A Bit About Conditional Expectation

Kexing Ying

July 24, 2021

Measure-theoretically, conditional expectation is defined in terms of conditioning on a sub-σ algebra.

Definition 1 (Conditional Expectation)**.** Let $f : \Omega \to \mathbb{R}^+$ be a random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\int f \mathrm{d}\mathbb{P} < \infty$. Then, given a sub-$\sigma$-algebra $\mathcal{G}$ of $\mathcal{F}$, the conditional expectation of f with respect to $\mathcal{G}$ is the essentially unique $\mathcal{G}$-random variable f', such that

$$\int_A f \mathrm{d}\mathbb{P} = \int_A f' \mathrm{d}\mathbb{P}, \tag{1}$$

for all $A \in \mathcal{G}$. We denote f' by $\mathbb{E}(f \mid \mathcal{G})$.

We note that the integrals within this definition are computed with respect to the larger $\mathcal{F}$ rather than the smaller $\mathcal{G}$. Indeed, while f' is measurable with respect to $\mathcal{F}$, f is not measurable with respect to $\mathcal{G}$.

We also note that, if f is already measurable with respect to $\mathcal{G}$, then $f' = f$, $\mathbb{P}$-a.e. This follows since the Radon-Nikodym derivative is essentially unique and clearly, f satisfy the integral equation 1. Written, using the Radon-Nikodym derivative notation, we have

$$f' = \frac{\mathrm{d}(f\mathbb{P})|_{\mathcal{G}}}{\mathrm{d}\mathbb{P}|_{\mathcal{G}}}.$$

Thus, if f is $\mathcal{G}$-measurable, we simply exploiting the fact that,

$$(f\mathbb{P})|_{\mathcal{G}} = f(\mathbb{P}|_{\mathcal{G}}).$$

Geometrically, we may think about conditioning on a sub-σ-algebra as projection onto the sub-σ-algebra. This notion is made rigorous in the case of L^2-random variables.

Definition 2 (Conditional Expectation in L^2)**.** Let $f : \Omega \to \mathbb{R}^+$ be a L^2-random variable on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, given a sub-$\sigma$-algebra $\mathcal{G}$ of $\mathcal{F}$, the conditional expectation of f with respect to $\mathcal{G}$ is the orthogonal projection of f into $L^2(\Omega, \mathcal{G}, \mathbb{P}|_{\mathcal{G}})$.

In this definition, the conditional expectation f' is the $\mathcal{G}$-random variable which minimizes

$$\mathbb{E}[(f - f')^2] = \int (f - f')^2 \mathrm{d}\mathbb{P}.$$

By now, one might wonder how does this definition relate to the classical definition of conditional expectation where we condition on an event. In particular, how does this notion relate to $\mathbb{E}(X \mid B) = \mathbb{E}(X1|_B) = \frac{1}{\mathbb{P}(B)} \int_B X \mathrm{d}\mathbb{P}$ where $B \in \mathcal{F}, \mathbb{P}(B) > 0$.

Proposition 1. Given $B \in \mathcal{F}$,

$$\mathbb{E}(X \mid \sigma(\{B\})) \overset{\text{a.e.}}{=} \omega \mapsto \begin{cases} \mathbb{E}(X \mid B) & \text{if } \omega \in B \\ \mathbb{E}(X \mid B^c) & \text{if } \omega \notin B. \end{cases}$$

Proof. Denote Y as the random variable defined by the right hand side.

Since $\sigma(\{B\}) = \{\varnothing, B, B^c, \Omega\}$, it is clear that the Y is $\sigma(\{B\})$-measurable. Furthermore, it is also clear that, for all $A \in \sigma(\{B\})$, $\int_A Y \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P}$. Thus, the two random variables are equal almost everywhere. $\square$

To obtain the other elementary definitions learnt during an introduction to probability course regarding conditional probabilities, we may simply define,

Definition 3. Given $A \in \mathcal{F}$, we write $\mathbb{P}(A \mid \mathcal{G}) := \mathbb{E}(1|_A \mid \mathcal{G})$.

Definition 4. Given f, g $\mathcal{F}$-random variables, we write $\mathbb{E}(f \mid g) := \mathbb{E}(f \mid \sigma(g))$.

Using this measure-theoretic definition of conditional expectation, some standard results are extremely simple to derive.

Proposition 2. If $\{A_i \mid i \in \mathbb{N}\}$ form a partition of $\mathcal{F}$, then

$$\mathbb{E}(X \mid \mathcal{G}) = \sum_{i \in \mathbb{N}} \mathbb{E}(X \mid A_i) 1|_{A_i},$$

where $\mathcal{G} = \sigma(\{A_i \mid i \in \mathbb{N}\})$.

Proof. Clear by considering for all $G \in \mathcal{G}$, there exists some $I \subseteq \mathbb{N}$ such that $G = \bigcup_{i \in I} A_i$. $\square$

Proposition 3 (Law of Total Expectation)**.** Let $\mathcal{H} \leq \mathcal{G} \leq \mathcal{F}$, and X a $\mathcal{F}$-random variable, then

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \mathbb{E}(X)$$

and

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}), \mathbb{P}\text{-a.e.}$$

Proof. The first result is immediate as,

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G})) = \int_\Omega \mathbb{E}(X \mid \mathcal{G}) \mathrm{d}\mathbb{P} = \int_\Omega X \mathrm{d}\mathbb{P} = \mathbb{E}(X).$$

For the second result, we see, for all $A \in \mathcal{H}$,

$$\int_A \mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) \mathrm{d}\mathbb{P} = \int_A \mathbb{E}(X \mid \mathcal{G}) \mathrm{d}\mathbb{P} = \int_A X \mathrm{d}\mathbb{P}.$$

Thus, as the conditional expectation is essentially unique, we have

$$\mathbb{E}(\mathbb{E}(X \mid \mathcal{G}) \mid \mathcal{H}) = \mathbb{E}(X \mid \mathcal{H}), \mathbb{P}\text{-a.e.}$$

$\square$