)$-mesurable random variable, we deduce from Proposition I.33 there exists a real-valued measurable function g defined on $(E, \mathcal{E})$ such that $\mathbb{E}[X|V] = g(V)$. In the next two paragraphs we give an explicit formula for g when V is a discrete random variable and when $X = \varphi(Y, V)$ with Y some random variable taking values in a measurable space $(S, \mathcal{S})$ such that (Y, V) has a density with respect to some product measure on $S \times E$.

If $A \in \mathcal{F}$, we denote the probability of A conditionally on V by:

$$\mathbb{P}(A|V) = \mathbb{E}[\mathbf{1}_A|V]. \tag{II.5}$$

II.4.1 The discrete case

Let $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$. By considering the probability measure:

$$\frac{1}{\mathbb{P}(B)} \mathbf{1}_B \, \mathbb{P} : A \mapsto \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

see Corollary I.49, we define the expectation conditionally on B by, for all real-valued random variable Y such that $\mathbb{E}[Y]$ is well defined:

$$\mathbb{E}[Y|B] = \frac{\mathbb{E}[Y\mathbf{1}_B]}{\mathbb{P}(B)}. \tag{II.6}$$

The following corollary provides an explicit formula for the expectation conditionally on a discrete random variable.

Corollary II.13. *Let $(E, \mathcal{E})$ be a measurable space, and assume that all the singletons of E are measurable. Let V be a discrete random variable taking values in $(E, \mathcal{E})$, that is $\mathbb{P}(V \in \Delta_V) = 1$ where $\Delta_V = \{v \in E, \, \mathbb{P}(V = v) > 0\}$. Let X be a real-valued random variable such that $\mathbb{E}[X]$ is well defined. Then, we have that a.s. $\mathbb{E}[X|V] = g(V)$ with:*

$$g(v) = \frac{\mathbb{E}[X\mathbf{1}_{\{V=v\}}]}{\mathbb{P}(V = v)} = \mathbb{E}[X|V = v] \quad \text{for } v \in \Delta_V, \quad \text{and} \quad g(v) = 0 \quad \text{otherwise}. \tag{II.7}$$

Proof. Set $g(V) = \mathbb{E}[X|V]$, with g measurable. We deduce from (II.1) with $A = \{V = v\}$ that $\mathbb{E}[X\mathbf{1}_{\{V=v\}}] = g(v)\mathbb{P}(V = v)$. If $\mathbb{P}(V = v) > 0$, we get:

$$g(v) = \frac{\mathbb{E}[X\mathbf{1}_{\{V=v\}}]}{\mathbb{P}(V = v)} = \mathbb{E}[X|V = v].$$

The value of $\mathbb{E}[X|V = v]$ when $\mathbb{P}(V = v) = 0$ is unimportant, and it is set to 0. □

Exercise II.1. Let $A, B \in \mathcal{F}$ such that $\mathbb{P}(B) \in (0, 1)$. Compute $\mathbb{E}[\mathbf{1}_A|\mathbf{1}_B]$. △

Denote by $\mathbb{P}_v$ the probability measure on $(\Omega, \mathcal{F})$ defined by $\mathbb{P}_v(A) = \mathbb{P}(A|V = v)$ for $A \in \mathcal{F}$. The law of X conditionally on $\{V = v\}$, denoted by $\mathrm{P}_{X|v}$ is the image of the probability measure $\mathbb{P}_v$ by X, and we define the law of X conditionally on V as the collection of probability measure $\mathrm{P}_{X|V} = (\mathrm{P}_{X|v}, v \in \Delta_V)$. An illustration is given in the next example.

Example II.14. Let $(X_i, i \in [\![1, n]\!])$ be independent Bernoulli random variables with the same parameter $p \in (0, 1)$. We set $S_n = \sum_{i=1}^n X_i$, which has a binomial distribution with parameter (n, p). We shall compute the law of X_1 conditionally on S_n. W get for $k \in [\![1, n]\!]$:

$$\mathbb{P}(X_1 = 1|S_n = k) = \frac{\mathbb{P}(X_1 = 1, S_n = k)}{\mathbb{P}(S_n = k)} = \frac{\mathbb{P}(X_1 = 1)\mathbb{P}(X_2 + \cdots + X_n = k - 1)}{\mathbb{P}(S_n = k)} = \frac{k}{n},$$

where we used independence for X_1 and $(X_2, \ldots, X_n)$ for the second equality and that $X_2 + \cdots + X_n$ has a binomial distribution with parameter $(n - 1, p)$ for the last. For $k = 0$, we get directly that $\mathbb{P}(X_1 = 1|S_n = k) = 0$. We deduce that X_1 conditionally on $\{S_n = k\}$ is a Bernoulli with parameter k/n for all $k \in [\![0, n]\!]$. We shall say that, conditionally on S_n, X_1 has the Bernoulli distribution with parameter S_n/n.

Using Corollary II.13, we get that $\mathbb{E}[X_1|S_n] = S_n/n$, which could have been obtained directly as the expectation of a Bernoulli random variable is given by its parameter. △

II.4.2 The continuous case

Let Y be a random variable taking values in $(S, \mathcal{S})$ such that (Y, V) has a density with respect to some product measure on $S \times E$. More precisely the probability distribution of (Y, V) is given by $f_{(Y,V)}(y, v)\, \mu(dy)\nu(dv)$, where μ and ν are respectively measures on $(S, \mathcal{S})$ and $(E, \mathcal{E})$ and $f_{Y,V}$ is a $[0, +\infty]$-valued measurable function such that $\int f_{(Y,V)}\, \mu\otimes\nu = 1$. In this setting, we give a closed formula for $\mathbb{E}[X|V]$ when $X = \varphi(Y, V)$, with φ a real-valued measurable function defined on $S \times E$ endowed with the product σ-field.

According to Fubini theorem, V has probability distribution $f_V\nu$ with density (with respect to the measure ν) given by $f_V(v) = \int f_{(Y,V)}(y, v)\, \mu(dy)$ and Y has probability distribution $f_Y\mu$ with density (with respect to the measure μ) given by $f_Y(y) = \int f_{(Y,V)}(y, v)\, \nu(dv)$. We now define the law of Y conditionally on V.

Definition II.15. *The probability distribution of Y conditionally on $\{V = v\}$, with $v \in E$ such that $f_V(v) > 0$, is given by $f_{Y|V}(y|v)\, \mu(dy)$ with density $f_{Y|V}$ (with respect to the measure μ) given by:*

$$f_{Y|V}(y|v) = \frac{f_{(Y,V)}(y, v)}{f_V(v)}, \quad y \in S.$$

By convention, we set $f_{Y|V}(y|v) = 0$ *if* $f_V(v) = 0$*.*

Thanks to Fubini theorem, we get that, for v such that $f_V(v) > 0$, the function $y \mapsto f_{Y|V}(y|v)$ is a density as it is non-negative and we have that $\int f_{Y|V}(y|v)\, \mu(dy) = 1$.

We now give the expectation of $X = \varphi(Y, V)$, for some function φ, conditionally on V.

Proposition II.16. *Let* $(E, \mathcal{E}, \nu)$ *and* $(S, \mathcal{S}, \mu)$ *be measured space. Let* (Y, V) *be an* $S \times E$*-valued random variable with density* $(y, v) \mapsto f_{(Y,V)}(y, v)$ *with respect to the product measure* $\mu(dy)\nu(dv)$*. Let* φ *be a real-valued measurable function defined on* $S \times E$ *and set* $X = \varphi(Y, V)$*. Assume that* $\mathbb{E}[X]$ *is well defined. Then we have that a.e.* $\mathbb{E}[X|V] = g(V)$*, with:*

$$g(v) = \int \varphi(y, v) f_{Y|V}(y|v)\, \mu(dy). \tag{II.8}$$

Proof. Let $A \in \sigma(V)$. The function $\mathbf{1}_A$ is $\sigma(V)$-measurable, and thus, thanks to Proposition I.33, there exists a measurable function h such that $\mathbf{1}_A = h(V)$. Using that $f_{(Y,V)}(y, v) = f_{Y|V}(y|v) f_V(v)$, Fubini theorem and g given by (II.8), we get:

$$\begin{aligned}
\mathbb{E}[X\mathbf{1}_A] = \mathbb{E}[\varphi(Y, V)h(V)] &= \int \varphi(y, v)h(v) f_{(Y,V)}(y, v)\, \mu(dy)\nu(dv) \\
&= \int h(v) \left(\int \varphi(y, v) f_{Y|V}(y|v)\, \mu(dy) \right) f_V(v)\, \nu(dv) \\
&= \int h(v) g(v) f_V(v)\, \nu(dv) \\
&= \mathbb{E}[g(V)h(V)] = \mathbb{E}[g(V)\mathbf{1}_A].
\end{aligned}$$

Using (II.1) and Lemma II.5, we deduce that a.s. $g(V) = \mathbb{E}[X|V]$. □

Exercise II.2. Let (Y, V) be an $\mathbb{R}^2$-valued random variable whose law has density with respect to the Lebesgue measure on $\mathbb{R}^2$ given by $f_{(Y,V)}(y, v) = \lambda v^{-1} \mathrm{e}^{-\lambda v}\, \mathbf{1}_{\{0<y<v\}}$. Check that the law of Y conditionally on V is the uniform distribution on $[0, V]$. For a real-valued measurable bounded function φ defined on $\mathbb{R}$, deduce that $\mathbb{E}[\varphi(Y)|V] = V^{-1} \int_0^V \varphi(y)\, dy$.

△