Lecture-03: Conditional Expectation

1 Conditional expectation

Consider a probability space $(\Omega, \mathcal{F}, P)$, a random variable $X : \Omega \to \mathbb{R}$ defined on this probability space, and a nontrivial event $E \in \mathcal{F}$ such that $P(E) > 0$.

Definition 1.1. The conditional distribution of X conditioned on a nontrivial event E is defined for each $x \in \mathbb{R}$ as

$$F_{X|E}(x) \triangleq \frac{P(\{X \leqslant x\} \cap E)}{P(E)}.$$

Remark 1. We can verify that $F_{X|E} : \mathbb{R} \to [0,1]$ is a distribution function for any nontrivial $E \in \mathcal{F}$.

Definition 1.2. Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function. The conditional expectation of random variable $g(X)$ given a nontrivial event E is defined as

$$\mathbb{E}[g(X) \mid E] \triangleq \int_{x \in \mathbb{R}} g(x) dF_{X|E}(x).$$

Example 1.3. Consider two random variables X, Y defined on the same probability space $(\Omega, \mathcal{F}, P)$ with the joint distribution $F_{X,Y}(x,y) = P(\{X \leqslant x, Y \leqslant y\})$. For each $y \in \mathbb{R}$, we define event $A_Y(y) \triangleq Y^{-1}(-\infty, y] \in \mathcal{F}$ such that $F_Y(y) = P(A_Y(y))$. Then, for each $y \in \mathbb{R}$ such that $P(A_Y(y)) > 0$, we can write the conditional distribution of X given the event $A_Y(y)$ as

$$F_{X|A_Y(y)}(x) = \frac{F_{X,Y}(x,y)}{F_Y(y)}.$$

The conditional expectation of X given the event $A_Y(y)$ is defined as

$$\mathbb{E}[X|A_Y(y)] = \int_{x \in \mathbb{R}} x dF_{X|A_Y(y)}(x) = \frac{1}{F_Y(y)} \int_{x \in \mathbb{R}} x \frac{\partial}{\partial x} F_{X,Y}(x,y).$$

Definition 1.4. The *conditional expectation* of X given an event subspace $\mathcal{E} \subseteq \mathcal{F}$ is denoted $\mathbb{E}[X \mid \mathcal{E}]$ and is a random variable $Z \triangleq \mathbb{E}[X \mid \mathcal{E}] : \Omega \to \mathbb{R}$ that satisfies

(i) *measurability:* for each $B \in \mathcal{B}(\mathbb{R})$ we have $Z^{-1}(B) \in \mathcal{E}$,
(ii) *integrability:* $\mathbb{E}|Z| < \infty$, and
(iii) *orthogonality:* for each event $E \in \mathcal{E}$, we have $\mathbb{E}[X \mathbb{1}_E] = \mathbb{E}[Z \mathbb{1}_E]$.

Proposition 1.5. *Conditional expectation is unique almost surely.*

Proof. Consider a random variable $X : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ and a sub event space $\mathcal{E} \subseteq \mathcal{F}$. Let Z_1 and Z_2 be conditional expectations of X given $\mathcal{E}$. It suffices to show that $A_\epsilon \triangleq \{\omega \in \Omega : Z_1 - Z_2 > \epsilon\} \in \mathcal{E}$ and $B_\epsilon \triangleq \{\omega \in \Omega : Z_2 - Z_1 > \epsilon\} \in \mathcal{E}$ defined for each $\epsilon > 0$ has measure $P(A_\epsilon) = P(B_\epsilon) = 0$. From the definition of conditional expectation and linearity of expectation, we can write

$$0 \leqslant \epsilon P(A_\epsilon) < \mathbb{E}[(Z_1 - Z_2)\mathbb{1}_{A_\epsilon}] = \mathbb{E}[X \mathbb{1}_{A_\epsilon}] - \mathbb{E}[X \mathbb{1}_{A_\epsilon}] = 0.$$

Similarly, we can show that $P(B_\epsilon) = 0$, and the result follows. $\square$

Remark 2. Any random variable $Z : \Omega \to \mathbb{R}$ that satisfies the measurability, orthogonality, and integrability, is the conditional expectation of X given the sub-event space $\mathcal{E}$ from the a.s. uniqueness of conditional expectations.

Remark 3. Intuitively, we think of the event subspace $\mathcal{E}$ as describing the information we have. For each $A \in \mathcal{E}$, we know whether or not A has occurred. The conditional expectation $\mathbb{E}[X|\mathcal{E}]$ is the "best guess" of the value of X given the information $\mathcal{E}$.

Definition 1.6. Consider a random variable $X : \Omega \to \mathbb{R}$ and a random vector $Y : \Omega \to \mathbb{R}^n$ defined on the same probability space $(\Omega, \mathcal{F}, P)$. The conditional expectation of X given Y is defined as

$$\mathbb{E}[X \mid Y] \triangleq \mathbb{E}[X \mid \sigma(Y)].$$

Proposition 1.7. *For two random variables $X, Y : \Omega \to \mathbb{R}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, the conditional expectation $\mathbb{E}[X \mid Y]$ is a function of Y.*

Proof. We denote the conditional expectation $\mathbb{E}[X \mid Y]$ by a $\sigma(Y)$-measurable random variable $Z : \Omega \to \mathbb{R}$. It suffices to show that for any $y \in \mathbb{R}$, the conditional expectation $Z(\omega)$ remains constant on the set of outcomes $\omega \in Y^{-1}\{y\}$. First, we show that for any event $A \in \sigma(Y)$, either $Y^{-1}\{y\} \subseteq A$ or $A \cap Y^{-1}\{y\} = \emptyset$. This follows from the fact that either $y \in A$ or $y \notin A$. Next, we suppose that there exists a $y \in \mathbb{R}$ and $\omega_1, \omega_2 \in Y^{-1}\{y\}$ such that $Z(\omega_1) \neq Z(\omega_2)$. It follows that there exists an event $B \triangleq Z^{-1}\{Z(\omega_1)\} \in \sigma(Z)$ such that $\omega_1 \in B$ and $\omega_2 \notin B$. Since Z is $\sigma(Y)$-measurable, it follows that $B \in \sigma(Z) \subseteq \sigma(Y)$. This leads to a contradiction. □

Proposition 1.8. *Let X, Y be random variables on the probability space $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Let $\mathcal{G}$ and $\mathcal{H}$ be sub-event spaces of $\mathcal{F}$. Then, conditional expectation has the following properties.*

1. Linearity: $\mathbb{E}[\alpha X + \beta Y \mid \mathcal{G}] = \alpha \mathbb{E}[X \mid \mathcal{G}] + \beta E[Y \mid \mathcal{G}]$, *a.s.*
2. Monotonicity: *If $X \leqslant Y$ a.s., then $\mathbb{E}[X \mid \mathcal{G}] \leqslant E[Y \mid \mathcal{G}]$, a.s.*
3. Identity: *If X is $\mathcal{G}$-measurable and $\mathbb{E}|X| < \infty$, then $X = \mathbb{E}[X \mid \mathcal{G}]$ a.s. In particular, $c = \mathbb{E}[c \mid \mathcal{G}]$, for any constant $c \in \mathbb{R}$.*
4. Conditional Jensen's inequality: *If $\psi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|\psi(X)| < \infty$, then $\mathbb{E}[\psi(X) \mid \mathcal{G}] \geqslant \psi(\mathbb{E}[X \mid \mathcal{G}])$, a.s.*
5. Pulling out what's known: *If Y is $\mathcal{G}$-measurable and $\mathbb{E}|XY| < \infty$, then $\mathbb{E}[XY \mid \mathcal{G}] = Y\mathbb{E}[X \mid \mathcal{G}]$, a.s.*
6. Tower property: *If $\mathcal{H} \subseteq \mathcal{G}$, then $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}] = \mathbb{E}[X \mid \mathcal{H}]$, a.s..*
7. L^2-projection: *If $\mathbb{E}|X|^2 < \infty$, then $\zeta^* = \mathbb{E}[X \mid \mathcal{G}]$ minimizes $\mathbb{E}[(X - \zeta)^2]$ over all $\mathcal{G}$-measurable random variables ζ such that $\mathbb{E}|\zeta|^2 < \infty$.*
8. Irrelevance of independent information: *If $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then $\mathbb{E}[X|\sigma(\mathcal{G}, \mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}]$, a.s. In particular, if X is independent of $\mathcal{H}$, then $\mathbb{E}[X \mid \mathcal{H}] = \mathbb{E}[X]$ a.s.*

Proof. Let X, Y be random variables on the probability space $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}|X|, \mathbb{E}|Y| < \infty$. Let $\mathcal{G}$ and $\mathcal{H}$ be event spaces such that $\mathcal{G}, \mathcal{H} \subseteq \mathcal{F}$.

1. *Linearity:* Let $Z \triangleq \alpha \mathbb{E}[X \mid \mathcal{G}] + \beta \mathbb{E}[Y \mid \mathcal{G}]$, then since $\mathbb{E}[X \mid \mathcal{G}], \mathbb{E}[Y \in \mathcal{G}]$ are $\mathcal{G}$-measurable, it follows that their linear combination Z is also $\mathcal{G}$-measurable. The integrability follows from the following triangle inequality and the monotonicity of expectation

   $$|Z| \leqslant |\alpha| |\mathbb{E}[X \mid \mathcal{G}]| + |\beta| |\mathbb{E}[Y \mid \mathcal{G}]|.$$

   Further, for any event $F \in \mathcal{G}$, from the linearity of expectation and definition of conditional expectation, we have

   $$\mathbb{E}[Z \mathbb{1}_G] = \alpha \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mathbb{1}_G] + \beta \mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}] \mathbb{1}_G] = \mathbb{E}[(\alpha X + \beta Y) \mathbb{1}_G].$$

2. *Monotonicity:* Let $\epsilon > 0$ and define $A_\epsilon \triangleq \{\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[Y \mid \mathcal{G}] > \epsilon\} \in \mathcal{G}$. Then from the definition of conditional expectation, we have

   $$0 \leqslant \mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[Y \mid \mathcal{G}]) \mathbb{1}_{A_\epsilon}] = \mathbb{E}[(X - Y) \mathbb{1}_{A_\epsilon}] \leqslant 0.$$

   Thus, we obtain that $P(A_\epsilon) = 0$ for all $\epsilon > 0$. Taking limit $\epsilon \downarrow 0$, we get $0 = \lim_{\epsilon \downarrow 0} P(A_\epsilon) = P(\lim_\epsilon A_\epsilon) = P(A_0)$.
3. *Identity:* It follows from the definition that X satisfies all three conditions for conditional expectation. The event space generated by any constant function is the trivial event space $\{\emptyset, \Omega\} \subseteq \mathcal{G}$ for any event space. Hence, $\mathbb{E}[c \mid \mathcal{G}] = c$.
4. *Conditional Jensen's inequality:* We will use the fact that a convex function can always be represented by the supremum of a family of affine functions. Accordingly, we will assume for a convex function $\psi : \mathbb{R} \to \mathbb{R}$, we have linear functions $\phi_i : \mathbb{R} \to \mathbb{R}$ and constants $c_i \in \mathbb{R}$ for all $i \in I$ such that $\psi = \sup_{i \in I}(\phi_i + c_i)$.

For each $i \in I$, we have $\phi_i(\mathbb{E}[X \mid \mathcal{G}]) + c_i = \mathbb{E}[\phi_i(X) \mid \mathcal{G}] + c_i \leqslant \mathbb{E}[\psi(X) \mid \mathcal{G}]$ from the linearity and monotonicity of conditional expectation. It follows that

$$\psi(\mathbb{E}[X \mid \mathcal{G}]) = \sup_{i \in I}(\phi_i(\mathbb{E}[X \mid \mathcal{G}]) + c_i) \leqslant \mathbb{E}[\psi(X) \mid \mathcal{G}].$$

5. *Pulling out what's known:* Let Y be $\mathcal{G}$-measurable and $\mathbb{E}|XY| < \infty$. Since Y is given to be $\mathcal{G}$-measurable, conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable by definition, and product function is Borel measurable, it follows that $Y\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable.
It suffices to show that $\mathbb{E}[XY\mathbb{1}_G] = \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G]$ for all events $G \in \mathcal{G}$ and $\mathbb{E}|Y\mathbb{E}[X \mid \mathcal{G}]| < \infty$, when Y is a simple $\mathcal{G}$-measurable random variable such that $\mathbb{E}|XY| < \infty$. It follows that, we can write $Y = \sum_{y \in \mathcal{Y}} y\mathbb{1}_{E_y}$ for finite $\mathcal{Y}$ and $E_y \triangleq Y^{-1}\{y\} \in \mathcal{G}$ for all $y \in \mathcal{Y}$. From the definition of conditional expectation and linearity, we obtain for any $G \in \mathcal{G}$

$$\mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G] = \sum_{y \in \mathcal{Y}} y\mathbb{E}[\mathbb{1}_{G \cap E_y}\mathbb{E}[X \mid \mathcal{G}]] = \sum_{y \in \mathcal{Y}} y\mathbb{E}[X\mathbb{1}_{G \cap E_y}] = \mathbb{E}[X\sum_{y \in \mathcal{Y}} y\mathbb{1}_{G \cap E_y}] = \mathbb{E}[XY\mathbb{1}_G].$$

Conditional Jensen's inequality applied to convex function $|| : \mathbb{R} \to \mathbb{R}_+$, we get $|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}[|X| \mid \mathcal{G}]$. Therefore,

$$\mathbb{E}[|Y|\,|\mathbb{E}[X \mid \mathcal{G}]|] = \sum_{y \in \mathcal{Y}} |y|\,\mathbb{E}[|\mathbb{E}[X \mid \mathcal{G}]|\,\mathbb{1}_{E_y}] \leqslant \sum_{y \in \mathcal{Y}} |y|\,\mathbb{E}[|X|\,\mathbb{1}_{E_y}] = \mathbb{E}\,|XY|.$$

6. *Tower property:* Measurability follows from the definition of conditional expectation, since $\mathbb{E}[X \mid \mathcal{H}]$ is $\mathcal{H}$ measurable. Integrability follows from the application of conditional Jensen's inequality to convex function $|| : \mathbb{R} \to \mathbb{R}_+$ to get $|\mathbb{E}[X \mid \mathcal{H}]| \leqslant \mathbb{E}[|X| \mid \mathcal{H}]$, which implies $\mathbb{E}\,|\mathbb{E}[X \mid \mathcal{H}]| \leqslant \mathbb{E}\,|X| < \infty$. Orthogonality follows from the definition of conditional expectation, since for any $H \in \mathcal{H} \subseteq \mathcal{G}$, we have
$$\mathbb{E}[\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mid \mathcal{H}]\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_H] = \mathbb{E}[X\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{H}]\mathbb{1}_H].$$

7. *L^2-projection:* We define $L^2(\mathcal{G}) \triangleq \left\{\zeta \text{ a } \mathcal{G} \text{ measurable random variable} : \mathbb{E}\zeta^2 < \infty\right\}$. From the conditional Jensen's inequality applied to convex function $()^2 : \mathbb{R} \to \mathbb{R}_+$, we get that $\mathbb{E}(\mathbb{E}[X \mid \mathcal{G}])^2 \leqslant \mathbb{E}[X^2 \mid \mathcal{G}]$. Since $X \in L^2$, it follows that $X^2 \in L^1$ and hence $\mathbb{E}[X \mid \mathcal{G}] \in L^2$. It follows that $\zeta^* \triangleq \mathbb{E}[X \mid \mathcal{G}] \in L^2(\mathcal{G})$ from the definition of conditional expectation.
We first show that $X - \zeta^*$ is uncorrelated with all $\zeta \in L^2(\mathcal{G})$. Towards this end, we let $\zeta \in L^2(\mathcal{G})$ and observe that

$$\mathbb{E}[(X - \zeta^*)\zeta] = \mathbb{E}[\zeta X] - \mathbb{E}[\zeta\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\zeta X] - \mathbb{E}[\mathbb{E}[\zeta X \mid \mathcal{G}]] = 0.$$

The above equality follows from the linearity of expectation, the $\mathcal{G}$-measurability of ζ, and the definition of conditional expectation. Since $\zeta^* \in L^2(\mathcal{G})$, we have $(\zeta - \zeta^*) \in L^2(\mathcal{G})$. Therefore, $\mathbb{E}[(X - \zeta^*)(\zeta - \zeta^*)] = 0$. For any $\zeta \in L^2(\mathcal{G})$, we can write from the linearity of expectation

$$\mathbb{E}(X - \zeta)^2 = \mathbb{E}(X - \zeta^*)^2 + \mathbb{E}(\zeta - \zeta^*)^2 - 2\mathbb{E}(X - \zeta^*)(\zeta - \zeta^*) \geqslant \mathbb{E}(X - \zeta^*)^2.$$

8. *Irrelevance of independent information:* Measurability follows from the definition of conditional expectation and the definition of $\sigma(\mathcal{G}, \mathcal{H})$. Since $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable, it is $\sigma(\mathcal{G}, \mathcal{H})$ measurable.
Integrability follows from the conditional Jensen's inequality applied to convex function $|| : \mathbb{R} \to \mathbb{R}_+$ to get $|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}[|X| \mid \mathcal{G}]$, which implies that $\mathbb{E}\,|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}\,|X| < \infty$.
Orthogonality follows from the fact that it suffices to show for events $A = G \cap H \in \sigma(\mathcal{G}, \mathcal{H})$ where $G \in \mathcal{G}$ and $H \in \mathcal{H}$. In this case,

$$\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_{G \cap H}] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G\mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G]E[\mathbb{1}_H] = \mathbb{E}[X\mathbb{1}_G]\mathbb{E}[\mathbb{1}_H] = \mathbb{E}[X\mathbb{1}_{G \cap H}].$$

$\square$