Note 2

Conditional expectation.

see also A. Bovier's script for SS17, Chapter 2 [pdf].

1 Motivation

Recall the elementary definition of conditional probability of the event $\{Y = y\}$ given the event $\{X = x\}$ for a pair of discrete random variables X, Y:

$$\mathbb{P}(Y = y|X = x) := \frac{\mathbb{P}(Y = y, X = x)}{\mathbb{P}(X = x)}, \text{ if } \mathbb{P}(X = x) > 0. \tag{1}$$

Conditioning the original probability $\mathbb{P}$ on the event $\{X = x\}$ gives rise to a new probability $\mathbb{P}(\cdot|X = x)$ provided the event $\{X = x\}$ has a positive probability to happen. We could also consider the associated *conditional expectation* of any (bounded, measurable) function $f(Y)$ of Y, and denote it by

$$\mathbb{E}[f(Y)|X = x] = \sum_y f(y)\mathbb{P}(Y = y|X = x).$$

These elementary definitions cannot be easily generalised to the case where the random variable X is not discrete, because it could happen that all the events of the form $\{X = x\}$ are of zero probability and therefore eq. (1) does not make sense.

The standard way out of the problem is to generalise the notion of conditional expectation and then derive a notion of conditional probability as a by-product, the generalisation goes via considering the conditional value not as a deterministic quantity but as a random quantity itself, namely we will make the conditional expectation depend on the elementary event $\omega \in \Omega$ itself.

Somehow we would like to see the conditional expectation of $f(Y)$ with respect to X as our best prediction of $f(Y)$ given the informations contained in the observation of X (without specifying which value of X has been actually observed). If we note it as

$$\mathbb{E}[f(Y)|X],$$

it is natural to assume that this quantity depends on the outcome of X, therefore that there exists a function $u: \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[f(Y)|X] = u(X)$, in such a way that in the discrete setting we would have

$$u(x) = \mathbb{E}[f(Y)|X = x].$$

In order to find a condition on the function u let us observe that in the discrete setting we have

$$\mathbb{E}[u(X)h(X)] = \sum_{x:\mathbb{P}(X=x)>0} h(x)u(x)\mathbb{P}(X = x) = \sum_{x,y} h(x)f(y)\mathbb{P}(Y = y, X = x) \tag{2}$$

for all $h: \mathbb{R} \to \mathbb{R}$ measurable and bounded. This equality can be stated in general as

$$\mathbb{E}[h(X)\underbrace{u(X)}_{=\mathbb{E}[f(Y)|X]}] = E[h(X)f(Y)], \qquad \forall h. \tag{3}$$

This family of equalities will play the role of our definition of the conditional expectation $\mathbb{E}[f(Y)|X]$. Indeed note that if g is another function such that $\mathbb{E}[h(X)g(X)] = \mathbb{E}[h(X)f(Y)]$ for all h bounded and measurable, then letting $r(x) = g(x) - u(x)$ and choosing $h(x) = \operatorname{sign} r(x)$ we have $\mathbb{E}[|r(X)|] = 0$ which implies $g(x) = u(x)$ whenever $\mathbb{P}(X = x) > 0$. Therefore $\mathbb{P}(g(X) \neq u(X)) = 0$ and the condition (3) identifies $u(X)$ almost surely.

If $X: \Omega \to \{x_1, x_2, \ldots\}$ is a discrete random variable and $A_k = \{X = x_k\} = \{\omega \in \Omega: X(\omega) = x_k\}$, then $\sigma(X) = \sigma(A_1, A_2, \ldots)$. In this case the conditional expectation $Z = \mathbb{E}[f(Y)|X]$ satisfies

$$Z(\omega) = u(X(\omega)) = u(x_k) = \sum_y f(y) \frac{\mathbb{P}(A_k, Y = y)}{\mathbb{P}(A_k)} = \frac{\mathbb{E}[f(Y)\mathbb{1}_{A_k}]}{\mathbb{E}[\mathbb{1}_{A_k}]}$$

for all $\omega \in A_k$ such that $\mathbb{P}(A_k) > 0$. Therefore

$$Z(\omega) = \sum_{k: \mathbb{P}(A_k) > 0} \frac{\mathbb{E}[f(Y)\mathbb{1}_{A_k}]}{\mathbb{E}[\mathbb{1}_{A_k}]} \mathbb{1}_{A_k}(\omega), \qquad \text{for } \mathbb{P}\text{-almost all } \omega \in \Omega.$$

This shows that conditional expectation depends only on $\sigma(X)$ and not on the r.v. X (note that two random variables could generated the same σ-algebra). This observation then gives us the last motivation for the general definition of conditional expectation wrt. to a sub-σ-algebra of $\mathscr{F}$.

Definition 1. *Let $(\Omega, \mathscr{F}, \mathbb{P})$ a probability space and $\mathscr{G} \subseteq \mathscr{F}$ a sub-σ-algebra of $\mathscr{F}$. Let X a real integrable random variable (i.e. $\mathbb{E}[|X|] < \infty$). The conditional expectation of X given $\mathscr{G}$ is a $\mathscr{G}$-measurable random variable Z such that*

$$\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A Z] \qquad \forall A \in \mathscr{G} \tag{4}$$

The first properties of any conditional expectation are estabilished as follows.

Proposition 2. *If Z is a conditional expectation for X given $\mathscr{G}$, we have $\mathbb{E}|Z| \leqslant \mathbb{E}[|X|] < \infty$ and if Z, Z' are two conditional expectations for X given $\mathscr{G}$ then $Z = Z'$ almost surely.*

Proof. Let $H = \operatorname{sgn}(Z) = \mathbb{1}_{Z \geqslant 0} - \mathbb{1}_{Z < 0}$, then by (4)

$$0 \leqslant \mathbb{E}[|Z|] = \mathbb{E}[(\mathbb{1}_{Z \geqslant 0} - \mathbb{1}_{Z < 0}) Z] = \mathbb{E}[(\mathbb{1}_{Z \geqslant 0} - \mathbb{1}_{Z < 0}) X] = |\mathbb{E}[HX]| \leqslant \mathbb{E}[|X|] < \infty,$$

since $\{Z \geqslant 0\}, \{Z < 0\} \in \mathscr{G}$ and $|H(\omega)| \leqslant 1$. If Z, Z' are two conditional expectations, again by equation (4) we see that $Z - Z'$ is a conditional expectation for 0 given $\mathscr{G}$ and as a consequence $\mathbb{E}|Z - Z'| = 0$. Therefore $\mathbb{P}(Z = Z') = 1$, indeed $\mathbb{P}(|Z - Z'| \geqslant \varepsilon) \leqslant \varepsilon^{-1}\mathbb{E}[|Z - Z'|] = 0$ from which we deduce that $\mathbb{P}(Z \neq Z') = \mathbb{P}(\cup_n \{|Z - Z|' \geqslant 1/n\}) \leqslant \sum_n \mathbb{P}(|Z - Z|' \geqslant 1/n) = 0$. □

Remark 3. The condition (4) is indeed equivalent (via the monotone class theorem) to

$$\mathbb{E}[HX] = \mathbb{E}[HY] \qquad \forall H \hat{\in} \mathscr{B} \text{ bounded} \tag{5}$$

where we introduce the useful notation $H \hat{\in} \mathscr{B}$ to mean that H is a $\mathscr{B}$ measurable r.v.

We have still to show that such a conditional expectation Z always exists (see below).

By Prop. 2 we know that if the conditional expectation exists then is unique a.s.. We will denote some representative of the equivalence class by $Z = \mathbb{E}[X|\mathscr{G}]$, and also let $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$ when Y is another random variable. Moreover we will define the conditional probability given $\mathscr{G}$ by $\mathbb{P}(A|\mathscr{G}) = \mathbb{E}[\mathbb{1}_A|\mathscr{G}]$ for all $A \in \mathscr{F}$. Note that both conditional expectation and conditional probability are actually (equivalence classes of) random variables and not numerical quantities. Note also for the same reason that the map $A \mapsto P(A|\mathscr{G})$ is not a probability measure, so a conditional probability is not a probability... (more on this later).

Example 4. Let $X: \Omega \to \{0,1\}$, then

$$\sigma(X) = \{\emptyset, \Omega, X^{-1}(\{0\}), X^{-1}(\{1\})\}.$$

Sub σ-algebras of a probability space $(\Omega, \mathscr{F})$ model partial informations about the probabilistic situation. In this context $\sigma(X)$ is interpreted as the information gained by the observation of the random variable X. The trivial σ-algebra $\{\emptyset, \Omega\}$ then corresponds to absence of any information and $\mathscr{F}$ to a complete knowledge of the model.

Example 5. Let $\Omega = [0,1]$, et $\mathscr{F} = \mathscr{B}([0,1])$. let

$$\mathscr{F}_1 = \sigma([0,1/2],(1/2,1]) = \{[0,1/2],(1/2,1],[0,1],\emptyset\}.$$

Then $\mathscr{F}_1$ encodes the information whether ω is at the left or the right of $1/2$. In particular, if $X_1 = \mathbb{1}_{[0,1/2]}$, then $\mathscr{F}_1 = \sigma(X_1)$. Let now $X_2 = \mathbb{1}_{[0,1/4]} + \mathbb{1}_{(1/2,3/4]}$, and $\mathscr{F}_2 = \sigma(X_1, X_2)$. Then

$$\mathscr{F}_2 = \sigma([0,1/4],(1/4,1/2],(1/2,3/4],(3/4,1]),$$

but $\sigma(X_2) \neq \sigma(X_1, X_2)$. Knowledge of the value of $X_1(\omega)$ put ω at left or right of $1/2$. Knowledge of $X_2(\omega)$ put ω either in $[0,1/4] \cup (1/2,3/4]$ or in its complement. Knowledge of $X_1(\omega), X_2(\omega)$ allow to put ω in one of the sets $[0,1/4],(1/4,1/2],(1/2,3/4],(3/4,1]$. En passant we remark that if we consider the uniform probability $\mathbb{P}$ on [0, 1] then the random variables X_1 and X_2 are independent and Bernoulli with parameter $1/2$.

Example 6. For the trivial σ-algebra $\mathscr{G} = \{\emptyset, \Omega\}$ we have $\mathbb{E}[X|\mathscr{G}] = \mathbb{E}[X]$: is enough to verify that this guess satisfies the definition (4).

Theorem 7. *Let X a random variable with values in the measurable space $(\Theta, \mathscr{H})$ and Y a another r.v. with values in another measurable space $(\Upsilon, \mathscr{G})$, such that it is also $\sigma(X)$ measurable. Then there exists a mesurable function $h: (\Theta, \mathscr{H}) \to (\Upsilon, \mathscr{G})$ such that $Y = h(X)$.*

$$\begin{array}{ccc} (\Omega, \sigma(X)) & \overset{X}{\longrightarrow} & (\Theta, \mathscr{H}) \\ {}_{Y}\searrow & & \swarrow_{h\circ X} \\ & (\Upsilon, \mathscr{G}) & \end{array}$$

Thanks to Theorem 7, if we condition wrt. the σ-algebra generated by a random variable X we have some more information on the structure of the conditional expectation:

Proposition 8. *If $Z \in L^1$ and X is another real random variable, then there exists a measurable function $h: \mathbb{R} \to \mathbb{R}$ such that $\mathbb{E}[Z|X] = h(X)$ almost surely.*

2 Existence

Let $\mathscr{G}$ a σ-algebra contained in $\mathscr{F}$, $X \in L^2(\mathscr{F})$ and let $Y = \mathbb{E}[X|\mathscr{G}]$. Assume that $Y \in L^2$ (it is not difficult to prove it, we will do it later), then by an explicit computation it holds that

$$\mathbb{E}[|X-Z|^2] = \mathbb{E}[|X-Y|^2] + \mathbb{E}[|Y-Z|^2],$$

for any $Z \in L^2(\mathscr{G})$ and therefore

$$\mathbb{E}[|X-Y|^2] = \inf_{Z \in L^2(\mathscr{G})} \mathbb{E}[|X-Z|^2]. \tag{6}$$

This shows that a conditional expectation of an $L^2(\mathscr{F})$ random variable is the best $\mathscr{G}$-measurable estimator for X, according to a quadratic risk. Eq. (6) then gives us a strategy to prove the existence of the conditional expectation in the L^2 setting.

Recall that $L^2(\Omega, \mathscr{F}, \mathbb{P}) = L^2(\mathscr{F})$ is the completion of the family of simple functions by the norm $\|\cdot\|_2 = (\mathbb{E}[|\cdot|^2])^{1/2}$. Elements of $L^2(\mathscr{F})$ are equivalence classes of square-integrable measurable functions according to the equivalence relation $X \sim Y \Leftrightarrow \mathbb{P}(X \neq Y) = 0$.

Corollary 9. *If $\mathscr{B} \subseteq \mathscr{F}$ is a sub-σ-algebra of $\mathscr{F}$ then $L^2(\mathscr{B})$ is a closed vector subspace of $L^2(\mathscr{F})$ and for all $X \in L^2(\mathscr{F})$ there exists a unique $Y \in L^2(\mathscr{B})$ such that:*

a) $\mathbb{E}[|X-Y|^2] = \inf_{Z \in L^2(\mathscr{B})} \mathbb{E}[|X-Z|^2]$ *;*

b) $X - Y \perp L^2(\mathscr{B})$.

We call Y the orthogonal projection of X on $L^2(\mathscr{G})$.

Proof. The set $L^2(\mathscr{B})$ is complete with the L^2 norm, so it is also closed in $L^2(\mathscr{F})$. Let $\Delta = \inf_{Z \in L^2(\mathscr{B})} \mathbb{E}[|X-Z|^2]$ and $(Y_n)_n$ a minimizing sequence: $\mathbb{E}[|X-Y_n|^2] \to \Delta$ when $n \to \infty$. We have

$$\mathbb{E}[|X-Y_n|^2] + \mathbb{E}[|X-Y_m|^2] = 2\,\mathbb{E}[|X-(Y_n+Y_m)/2|^2] + \mathbb{E}[|Y_n-Y_m|^2]/2$$

(use $\mathbb{E}[|A+B|^2] + \mathbb{E}[|A-B|^2] = 2\mathbb{E}[A^2] + 2\mathbb{E}[B^2]$). But $(Y_n+Y_m)/2 \in L^2(\mathscr{B})$ which gives that

$$\mathbb{E}[|Y_n-Y_m|^2]/2 \leqslant \mathbb{E}[|X-Y_n|^2] + \mathbb{E}[|X-Y_m|^2] - 2\Delta \to 0,$$

for $n, m \to \infty$. Therefore the sequence $(Y_n)_n$ is Cauchy. Let $Y = L^2-\lim_n Y_n \in L^2(\mathscr{B})$. We have that $\|X-Y\|_2 \leqslant \|X-Y_n\|_2 + \|Y_n-Y\|_2$ and then that $\|X-Y\|_2 = \sqrt{\Delta}$ since $\|Y_n-Y\|_2 \to 0$.

For all $t \in \mathbb{R}$ and $Z \in L^2(\mathscr{B})$ consider $Y + tZ \in L^2(\mathscr{B})$ and observe that

$$0 \leqslant \mathbb{E}[|X - Y - tZ|^2] - \mathbb{E}[|X - Y|^2] = -2t\mathbb{E}[(X - Y)Z] + t^2\mathbb{E}[Z^2].$$

The polynomial $P(t) = at^2 + bt$ satisfy $P(t) \geqslant 0$ for all $t \geqslant 0$ which implies $b = 0$, and in particular $\mathbb{E}[(X - Y)Z] = 0$ for all $Z \in L^2(\mathscr{B})$. The converse implication is easy to show. To show uniqueness of the orthogonal projection assume that Y' is another projection. We have $\mathbb{E}[(Y - Y')Z] = 0$ for all $Z \in L^2(\mathscr{G})$ and therefore also for $Z = Y - Y'$, but then $\mathbb{E}[(Y - Y')^2] = 0 \Rightarrow Y - Y' = 0$ (a.s.). □

Theorem 10. *For all $X \in L^1(\mathscr{F})$ and σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ the conditional expectation $\mathbb{E}[X|\mathscr{G}]$ exists.*

Proof. The orthogonal projection Y of X on $L^2(\mathscr{G})$ satisfait $\mathbb{E}[XZ] = \mathbb{E}[YZ]$ for all $Z \in L^2(\mathscr{G})$ and in particular for all bounded $\mathscr{G}$-mesurable Z. Therefore $Y = \mathbb{E}[X|\mathscr{G}]$ a.s. which shows the existence of the conditional expectation when $X \in L^2(\mathscr{F})$.

To prove existence for all $X \in L^1(\mathscr{F})$ we proceed by approximation. Let $X \geqslant 0$ and in L^1. Let $X_n = \min(X, n)$ and Y_n the orthogonal projection of X_n onto $L^2(\mathscr{G})$. Then, for $n \geqslant m$ we have that $0 \leqslant \mathbb{E}[\mathbb{1}_A(X_n - X_m)] = \mathbb{E}[\mathbb{1}_A(Y_n - Y_m)]$ for all $A \in \mathscr{G}$ which implies that $Y_n \geqslant Y_m$ a.s. (check) and that it exists a null set $N \in \mathscr{G}$ off which the sequence $(Y_n(\omega))_n$ is increasing for all $\omega \in N^c$. Let $Y = \sup_n Y_n$. We have $\mathbb{E}[\mathbb{1}_A Y] = \sup_n \mathbb{E}[\mathbb{1}_A Y_n] = \sup_n \mathbb{E}[\mathbb{1}_A X_n] = \mathbb{E}[\mathbb{1}_A X]$ by monotone convergence and therefore, we have also $Y \in L^1(\mathscr{G})$ and $Y = \mathbb{E}[X|\mathscr{G}]$. For a generic $X \in L^1$ we decompose $X = X_+ - X_-$ with $X_+, X_- \geqslant 0$ and in L^1 and we let $Y_\pm = \mathbb{E}[X_\pm|\mathscr{B}]$ and $Y = Y_+ - Y_-$. We obtain $Y \in L^1(\mathscr{B})$ such that $\mathbb{E}[\mathbb{1}_A X] = \mathbb{E}[\mathbb{1}_A Y]$ for all $A \in \mathscr{B}$ as required. □

3 Properties

Proposition 11. *For all $X, Y \in L^1(\mathscr{F})$ and all sub-σ-algebras $\mathscr{G}, \mathscr{H} \subseteq \mathscr{F}$ we have the following properties of the conditional expectation:*

1. *Linearity: $\mathbb{E}[\lambda X + \mu Y|\mathscr{G}] = \lambda\mathbb{E}[X|\mathscr{G}] + \mu\mathbb{E}[Y|\mathscr{G}]$ for all $\lambda, \mu \in \mathbb{R}$;*
2. *Positivity: $X \geqslant 0 a.s. \Rightarrow \mathbb{E}[X|\mathscr{G}] \geqslant 0$ a.s. ;*
3. *Monotone convergence: $0 \leqslant X_n \nearrow X a.s. \Rightarrow \mathbb{E}[X_n|\mathscr{G}] \nearrow \mathbb{E}[X|\mathscr{G}]$ a.s. ;*
4. *Jensen's inequality: for all convex $\varphi: \mathbb{R} \to \mathbb{R}$: $\mathbb{E}[\varphi(X)|\mathscr{G}] \geqslant \varphi(\mathbb{E}[X|\mathscr{G}])$;*
5. *Contractivity in L^p: $\|\mathbb{E}[X|\mathscr{G}]\|_p \leqslant \|X\|_p$ for all $p \in [1, \infty]$,*
6. *Telescoping: If $\mathscr{H}$ is a sub-σ-algebra of $\mathscr{G}$ then*

$$\mathbb{E}[\mathbb{E}[X|\mathscr{G}]|\mathscr{H}] = \mathbb{E}[X|\mathscr{H}] = \mathbb{E}[\mathbb{E}[X|\mathscr{H}]|\mathscr{G}];$$

7. *If $Z \hat{\in} \mathscr{G}$, $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|XZ|] < +\infty$ then $\mathbb{E}[XZ|\mathscr{G}] = Z\,\mathbb{E}[X|\mathscr{G}]$.*

Proof.

1. Exercise.

2. We note that if $\mathbb{E}[X|\mathscr{G}] \leqslant \varepsilon < 0$ on $A \in \mathscr{G}$ such that $\mathbb{P}(A) > 0$ then $0 < \mathbb{E}[X 1_A] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}] 1_A] \leqslant \varepsilon \mathbb{P}(A) < 0$ which is impossible.

3. Let $Y_n = \mathbb{E}[X_n|\mathscr{G}]$. By positivity of conditional expectation we have that $(Y_n)_n$ is an increasing sequence. More precisely, there exist a probability 1 event A Let $Y = \limsup_n Y_n$, then $Y \hat{\in} \mathscr{G}$ and the monotone convergence theorem allows us to pass to the limit in $\mathbb{E}[X_n \mathbb{1}_A] = \mathbb{E}[Y_n 1_A]$ to obtain $\mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A]$ for all $A \in \mathscr{G}$. Therefore $Y = \mathbb{E}[X|\mathscr{G}]$ a.s.

4. Exercise.

5. Use property (4). Exercice.

6. Exercise.

7. Exercise. (Easy for simple functions and then use monotone limits for $X, Z \geqslant 0$). □

The following lemma will be useful later on in the study of martingales.

Lemma 12. *Let $X \in L^1$, then the family $(\mathbb{E}[X|\mathscr{G}]: \mathscr{G} \subseteq \mathscr{F}$ is a σ-algebra$)$ is uniformly integrable.*

Proof. Let $X_{\mathscr{G}} = \mathbb{E}[X|\mathscr{G}]$. We need to prove that for all $\varepsilon > 0$ there exists $K > 0$ such that

$$\mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| > K}] \leqslant \varepsilon,$$

for all $\mathscr{G}$. Observe that

$$\mathbb{E}[|X_{\mathscr{G}}| \mathbb{1}_{|X_{\mathscr{G}}| > K}] \leqslant \mathbb{E}[\mathbb{E}[|X||\mathscr{G}] \mathbb{1}_{|X_{\mathscr{G}}| > K}] = \mathbb{E}[\mathbb{E}[|X| \mathbb{1}_{|X_{\mathscr{G}}| > K}|\mathscr{G}]] = \mathbb{E}[|X| \mathbb{1}_{|X_{\mathscr{G}}| > K}]$$

and we can choose K large enough so that since $\mathbb{P}(|X_{\mathscr{G}}| > K) \leqslant \mathrm{KL}^{-1}\mathbb{E}[|X_{\mathscr{G}}|] \leqslant K^{-1}\mathbb{E}[|X|] \leqslant \delta$ uniformly in $\mathscr{G}$ for some prescribed $\delta > 0$. As a consequence, since X is uniformly integrable, we have $\mathbb{E}[|X| \mathbb{1}_{|X_{\mathscr{G}}| > K}] \leqslant \varepsilon$ for all $\mathscr{G}$ and we conclude. □

4 Relations with independence

Recall the basic notion of independence: two events $A, B \in \mathscr{F}$ are independent (wrt. the given probability measure $\mathbb{P}$) if

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

There are obvious generalisation to families of σ-algebras and to families of random variables.

Definition 13.

a) We say that a family $(\mathscr{A}_i)_{i \in I}$ of sub-σ-algebras of $\mathscr{F}$ is independent iff $\mathbb{P}(\cap_{i \in J} A_i) = \prod_{i \in J} \mathbb{P}(A_i)$ for all choices of $J \subseteq I$ and $A_i \in \mathscr{A}_i$, $i \in J$.

b) We say that X is a random variable independent of the σ-algebra $\mathscr{G}$ if $\sigma(X)$ is independent of $\mathscr{G}$, namely if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ for all $A \in \sigma(X)$ and $B \in \mathscr{G}$.

c) A family of random variables $(X_i)_{i \in I}$ is independent if is so the family of respective σ-algebras.

Proposition 14.

a) If X is independent of the σ-algebra $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}]$ is almost surely constant and

$$\mathbb{E}[X|\mathcal{G}] = \mathbb{E}(X), \qquad a.s.$$

b) If $\mathcal{H}$ and $\mathcal{G}$ are independent, X is $\mathcal{G}$-mesurable and $\mathcal{G}' \subseteq \mathcal{G}$, then

$$\mathbb{E}[X|\mathcal{H}, \mathcal{G}'] = \mathbb{E}[X|\mathcal{G}'].$$

c) If $X_1, \ldots, X_n$ is a family of independent r.v. and $f(X_1, \ldots, X_n) \in L^1$ then

$$\mathbb{E}[f(X_1, \ldots, X_n)|X_1] = \varphi(X_1)$$

where $\varphi(x) \coloneqq \mathbb{E}[f(x, X_2, \ldots, X_n)]$.

Proof. a) Easy.

b) Let us assume that $X \geqslant 0$ and is in L^1. Let $G \hat{\in} \mathcal{G}'$ and $H \hat{\in} \mathcal{H}$. By assumption $X \mathbb{1}_G \hat{\in} \mathcal{G}$ and $1_H \in \mathcal{H}$ are independent, therefore $\mathbb{E}[X \mathbb{1}_G \mathbb{1}_H] = \mathbb{E}[X \mathbb{1}_G]\, \mathbb{E}[\mathbb{1}_H]$ and if we denote $Y = \mathbb{E}[X|\mathcal{G}']$ we also have $\mathbb{E}[Y \mathbb{1}_G \mathbb{1}_H] = \mathbb{E}[Y \mathbb{1}_G]\mathbb{E}[\mathbb{1}_H]$ which tells us that $\mathbb{E}[X \mathbb{1}_G \mathbb{1}_H] = \mathbb{E}[Y \mathbb{1}_G \mathbb{1}_H]$. As a consequence, the measures

$$\mu_X(F) = \mathbb{E}[X \mathbb{1}_F] \text{ and } \mu_Y(F) = \mathbb{E}[X \mathbb{1}_F],$$

defined on $\sigma(\mathcal{G}', \mathcal{H})$ have the same mass and verify $\mu_X(G \cap H) = \mu_Y(G \cap H)$ for all $G \in \mathcal{G}'$ and $H \in \mathcal{H}$. But the family of events of the form $G \cap H$ is a π-system and therefore we can conclude that the measures coincide on $\sigma(\mathcal{G}', \mathcal{H})$.

c) Use Fubini's theorem on the joint law of $(X_1, X_2, \ldots, X_n)$. □

Remark 15. Note that in case c) above we do not have

$$\varphi(X_1) \neq \mathbb{E}[f(X_1, X_2, \ldots, X_n)].$$

There is no simple way to write down explicitly $\varphi(X_1)$, the correct expression

$$\varphi(X_1(\omega)) = \int_\Omega f(X_1(\omega), X_2(\omega'), \ldots, X_n(\omega'))\, \mathbb{P}(\mathrm{d}\omega')$$

make use of two different variables ω, ω' where the second is integrated over wrt. $\mathbb{P}$.

5 Regular conditional probabilities

Given a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ we can consider the random variables

$$A \in \mathcal{F} \mapsto \mathbb{P}(A|\mathcal{G}) = \mathbb{E}[\mathbb{1}_A|\mathcal{G}] \in L^1(\Omega, \mathcal{G}, \mathbb{P})$$

which are such that (check)

a) $\mathbb{P}(\emptyset|\mathcal{G}) = 0$, $\mathbb{P}(A^c|\mathcal{G}) = 1 - \mathbb{P}(A|\mathcal{G})$;

b) $\mathbb{P}(\cup_n A_n|\mathscr{G}) = \sum_n \mathbb{P}(A_n|\mathscr{G})$ for all countable families $(A_n)_n \subseteq \mathscr{F}$ of pairwise disjoint events.

However is in general not possible to guarantee that for fixed $\omega \in \Omega$ the set-function $A \in \mathscr{F} \mapsto \mathbb{P}(A|\mathscr{G})(\omega) \in [0,1]$ is a bona-fide probability measure on $(\Omega, \mathscr{F})$. The difficulty lies in the fact that property b) above holds $\mathbb{P}$-a.s., namely outside an exceptional set of probability 0. Given that there exists uncountably many families $(A_n)_n \subseteq \mathscr{F}$ in b), we cannot guarantee that the additivity holds with probability one, i.e. that we can find a *common* exceptional set for the property b).

This discussion motivates the introduction of the concept of *regular conditional probability*. A regular conditional probability given the σ-algebra $\mathscr{G} \subseteq \mathscr{F}$ is the assignment $\mathbb{P}_{\mathscr{G}}: \Omega \to \mathscr{P}(\Omega, \mathscr{F})$ where $\mathscr{P}(\Omega, \mathscr{F})$ is the space of probability measures on the measure space $(\Omega, \mathscr{F})$, such that for all $A \in \mathscr{F}$ the map $\omega \mapsto \mathbb{P}_{\mathscr{G}}(\omega, A)$ is $\mathscr{G}$-measurable and is a version of the conditional probability $\mathbb{P}(A|\mathscr{G})$.

Existence of a regular conditional probability $\mathbb{P}_{\mathscr{G}}$ for $\mathscr{G}$ ensures that the conditional expectation is indeed the expectation wrt. a standard probability measure: a monotone class argument allows to prove that for all $X \in L^1(\Omega, \mathscr{F}, \mathbb{P})$ we have

$$\mathbb{E}[X|\mathscr{G}](\omega) = \int_\Omega X(\omega')\mathbb{P}_{\mathscr{G}}(\omega, \mathrm{d}\omega'), \quad \text{for } \mathbb{P}\text{-almost all } \omega \in \Omega.$$

Remark 16. Existence of a regular conditional probability can be guaranteed when $(\Omega, \mathscr{F})$ is a Polish space (complete, metrisable space) endowed with the Borel σ-algebra. This is another pleasant feature of the Polish setting.

6 Some examples

Example 17. Let $(X_i)_{1 \leqslant i \leqslant n}$ a vector of i.i.d. random variables and let $X = \sum_i X_i$, then

$$\mathbb{E}[X_1|X] = \frac{X}{n}, \quad \mathbb{E}[X|X_1] = (n-1)\mathbb{E}[X_1] + X_1.$$

In order to prove these equalities one has to remark first that we have

$$\mathbb{E}[X_1|X] = \mathbb{E}[X_2|X] = \cdots = \mathbb{E}[X_n|X] = h(X)$$

for some measurable function $h: \mathbb{R} \to \mathbb{R}$. Indeed since the variables are i.i.d. we have

$$\mathbb{E}[f(X_1, \ldots, X_n)] = \mathbb{E}[f(X_{\sigma(1)}, \ldots, X_{\sigma(n)})]$$

for any bounded measurable function $f: \mathbb{R}^n \to \mathbb{R}$, which implies that for any bounded measurable function X we have (think why)

$$\mathbb{E}[X_1 h(X)] = \mathbb{E}[X_2 h(X)] = \cdots = \mathbb{E}[X_n h(X)]$$

and therefore one can conclude using the definition of conditional expectation. By linearity we have

$$X = \mathbb{E}[X|X] = \mathbb{E}[X_1 + \cdots + X_n|X] = [X_1|X] + \mathbb{E}[X_2|X] + \cdots + \mathbb{E}[X_n|X] = nh(X)$$

and we conclude

$$\mathbb{E}[X_1|X] = \frac{X}{n}.$$

The other equation follows more easily.

Example 18. Let $(X_{n,m})_{n,m\geqslant 0}$ a double suite of i.i.d. r.v. with values in $\mathbb{R}_+$. Let $Z_0 = 1$ and $Z_n = X_{n,1} + \cdots + X_{n,Z_{n-1}}$ for $n \geqslant 1$. Using the conditional expectation one can show that the generating function $f_n(\theta) = \mathbb{E}[\theta^{Z_n}]$ satisfies

$$f_0(\theta) = 1 \qquad f_n = f_{n-1}(f(\theta)) \quad \text{for } n \geqslant 1.$$

For any $\theta \in [0,1]$ let $f(\theta) = \mathbb{E}[\theta^{X_{1,1}}] = \mathbb{E}[\theta^{Z_1}] = f_1(\theta)$. Then observe that Z_{n-1} is independent of $(X_{n,k})_k$ since it is measurable wrt the σ-algebra generated by $(X_{\ell,k})_{\ell\leqslant n-1,k}$ which is independent of that generated by $(X_{n,k})_k$. Therefore

$$f_n(\theta) = \mathbb{E}[\theta^{Z_n}] = \mathbb{E}[\theta^{X_{n,1}+\cdots+X_{n,Z_{n-1}}}] = \mathbb{E}[\mathbb{E}[\theta^{X_{n,1}+\cdots+X_{n,Z_{n-1}}}|Z_{n-1}]] = \mathbb{E}[\varphi(Z_{n-1})]$$

where

$$\varphi(z) = \mathbb{E}[\theta^{X_{n,1}+\cdots+X_{n,z}}]$$

by independence. We can compute $\varphi(z)$ again using independece of the vector $(X_{n,k})_k$ as

$$\varphi(z) = \mathbb{E}[\theta^{X_{n,1}}\cdots\theta^{X_{n,z}}] = \mathbb{E}[\theta^{X_{n,1}}]\cdots\mathbb{E}[\theta^{X_{n,z}}] = \underbrace{f(\theta)\cdots f(\theta)}_{z \text{ times}} = f(\theta)^z$$

using also the fact that $(X_{n,k})_k$ have the same law, namely that of $X_{1,1}$. We conclude therefore that

$$f_n(\theta) = \mathbb{E}[\varphi(Z_{n-1})] = \mathbb{E}[f(\theta)^{Z_{n-1}}] = f_{n-1}(f(\theta)).$$