A.6 Conditional expectations

Let a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ be given. Consider two random variables or vectors X and Y that both assume finitely many values in sets $\mathcal{X}$ and $\mathcal{Y}$ respectively. Assume that $\mathbb{P}(Y = y) > 0$ for all $y \in \mathcal{Y}$. Then the conditional probabilities $\mathbb{P}(X = x|Y = y)$ are all well defined as well as for any function f on $\mathcal{X}$ the *conditional expectation* $\mathbb{E}[f(X)|Y = y] := \sum_{x\in\mathcal{X}} f(x)\mathbb{P}(X = x|Y = y)$.

Consider the function $\hat{f}$ defined by $\hat{f}(y) = \mathbb{E}[f(X)|Y = y]$. With the aid of $\hat{f}$ we define the *conditional expectation of* $f(X)$ *given* Y, denoted by $\mathbb{E}[f(X)|Y]$, as $\hat{f}(Y)$. A simple calculation suffices to check the relation

$$\mathbb{E}[f(X)|Y] = \sum_y \hat{f}(y)\mathbf{1}_{\{Y=y\}} = \sum_y \frac{\mathbb{E}(f(X)\mathbf{1}_{\{Y=y\}})}{\mathbb{P}(Y = y)}\mathbf{1}_{\{Y=y\}}. \tag{A.5}$$

Elementary properties of conditional expectation like linearity are in this case easy to prove. Other properties are equally fundamental and easy to prove. They are listed in Proposition A.10. Replacing X above by (X, Y) we can also define conditional expectations like $\mathbb{E}[f(X, Y)|Y]$.

Proposition A.10 *The following properties hold for conditional expectations.*

(i) *If X and Y are independent, then $\mathbb{E}[f(X)|Y] = \mathbb{E}[f(X)]$, the unconditional expectation.*

(ii) *Let $f(X, Y)$ be the product $f_1(X)f_2(Y)$. Then*

$$\mathbb{E}[f_1(X)f_2(Y)|Y] = f_2(Y)\mathbb{E}[f_1(X)|Y].$$

(iii) *If X and Y are independent, then*

$$\mathbb{E}[f(X, Y)|Y] = \sum_{x\in\mathcal{X}} f(x, Y)\mathbb{P}(X = x).$$

Proof Exercise A.13. □

The random variable Y induces a sub-σ-algebra of $\mathcal{F}$ on Ω, call it $\mathcal{G}$, which is generated by the sets $\{Y = y\}$ and that $\mathcal{G} = \sigma(Y)$. Note that the sets $\{Y = y\}$ constitute a partition of Ω and hence every set in $\mathcal{G}$ is a finite union of some $\{Y = y\}$.

Consider again the function $\hat{f}$ above. It defines another function on Ω, F say, according to $F(\omega) = \hat{f}(y)$ on the set $\{\omega : Y(\omega) = y\}$, hence $F(\omega) = \hat{f}(Y(\omega))$, in short $F = \hat{f}(Y)$. In this way we can identify the conditional expectation $\hat{f}(Y)$ with the random variable F. Note that F is $\mathcal{G}$-measurable, it is constant on the sets $\{Y = y\}$. Moreover, one easily verifies (Exercise A.14) that

$$\mathbb{E}(F\mathbf{1}_{\{Y=y\}}) = \mathbb{E}(f(X)\mathbf{1}_{\{Y=y\}}). \tag{A.6}$$

In words, the expectation of $f(X)$ and its conditional expectation over the sets $\{Y = y\}$ are the same, and therefore we will also have

$$\mathbb{E}(F\mathbf{1}_G) = \mathbb{E}(f(X)\mathbf{1}_G) \text{ for every set } G \in \mathcal{G}. \tag{A.7}$$

Conditional expectation can also be defined for random variables that are not discrete. To that end we proceed directly to the general definition as it is used in modern probability theory. That it is possible to define conditional expectation as we will do below is a consequence of the Radon-Nikodym theorem in measure theory. The definition is motivated by (A.7).

Definition A.11 If X is a random variable with $\mathbb{E}|X| < \infty$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$, then the conditional expectation of X given $\mathcal{G}$ is any $\mathcal{G}$-measurable random variable $\hat{X}$ with the property that $\mathbb{E}X\mathbf{1}_G = \mathbb{E}\hat{X}\mathbf{1}_G$ for all $G \in \mathcal{G}$. We will use the notation $\mathbb{E}[X|\mathcal{G}]$ for any of the $\hat{X}$ above.

Note that the conditional expectation is not uniquely defined, but only up to almost sure equivalence, if $\hat{X}$ and $\hat{X}'$ both satisfy the requirements of Definition A.11 then $\mathbb{P}(\hat{X} = \hat{X}') = 1$, see Exercise A.23. Different $\hat{X}$ are called versions of the conditional expectation. Conditioning with respect to a random variable (vector) Y is obtained by taking $\mathcal{G}$ equal to the σ-algebra that Y induces on Ω. It is a theorem in probability theory that any version of the conditional expectation given Y can be represented by a (measurable) function of Y. We have seen this to be true at the beginning of this section. The most important properties of conditional expectation that we use in this course are collected in the following proposition.

Proposition A.12 *Let X be a random variable with $\mathbb{E}|X| < \infty$.*

(i) $\mathbb{E}(\mathbb{E}[X|\mathcal{G}]) = \mathbb{E}X$.

(ii) *If X is independent of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$.*

(iii) *If $\mathcal{H}$ is a sub-σ-algebra of $\mathcal{G}$, then $\mathbb{E}[X|\mathcal{H}] = \mathbb{E}[\mathbb{E}[X|\mathcal{G}]|\mathcal{H}]$ (iterated conditioning).*

(iv) *If Y is $\mathcal{G}$-measurable and $\mathbb{E}|XY| < \infty$, then $\mathbb{E}[XY|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}]Y$.*

(v) *If X is independent of $\mathcal{G}$, Y is $\mathcal{G}$-measurable and f is a measurable function on $\mathbb{R}^2$ with $\mathbb{E}|f(X,Y)| < \infty$, then $\mathbb{E}[f(X,Y)|\mathcal{G}] = \int f(x,Y)\,\mathbb{P}^X(dx)$, with $\mathbb{P}^X$ the distribution of X. Alternatively written, $\mathbb{E}[f(X,Y)|\mathcal{G}] = \hat{f}(Y)$, where $\hat{f}(y) = \mathbb{E}f(X,y)$.*

(vi) *The conditional expectation is a linear operator on the space of random variables with finite expectation.*

A special occurs of conditional expectation occurs if $\mathcal{G}$ is generated by a partition $\{G_1, \ldots, G_m\}$ of Ω with all $\mathbb{P}(G_i) > 0$. Then, completely analogous to (A.5),

$$\mathbb{E}[X|\mathcal{G}] = \sum_i \frac{\mathbb{E}(X\mathbf{1}_{G_i})}{\mathbb{P}(G_i)}\mathbf{1}_{G_i}. \tag{A.8}$$

Indeed, (A.5) is a special case of (A.8). Take $\mathcal{G} = \sigma(Y)$ and $G_i = \{Y = y_i\}$, where the y_i are the different values that Y assumes.

If the random variable X in Definition A.11 has the stronger integrability property $\mathbb{E}X^2 < \infty$, then one has (this is Exercise A.25)

$$\mathbb{E}(X - \hat{X})^2 \leq \mathbb{E}(X - Y)^2, \text{ for all } \mathcal{G}\text{-measurable } Y \text{ with } \mathbb{E}Y^2 < \infty. \tag{A.9}$$

Equation (A.9) offers a nice interpretation of conditional expectation as a projection, think of this! And with this interpretation in mind, property (iii) of Proposition A.12 should look familiar: it is analogous to repeated, iterated projection, first on a subspace, than on a smaller subspace, being equivalent to immediate projection on the smaller subspace.