Lecture-02: Conditional Expectation

1 Conditional expectation

Consider a probability space $(\Omega, \mathcal{F}, P)$.

Definition 1.1. For a random variable X, the conditional distribution conditioned on an event $E \in \mathcal{F}$ is given by

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

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

Definition 1.2. For any Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ and a random variable $X : \Omega \to \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, the conditional expectation of a random variable $g(X)$ given an event E is given by

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

Remark 2. For a random variable X and a sub-event space $\mathcal{E} \subseteq \mathcal{F}$, the conditional distribution of X conditioned on an event $E \in \mathcal{E}$ is given by $F_{X|E} \in [0,1]^{\mathbb{R}}$. Therefore, we have a a collection of distribution functions $(F_{X|E} \in [0,1]^{\mathbb{R}} : E \in \mathcal{E})$. Similarly, the conditional expectation of the random variable X given any event $E \in \mathcal{E}$ is $\mathbb{E}[X|E]$, and we have a collection $(\mathbb{E}[X \mid E] : E \in \mathcal{E})$ for the sub-event space $\mathcal{E}$.

We can generalize the conditional expectation definition for all events in an event subspace $\mathcal{E} \subseteq \mathcal{F}$.

Definition 1.3. The **conditional expectation** of X given event subspace $\mathcal{E}$ is denoted $\mathbb{E}[X|\mathcal{E}]$ and is a random variable $Z = \mathbb{E}[X|\mathcal{E}]$ where

1. **measurability:** For each $B \in \mathcal{B}(\mathbb{R})$, we have $Z^{-1}(B) \in \mathcal{E}$, and
2. **integrability:** for each event $E \in \mathcal{E}$, we have $\mathbb{E}[X \mathbb{1}_E] = \mathbb{E}[Z \mathbb{1}_E]$.

Remark 3. Any random variable $Z : \Omega \to \mathbb{R}$ that satisfies above two properties is the conditional expectation of X given the sub-event space $\mathcal{E}$ from the a.s. uniqueness of conditional expectations.

Remark 4. 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 $\mathcal{E}$.

Example 1.4. Consider two random variables X, Y defined on the same probability space $(\Omega, \mathcal{F}, P)$ with the joint distribution $F_{X,Y}(x,y) = P(\{X \leqslant x, Y \leqslant y\})$. The conditional expectation of X given Y is defined as

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

Since any Borel measurable set $B \in \mathcal{B}(\mathbb{R})$ is generated by half-open sets $(-\infty, y] \in \mathcal{B}(\mathbb{R})$ for $y \in \mathbb{R}$, any event $A \in \sigma(Y)$ is generated by the collection of events $(G_y \triangleq Y^{-1}(-\infty, y] \in \sigma(Y) : y \in \mathbb{R})$. For each $y \in \mathbb{R}$ such that $F_Y(y) = P(G_y) > 0$, we can write the conditional distribution of X given the event E_y as

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

The conditional expectation of X given the event G_y is defined as

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

We observe that $\mathbb{E}[X \mid E]$ can be evaluated for any event $E \in \sigma(Y)$ such that $P(E) > 0$. This implies that $\mathbb{E}[X \mid E] : \Omega \to \mathbb{R}$ is a $\sigma(Y)$-measurable random variable. We further observe that

$$\mathbb{E}[\mathbb{E}[X|G_y]\mathbb{1}_{G_y}] = \mathbb{E}[X|G_y]F_Y(y) \; = \int_{x\in\mathbb{R}} x d_x F_{X,Y}(x,y) = \mathbb{E}[X\mathbb{1}_{G_y}].$$

Proposition 1.5. *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*

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. ***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.*
5. ***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 $\mathbb{E}|\zeta|^2 < \infty$.*
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. ***irrelevance of independent information:*** *If $\mathcal{H}$ is independent of $\sigma(\mathcal{G},\sigma(X))$ then*

   $$\mathbb{E}[X|\sigma(\mathcal{G},\mathcal{H})] = \mathbb{E}[X \mid \mathcal{G}], \text{ a.s.}$$

   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. 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 $\varepsilon > 0$ and define $A_\varepsilon \triangleq \left\{\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[Y \mid \mathcal{G}] > \varepsilon\right\} \in \mathcal{G}$. Then from the definition of conditional expectation, we have

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

   Thus, we obtain that $P(A_\varepsilon) = 0$ for all $\varepsilon > 0$.

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. **pulling out what's known:** Let Y be $\mathcal{G}$-measurable and $\mathbb{E}|XY| < \infty$, then we need to show that $\mathbb{E}[XY\mathbb{1}_G] = \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G]$, for all events $G \in \mathcal{G}$.

   It suffices to show that $\mathbb{E}[ZX] = \mathbb{E}[Z\mathbb{E}[X \mid \mathcal{G}]]$ for any simple $\mathcal{G}$-measurable random variable Z with $\mathbb{E}|ZX| < \infty$, from which the previous statement follows for $Z = Y\mathbb{1}_G$.

   Let $Z = \sum_{k=1}^n \alpha_k\mathbb{1}_{A_k}$ for $(A_1,\dots,A_n) \subset \mathcal{G}$, then the result is a consequence of the definition of conditional expectation and linearity.

5. **L^2-projection:** We can write for $\mathcal{G}$ measurable functions ζ,ζ^* such that $\mathbb{E}\zeta^2, \mathbb{E}(\zeta^*)^2 < \infty$, from the linearity of expectation

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

   It is enough to show that $X - \mathbb{E}[X \mid \mathcal{G}]$ is orthogonal to all $\mathcal{G}$-measurable ζ such that $\mathbb{E}\zeta^2 < \infty$. Towards this end, we observe that for $\mathcal{G}$ measurable function ζ such that $\mathbb{E}\zeta^2 < \infty$, we have

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

   This implies that $\mathbb{E}(X-\zeta)^2 \geqslant \mathbb{E}(X-\zeta^*)^2$ for all $\mathcal{G}$ measurable random variables ζ that have finite second moment.

6. **tower property:** From the definition of conditional expectation, we know that $\mathbb{E}[X \mid \mathcal{H}]$ is $\mathcal{H}$ measurable, and we can verify that the mean of absolute value is finite. Let $H \in \mathcal{H} \subseteq \mathcal{G}$, then from the definition of conditional expectation, we see that

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

7. **irrelevance of independent information:** We assume $X \geqslant 0$ and show that

$$\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_A], \text{ a.s. for all } A \in \sigma(\mathcal{G},\mathcal{H}).$$

It suffices to show for $A = G \cap H$ where $G \in \mathcal{G}$ and $H \in \mathcal{H}$. We show that

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

□

Example 1.6 (Conditioning on indicator random variables). Let X be a random variable defined on the probability space $(\Omega,\mathcal{F},P)$, and $E \in \mathcal{F}$ be an event, then

$$\mathbb{E}[X|\mathbb{1}_E] = \mathbb{E}[X|E]\mathbb{1}_E + \mathbb{E}[X|E^c]\mathbb{1}_{E^c} \text{ a.s.}$$

The σ-algebra generated by the indicator random variable $\mathbb{1}_E$ is

$$\mathcal{E} \triangleq \sigma(\mathbb{1}_E) = \{\emptyset,\Omega,E,E^c\}.$$

We first observe that RHS is an $\mathcal{E}$ measurable random-variable. Second, we observe that

$$\mathbb{E}[X\mathbb{1}_E] = \mathbb{E}[\mathbb{E}[X|E]\mathbb{1}_E], \qquad \mathbb{E}[X\mathbb{1}_{E^c}] = \mathbb{E}[\mathbb{E}[X|E^c]\mathbb{1}_{E^c}.$$

Example 1.7 (Conditioning on simple random variables). Consider two random variables X,Y defined on the same probability space $(\Omega,\mathcal{F},P)$, where Y is a simple random variable such that

$$Y = \sum_{y\in\mathcal{Y}} y\mathbb{1}_{E_y} \text{ for a finite alphabet } \mathcal{Y} \subset \mathbb{R},$$

such that $E_y \triangleq Y^{-1}(\{y\}) \in \sigma(Y) \subseteq \mathcal{F}$ and $p_y \triangleq P(E_y) > 0$. The collection $(E_y \in \mathcal{F} : y \in \mathcal{Y})$ forms a finite partition of the outcome space Ω. Further, we observe that $\sigma(Y) = \left\{\cup_{y\in F} E_y \in \mathcal{F} : F \subseteq \mathcal{Y}\right\}$. The conditional distribution of X given the event E_y is

$$F_{X|E_y}(x) = \frac{P(\{X \leqslant x, Y = y\})}{p_y}.$$

The conditional expectation of X given the event E_y is defined as

$$\mathbb{E}[X|E_y] = \mathbb{E}[X|Y=y] = \int_{x\in\mathbb{R}} x dF_{X|E_y}(x) = \int_{x\in\mathbb{R}} x \int_{z=y} \frac{dF_{X,Y}(x,z)}{P(E_y)} = \frac{\mathbb{E}[X\mathbb{1}_{E_y}]}{P(E_y)}.$$

Since $P(E_y) > 0$, from the Example 1.4, we can see that

$$\mathbb{E}[X\mathbb{1}_{E_y}] = \mathbb{E}[\mathbb{E}[X|E_y]\mathbb{1}_{E_y}], \text{ for all } y \in \mathcal{Y}.$$

From the definition of conditional expectation, it follows that $\mathbb{E}[X|Y]\mathbb{1}_{E_y} = \mathbb{E}[X|E_y]\mathbb{1}_{E_y}$ for each $y \in \mathcal{Y}$. It follows that the conditional expectation of X given Y is

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