Lecture-13: Conditional expectation

1 Conditional expectation conditioned on a random vector

Let $X : \Omega \to \mathbb{R}^n$ be a random vector defined on a probability space $(\Omega, \mathcal{F}, P)$. Recall that the projection $\pi_i : \mathbb{R}^n \to \mathbb{R}$ of n-dimensional vector X on its ith component gives $X_i = \pi_i(X)$, and it is a Borel measurable function. It follows that $X_i = \pi_i \circ X$ is a random variable for each $i \in [n]$. Further the event space generated by random vectors X is given by

$$\sigma(X) = \sigma(X_1, \dots, X_n).$$

Remark 1. Let $(A_i \in \mathcal{F} : i \in [n])$ be an n-length sequence of events, then $X = (\mathbb{1}_{A_i} : i \in [n])$ is a random vector, and the smallest event space generated by this random vector is $\sigma(X) = \sigma(A_1, \dots, A_n)$.

Definition 1.1. Consider a random variable $X : \Omega \to \mathbb{R}$ such that $\mathbb{E}|X| < \infty$ and a random vector $Y : \Omega \to \mathbb{R}^n$ for some $n \in \mathbb{N}$, defined on the same probability space $(\Omega, \mathcal{F}, P)$. The **conditional expectation of random variable** X **given the random vector** Y is defined as $\mathbb{E}[X \mid Y] \triangleq \mathbb{E}[X \mid \sigma(Y)]$.

Lemma 1.2. *For a random sequence $X : \Omega \to \mathbb{R}^{\mathbb{N}}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, we have*

$$\sigma(X_1, \dots, X_n) \subseteq \sigma(X_1, \dots, X_{n+1}), \quad n \in \mathbb{N}.$$

Proof. For any $x \in \mathbb{R}^{n+1}$, any generating event for collection $\sigma(X_1, \dots, X_n)$ is of the form $\cap_{i=1}^n X_i^{-1}(-\infty, x_i] = \cap_{i=1}^n X_i^{-1}(-\infty, x_i] \cap X_{n+1}^{-1}(\mathbb{R})$, a generating event for collection $\sigma(X_1, \dots, X_{n+1})$. □

2 Properties of Conditional Expectation

Proposition 2.1. *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 $\mathcal{G}, \mathcal{H} \subset \mathcal{F}$. Then the following statements are true.*

1. ***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 $c \in \mathbb{R}$.*
2. ***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.*
3. ***Monotonicity:*** *If $X \leqslant Y$ a.s., then $\mathbb{E}[X \mid \mathcal{G}] \leqslant E[Y \mid \mathcal{G}]$ a.s.*
4. ***Conditional Jensen's inequality:*** *If $\psi : \mathbb{R} \to \mathbb{R}$ is convex and $\mathbb{E}|\psi(X)| < \infty$, then $\psi(\mathbb{E}[X \mid \mathcal{G}]) \leqslant \mathbb{E}[\psi(X) \mid \mathcal{G}]$ 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. ***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.*
8. ***L^2-projection:*** *If $\mathbb{E}|X|^2 < \infty$, then $\zeta^* \triangleq \mathbb{E}[X \mid \mathcal{G}]$ minimizes $\mathbb{E}[(X - \zeta)^2]$ over all $\mathcal{G}$-measurable random variables ζ such that $\mathbb{E}|\zeta|^2 < \infty$.*

Proof. In all the properties below, we have to show that conditional expectation given an event space equals another random variable almost surely. To this end, we will show that the right hand side satisfies the three properties of the conditional expectation random variable, and hence is the conditional expectation from almost sure uniqueness.

1. **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, $\mathbb{E}[c \mid \mathcal{G}] = c$.

2. **Linearity:** Let $Z \triangleq \mathbb{E}[(\alpha X + \beta Y) \mid \mathcal{G}]$ and $Z_1 \triangleq \mathbb{E}[X \mid \mathcal{G}]$ and $Z_2 \triangleq \mathbb{E}[Y \mid \mathcal{G}]$. We have to show that $Z = \alpha Z_1 + \beta Z_2$ almost surely.

   (i) From the definition of conditional expectation, we have $\mathbb{E}[X \mid \mathcal{G}], \mathbb{E}[Y \mid \mathcal{G}]$ are $\mathcal{G}$-measurable. In addition, the linear map $g : \mathbb{R}^2 \to \mathbb{R}$ defined by $g(x) \triangleq \alpha x_1 + \beta x_2$ is Borel measurable. Therefore, the linear combination $\alpha Z_1 + \beta Z_2$ is $\mathcal{G}$-measurable.

   (ii) For any $G \in \mathcal{G}$, from the linearity of expectation and definition of conditional expectation, we have
   $$\mathbb{E}[(\alpha Z_1 + \beta Z_2)\mathbb{1}_G] = \alpha \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G] + \beta \mathbb{E}[\mathbb{E}[Y \mid \mathcal{G}]\mathbb{1}_G] = \mathbb{E}[(\alpha X + \beta Y)\mathbb{1}_G].$$

   (iii) From the definition of conditional expectation, we have $\mathbb{E}\,|Z_1|, \mathbb{E}\,|Z_2|$ are finite. Therefore, we have
   $$\mathbb{E}\,|\alpha Z_1 + \beta Z_2| \leqslant |\alpha|\,\mathbb{E}\,|Z_1| + |\beta|\,\mathbb{E}\,|Z_2| < \infty.$$

   This implies that $\alpha Z_1 + \beta Z_2$ satisfies three properties of conditional expectation $\mathbb{E}[(\alpha X + \beta Y) \mid \mathcal{G}]$. From the almost sure uniqueness of conditional expectation, we have $Z = \alpha Z_1 + \beta Z_2$ almost surely.

3. **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
   $$0 \leqslant \epsilon P(A_\epsilon) = \mathbb{E}[\epsilon \mathbb{1}_{A_\epsilon}] < \mathbb{E}[(\mathbb{E}[X \mid \mathcal{G}] - \mathbb{E}[Y \mid \mathcal{G}])\mathbb{1}_{A_\epsilon}] = \mathbb{E}[(X - Y)\mathbb{1}_{A_\epsilon}] \leqslant 0.$$
   Thus, we obtain that $P(A_\epsilon) = 0$ for all $\epsilon > 0$. Taking limit $\epsilon \downarrow 0$, we get
   $$0 = \lim_{\epsilon \downarrow 0} P(A_\epsilon) = P(\lim_{\epsilon} A_\epsilon) = P(A_0).$$

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:** From the almost sure uniqueness of conditional expectation, it suffices to show that $Y\mathbb{E}[X \mid \mathcal{G}]$ satisfies following three properties of the conditional expectation $\mathbb{E}[XY \mid \mathcal{G}]$:

   (i) the random variable $Y\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable,

   (ii) $\mathbb{E}[XY\mathbb{1}_G] = \mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G]$ for any event $G \in \mathcal{G}$, and

   (iii) $\mathbb{E}\,|Y\mathbb{E}[X \mid \mathcal{G}]|$ is finite.

   Part (i) is true since Y is given to be $\mathcal{G}$-measurable, $\mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$-measurable by the definition of conditional expectation, and product function is Borel measurable.

   It suffices to show part (ii) and (iii) for simple $\mathcal{G}$-measurable random variables $Y = \sum_{y \in \mathcal{Y}} y \mathbb{1}_{E_y}$ where $E_y = Y^{-1}\{y\} \in \mathcal{G}$. For any $G \in \mathcal{G}$, we have $G \cap E_y \in \mathcal{G}$ for all $y \in \mathcal{Y}$ and $\cup_{y \in \mathcal{Y}}(G \cap E_y) = G$.

   Part (ii) follow from the linearity of expectation and definition of conditional expectation $\mathbb{E}[X \mid \mathcal{G}]$, such that
   $$\mathbb{E}[Y\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_G] = \sum_{y \in \mathcal{Y}} y \mathbb{E}[\mathbb{1}_{G \cap E_y}\mathbb{E}[X \mid \mathcal{G}]] = \sum_{y \in \mathcal{Y}} y \mathbb{E}[X\mathbb{1}_{G \cap E_y}] = \mathbb{E}[X \sum_{y \in \mathcal{Y}} y \mathbb{1}_{G \cap E_y}] = \mathbb{E}[XY\mathbb{1}_G].$$

Part (iii) follows from the fact that $\mathbb{E}|XY|$ is finite and the conditional Jensen's inequality applied to convex function $|\cdot| : \mathbb{R} \to \mathbb{R}_+$ to 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:** From the almost sure uniqueness of conditional expectation, it suffices to show that

   (i) $\mathbb{E}[X \mid \mathcal{H}]$ is $\mathcal{H}$-measurable,

   (ii) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{H}] \mathbb{1}_H] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mathbb{1}_H]$ for all $H \in \mathcal{H}$, and

   (iii) $\mathbb{E}[|\mathbb{E}[X \mid \mathcal{H}]|]$ is finite.

   Part (i) follows from the definition of conditional expectation, which implies that $\mathbb{E}[X \mid \mathcal{H}]$ is $\mathcal{H}$ measurable.

   Part (ii) follows from the fact that $\mathcal{H} \subseteq \mathcal{G}$, and hence any $H \in \mathcal{H}$ belongs to $\mathcal{G}$. Therefore, from the definition of conditional expectation, we have

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

   Part (iii) follows from the conditional Jensen's inequality applied to convex function $|\cdot| : \mathbb{R} \to \mathbb{R}_+$ to get $|\mathbb{E}[X \mid \mathcal{H}]| \leqslant \mathbb{E}[|X| \mid \mathcal{H}]$, and hence $\mathbb{E}\,|\mathbb{E}[X \mid \mathcal{H}]| \leqslant \mathbb{E}\,|X| < \infty$.

7. **Irrelevance of independent information:** From the almost sure uniqueness of conditional expectation, it suffices to show that

   (i) $\mathbb{E}[X \mid \mathcal{G}]$ is $\sigma(\mathcal{G}, \mathcal{H})$-measurable,

   (ii) $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}] \mathbb{1}_G] = \mathbb{E}[X \mathbb{1}_G]$ for all $G \in \sigma(\mathcal{G}, \mathcal{H})$, and

   (iii) $\mathbb{E}\,|\mathbb{E}[X \mid \mathcal{G}]|$ is finite.

   Part (i) 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.

   Part (ii) 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,

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

   Part (iii) follows from the conditional Jensen's inequality applied to convex function $|\cdot| : \mathbb{R} \to \mathbb{R}_+$ to get $|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}[|X| \mid \mathcal{G}]$. This implies that $\mathbb{E}\,|\mathbb{E}[X \mid \mathcal{G}]| \leqslant \mathbb{E}\,|X| < \infty$.

8. **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 $(\cdot)^2 : \mathbb{R} \to \mathbb{R}_+$, we get that $(\mathbb{E}[X \mid \mathcal{G}])^2 \leqslant \mathbb{E}[X^2 \mid \mathcal{G}]$ a.s.. 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

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

□