Lecture-10: Conditional expectation

1 Conditional Expectation

Definition 1.1 (Conditional expectation). Consider two random variables X, Y defined on the same probability space $(\Omega, \mathcal{F}, P)$.

(i) The **conditional distribution** of the random variable X given a non-trivial event B generated by the random variable Y is
$$F_{X \mid B}(x) = \frac{P(\{X \leqslant x\} \cap B)}{P(B)}.$$
We can define $F_{X \mid B} = 0$ for trivial events $B \in \sigma(Y)$ such that $P(B) = 0$.

(ii) The **conditional expectation** of random variable X given any event B generated by random variable Y is given by
$$\mathbb{E}[X \mid B] = \int_{x \in \mathbb{R}} x dF_{X \mid B}.$$

(iii) The **conditional distribution** of the random variable X given the random variable Y, is a function of the random variable Y, and hence a random variable. In particular, for each event $\{Y \leqslant y\}$
$$F_{X \mid Y} \mathbb{1}_{\{Y \leqslant y\}} = F_{X \mid \{Y \leqslant y\}} \mathbb{1}_{\{Y \leqslant y\}} = \frac{F_{X,Y}(\cdot, y)}{F_Y(y)} \mathbb{1}_{\{Y \leqslant y\}} \text{ for all } y \in \mathbb{R}.$$
That is, this conditional distribution is the collection $(F_{X \mid B} : B \in \sigma(Y))$ of conditional distributions for each event B generated by the random variable Y. The probability of the event B and the probability of the conditional distribution $F_{X \mid B} \mathbb{1}_B$ is $P(B)$.

(iv) The **conditional expectation** of random variable X given the random variable Y, is a function of the random variable Y, and hence a random variable. In particular, for each event $\{Y \leqslant y\}$, we have the event
$$\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}} = \mathbb{E}[X \mid \{Y \leqslant y\}] \mathbb{1}_{\{Y \leqslant y\}} = \mathbb{1}_{\{Y \leqslant y\}} \int_{x \in \mathbb{R}} x dF_{X \mid \{Y \leqslant y\}}(x), \text{ for all } y \in \mathbb{R}.$$
That is, this conditional expectation is the collection of events $(\mathbb{E}[X|B] : B \in \sigma(Y))$ corresponding to the conditional expectation given the events generated by the random variable Y. The probability of the event B and the probability of the conditional expectation $\mathbb{E}[X \mid B] \mathbb{1}_B$ is $P(B)$.

(v) The mean of random variable $\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}}$ for all $y \in \mathbb{R}$ is given by
$$\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}}] = \int_{t \in \mathbb{R}} dF_Y(t) \mathbb{1}_{\{Y \leqslant y\}} \int_{x \in \mathbb{R}} x dF_{X \mid \{Y \leqslant y\}}(x) = \int_{(x,t) \in \mathbb{R}^2} x \mathbb{1}_{\{t \leqslant y\}} dF_{X,Y}(x,t) = \mathbb{E}[X \mathbb{1}_{\{Y \leqslant y\}}].$$
Similarly, for all events $B \in \sigma(Y)$, we have the event
$$\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B].$$

That is, the conditional expectation $\mathbb{E}[X \mid Y]$ is a function of Y and $\sigma(Y)$ measurable, and for all events $B \in \sigma(Y)$, we have
$$\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B].$$

Definition 1.2 (Conditional expectation). Consider a probability space $(\Omega, \mathcal{F}, P)$, a smaller event space $\mathcal{G} \subset \mathcal{F}$, a random variable X such that $\mathbb{E}|X| < \infty$. Conditional expectation of X given $\mathcal{G}$ is denoted by $Y \triangleq \mathbb{E}[X \mid \mathcal{G}]$, a random variable on the same probability space such that

(i) Y is $\mathcal{G}$ measurable, i.e. $Y^{-1}(-\infty, y] \in \mathcal{G}$ for all $y \in \mathbb{R}$,

(ii) for all $A \in \mathcal{G}$, we have $\int_A X dP = \mathbb{E}[X \mathbb{1}_A] = \mathbb{E}[Y \mathbb{1}_A] = \int_A Y dP$,

(iii) $\mathbb{E}|Y| < \infty$.

Definition 1.3. A random variable is independent of an event space $\mathcal{G}$ if

$$P(X^{-1}(-\infty, x] \cap B) = F_X(x) P(B), \text{ for all } x \in \mathbb{R}, B \in \mathcal{G}.$$

Proposition 1.4. *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*

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 $A \in \mathcal{G}$, from the linearity of expectation and definition of conditional expectation, we have

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

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 $\epsilon > 0$.

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 $\mathbb{E}[XY \mathbb{1}_A] = \mathbb{E}[Y \mathbb{E}[X \mid \mathcal{G}] \mathbb{1}_A]$, for all $A \in \mathcal{G}$.

   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 $Z = Y \mathbb{1}_A$.

   Let $Z = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ for $(A_1, \ldots, A_n) \subset \mathcal{G}$, 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

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

□

1.1 Conditional expectation conditioned on a random vector

Consider the probability space $(\Omega, \mathcal{F}, P)$, and a random vector $X : \Omega \to \mathbb{R}^n$. Let $\pi_i : \mathbb{R}^n \to \mathbb{R}$ be the projection of an n-length vector to its ith component, i.e. $\pi_i(X) = X_i$.

Exercise 1.5. Show that the function $\pi_i : \mathbb{R}^n \to \mathbb{R}$ defined by $g : x \mapsto x_i$ is a Borel measurable function.

Since $\pi_i : \mathbb{R}^n \to \mathbb{R}$ is Borel measurable function, $X_i = \pi_i \circ X$ is a random variable for each $i \in [n]$.

Definition 1.6 (Event space generated by a random vector). Let $X : \Omega \to \mathbb{R}^n$ be a random vector, then the smallest event space generated by the events of the form $X^{-1}(-\infty, x_1] \times \cdots \times (-\infty, x_n]$ for $x \in \mathbb{R}^n$ is called the **event space generated** by this random vector X, and denoted by $\sigma(X)$.

Example 1.7 (Indicator functions). Let $A, B \in \mathcal{F}$ be events, then $X = (\mathbb{1}_A, \mathbb{1}_B)$ is a random vector and

$$X^{-1}(-\infty, x_1] \times (-\infty, x_2] = X_1^{-1}(-\infty, x_1] \cap X_2^{-1}(-\infty, x_2] = \begin{cases} \Omega, & \min\{x_1, x_2\} \geqslant 1, \\ A^c, & x_1 \in [0,1), x_2 \geqslant 1, \\ B^c, & x_1 \geqslant 1, x_2 \in [0,1), \\ A^c \cap B^c, & x_1, x_2 \in [0,1), \\ \emptyset, & \min\{x_1, x_2\} < 0. \end{cases}$$

This implies that the smallest event space generated by this random vector is

$$\sigma(X) = \{\emptyset, A, A^c, B, B^c, A \cup B, A \cup B^c, A \cap B, A \cap B^c, A^c \cup B, A^c \cup B^c, A^c \cap B, A^c \cap B^c, \Omega\} = \sigma(\emptyset, A, B, \Omega).$$

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 $\sigma(X) = \sigma(\emptyset, \Omega, A_1, \dots, A_n)$.

Lemma 1.8. *For a sequence of random variables $(X_i : i \in \mathbb{N})$ defined on the same probability space $(\Omega, \mathcal{F}, P)$, we have*

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

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

Definition 1.9 (Conditional expectation given a random vector). Consider a random variable $X : \Omega \to \mathbb{R}$ and a random vector $Y : \Omega \to \mathbb{R}^n$ for some $n \in \mathbb{N}$, defined on the same probability space $(\Omega, \mathcal{F}, P)$.

(i) The **conditional distribution** of the random variable X given a non-trivial event B generated by the random vector Y is

$$F_{X \mid B}(x) = \frac{P(\{X \leqslant x\} \cap B)}{P(B)}.$$

We can define $F_{X \mid B} = 0$ for trivial events $B \in \sigma(Y)$ such that $P(B) = 0$.

(ii) The **conditional expectation** of random variable X given any event B generated by random vector Y is given by

$$\mathbb{E}[X \mid B] = \int_{x \in \mathbb{R}} x dF_{X \mid B}.$$

(iii) The **conditional distribution** of the random variable X given the random vector Y, is a function of the random vector Y, and hence a random variable. In particular, for each event $\{Y \leqslant y\} = \cap_{i=1}^n \{Y_i \leqslant y_i\}$

$$F_{X \mid Y} \mathbb{1}_{\{Y \leqslant y\}} = F_{X \mid \{Y \leqslant y\}} \mathbb{1}_{\{Y \leqslant y\}} = \frac{F_{X,Y}(\cdot, y)}{F_Y(y)} \mathbb{1}_{\{Y \leqslant y\}} \text{ for all } y \in \mathbb{R}^n.$$

That is, this conditional distribution is the collection $(F_{X \mid B} : B \in \sigma(Y))$ of conditional distributions for each event B generated by the random vector Y. The probability of the event B and $F_{X \mid B}$ is $P(B)$.

(iv) The **conditional expectation** of random variable X given the random vector Y, is a function of the random vector Y, and hence a random variable. In particular, for each event $\{Y \leqslant y\} = \cap_{i=1}^n \{Y_i \leqslant y_i\}$, we have the event

$$\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}} = \mathbb{E}[X \mid \{Y \leqslant y\}] \mathbb{1}_{\{Y \leqslant y\}} = \mathbb{1}_{\{Y \leqslant y\}} \int_{x \in \mathbb{R}} x dF_{X \mid \{Y \leqslant y\}}(x), \text{ for all } y \in \mathbb{R}^n.$$

That is, this conditional expectation is the collection of events $(\mathbb{E}[X|B] : B \in \sigma(Y))$ corresponding to the conditional expectation given the events generated by the random vector Y. The probability of the event B and $\mathbb{E}[X \mid B]$ is $P(B)$.

(v) The mean of random variable $\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}}$ for all $y \in \mathbb{R}^n$ is given by

$$\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_{\{Y \leqslant y\}}] = \int_{t \in \mathbb{R}^n} dF_Y(t) \mathbb{1}_{\{Y \leqslant y\}} \int_{x \in \mathbb{R}} x dF_{X \mid \{Y \leqslant y\}}(x) = \int_{(x,t) \in \mathbb{R}^{n+1}} x \mathbb{1}_{\{t \leqslant y\}} dF_{X,Y}(x, t) = \mathbb{E}[X \mathbb{1}_{\{Y \leqslant y\}}].$$

Similarly, for all events $B \in \sigma(Y)$, we have the event

$$\mathbb{E}[\mathbb{E}[X \mid Y] \mathbb{1}_B] = \mathbb{E}[X \mathbb{1}_B].$$