Lecture-12: Conditional Expectation

1 Conditional expectation given a non trivial event

Consider a probability space $(\Omega, \mathcal{F}, P)$ and an event $B \in \mathcal{F}$ such that $P(B) > 0$. Then, the conditional probability of any event $A \in \mathcal{F}$ given an event B is defined as

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Consider a random variable $X : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$, with distribution function $F_X : \mathbb{R} \to [0,1]$, and a non trivial event $B \in \mathcal{F}$ such that $P(B) > 0$.

Definition 1.1. The **conditional distribution of** X **given event** B is denoted by $F_{X|B} : \mathbb{R} \to [0,1]$ and $F_{X|B}(x)$ is defined as the probability of event $A_X(x) \triangleq X^{-1}(-\infty, x]$ conditioned on event B for all $x \in \mathbb{R}$. That is,

$$F_{X|B}(x) \triangleq P(A_X(x)|B) = \frac{P(A_X(x) \cap B)}{P(B)} \text{ for all } x \in \mathbb{R}.$$

Remark 1. The conditional distribution $F_{X|B}$ is a distribution function. This follows from the fact that (i) $F_{X|B} \geqslant 0$, (ii) $F_{X|B}$ is right continuous, (iii) $\lim_{x \downarrow -\infty} F_{X|B}(x) = 0$ and $\lim_{x \uparrow \infty} F_{X|B}(x) = 1$.

Remark 2. For a discrete random variable $X : \Omega \to \mathcal{X}$, the conditional probability mass function of X given a non trivial event B is given by $P_{X|B}(x) = \frac{P(X^{-1}\{x\} \cap B)}{P(B)}$ for all $x \in \mathcal{X}$.

Remark 3. For a continuous random variable $X : \Omega \to \mathbb{R}$, the conditional density of X given a non trivial event B is given by $f_{X|B}(x) = \frac{dF_{X|B}(x)}{dx}$ for all $x \in \mathbb{R}$.

Example 1.2 (Conditional distribution). Consider the probability space $(\Omega, \mathcal{F}, P)$ corresponding to a random experiment where a fair die is rolled once. For this case, the outcome space $\Omega = [6]$, the event space $\mathcal{F} = \mathcal{P}([6])$, and the probability measure $P(\omega) = \frac{1}{6}$ for all $\omega \in \Omega$.

We define a random variable $X : \Omega \to \mathbb{R}$ such that $X(\omega) = \omega$ for all $\omega \in \Omega$, and an event $B \triangleq \{\omega \in \Omega : X(\omega) \leqslant 3\} = [3] \in \mathcal{F}$. We note that $P(B) = 0.5$ and the conditional PMF of X given B is

$$P_{X|B}(x) = \frac{1}{3}\mathbb{1}_{\{x=3\}} + \frac{1}{3}\mathbb{1}_{\{x=2\}} + \frac{1}{3}\mathbb{1}_{\{x=1\}}.$$

Definition 1.3. The **conditional expectation of** X **given event** B is given as $\mathbb{E}[X|B] \triangleq \int_{x \in \mathbb{R}} x dF_{X|B}(x)$.

Remark 4. For a discrete random variable $X : \Omega \to \mathcal{X}$, the conditional expectation of X given a non trivial event B is given by $\mathbb{E}[X|B] = \sum_{x \in \mathcal{X}} x P_{X|B}(x)$.

Example 1.4 (Conditional expectation). For the random variable X and event B defined in Example 1.2, the conditional expectation $\mathbb{E}[X|B] = 2$.

Remark 5. Consider two random variables X, Y defined on this probability space, then for $y \in \mathbb{R}$ such that $F_Y(y) > 0$, we can define events $A_X(x) \triangleq X^{-1}(-\infty, x]$ and $A_Y(y) = Y^{-1}(-\infty, y]$, such that

$$P\left(\{X \leqslant x\} \mid \{Y \leqslant y\}\right) = \frac{F_{X,Y}(x,y)}{F_Y(y)}.$$

The key observation is that $\{Y \leqslant y\}$ is a non-trivial event. How do we define conditional expectation based on events such as $\{Y = y\}$? When random variable Y is continuous, this event has zero probability measure.

2 Conditional expectation given an event space

Consider random variables $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ defined on the same probability space $(\Omega, \mathcal{F}, P)$ such that $\mathbb{E}|X| < \infty$, and a smaller event space $\mathcal{G} \subset \mathcal{F}$. For each non trivial event $G \in \mathcal{G}$, we know how to define the conditional distribution $F_{X|G}$ and $\mathbb{E}[X|G]$. For any trivial event $N \in \mathcal{G}$, these are undefined.

Definition 2.1. The **conditional expectation of random variable** X **given event space** $\mathcal{G}$ is a random variable $\mathbb{E}[X \mid \mathcal{G}] : \Omega \to \mathbb{R}$ defined on the same probability space, such that

(i) $Z \triangleq \mathbb{E}[X \mid \mathcal{G}]$ is $\mathcal{G}$ measurable,
(ii) for all $G \in \mathcal{G}$, we have $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[Z\mathbb{1}_G]$,
(iii) $\mathbb{E}|Z| < \infty$.

Lemma 2.2. *The conditional expectation of X given $\mathcal{G}$ is an a.s. unique random variable.*

Proof. Consider two random variables $Z_1 = \mathbb{E}[X|\mathcal{G}]$ and $Z_2 = \mathbb{E}[X|\mathcal{G}]$. Then from the definition, Z_1, Z_2 are $\mathcal{G}$ measurable random variables, and $Z_1 - Z_2$ is also $\mathcal{G}$ measurable. Therefore, $G_n \triangleq \left\{Z_1 - Z_2 > \frac{1}{n}\right\} \in \mathcal{G}$ and $\mathbb{E}[(Z_1 - Z_2)\mathbb{1}_{G_n}] = 0$ by definition. It follows from continuity of probability, that $P(\lim_n G_n) = 0$. Similarly, defining $F_n \triangleq \left\{Z_2 - Z_1 > \frac{1}{n}\right\}$, we can show that $P(\lim_n F_n) = 0$. □

Example 2.3 (Conditional expectation as averaging). Consider a random variable $X : \Omega \to \mathbb{R}$ defined on a probability space $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}|X| < \infty$, and the coarsest event space $\mathcal{G} = \{\emptyset, \Omega\} \subseteq \mathcal{F}$ and finest event space $\mathcal{F}$. We observe that $\mathbb{E}[X|\mathcal{G}] = \mathbb{E}X$ a.s. uniquely, since (i) $\mathbb{E}X$ is a constant and hence $\mathcal{G}$ measurable, (ii) $\mathbb{E}[\mathbb{E}X\mathbb{1}_\emptyset] = \mathbb{E}[X\mathbb{1}_\emptyset] = 0$ and $\mathbb{E}[\mathbb{E}X\mathbb{1}_\Omega] = \mathbb{E}[X\mathbb{1}_\Omega] = \mathbb{E}X$, and (iii) $\mathbb{E}|\mathbb{E}X| = |\mathbb{E}X| \leqslant \mathbb{E}|X| < \infty$ from the Jensen's inequality.

We also observe that $\mathbb{E}[X|\mathcal{F}] = X$ a.s. uniquely, since (i) X is $\mathcal{F}$ measurable random variable, (ii) $\mathbb{E}[X\mathbb{1}_G] = \mathbb{E}[X\mathbb{1}_G]$ for all events $G \in \mathcal{F}$, and (iii) $\mathbb{E}|X| < \infty$.

Lemma 2.4. *The mean of conditional expectation of random variable X given event space $\mathcal{G}$ is $\mathbb{E}X$.*

Proof. From the definition of event space $\Omega \in \mathcal{G}$, and from the definition of conditional expectation, we get $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]\mathbb{1}_\Omega] = \mathbb{E}[X\mathbb{1}_\Omega] = \mathbb{E}X$. □

Definition 2.5. The conditional expectation of X given Y is a random variable $\mathbb{E}[X \mid Y] \triangleq \mathbb{E}[X \mid \sigma(Y)]$ defined on the same probability space.

Example 2.6 (Conditioning on simple random variables). For a simple random variable $Y : \Omega \to \mathcal{Y} \subseteq \mathbb{R}$ defined on the probability space $(\Omega, \mathcal{F}, P)$, we define fundamental events $E_y \triangleq Y^{-1}\{y\} \in \mathcal{F}$ for all $y \in \mathcal{Y}$. Then the sequence of events $E \triangleq (E_y \in \mathcal{F} : y \in \mathcal{Y})$ partitions the sample space, and we can write the event space generated by random variable Y as $\sigma(Y) = (\cup_{y \in I} E_y : I \subseteq \mathcal{Y})$.

For a random variable $X : \Omega \to \mathbb{R}$ defined on the same probability space, the random variable $Z \triangleq \mathbb{E}[X \mid Y]$ is $\sigma(Y)$ measurable. Therefore, $\mathbb{E}[X|Y] = \sum_{y \in \mathcal{Y}} \alpha_y \mathbb{1}_{E_y}$ for some $\alpha \in \mathbb{R}^{\mathcal{Y}}$. We verify that $Z : \Omega \to \mathbb{R}$

is a $\sigma(Y)$ measurable random variable, since $\sigma(Z) \subseteq \sigma(Y)$. We can also check that $\mathbb{E}|Z| < \infty$. Further, we have $\mathbb{E}[Z\mathbb{1}_{E_y}] = \mathbb{E}[X\mathbb{1}_{E_y}]$ for any $y \in \mathcal{Y}$, which implies that $\alpha_y = \frac{\mathbb{E}[X\mathbb{1}_{E_y}]}{P_Y(y)}$ for any $y \in \mathcal{Y}$. Notice that

$$\mathbb{E}[X \mid E_y] = \int_{x\in\mathbb{R}} x dF_{X|E_y}(x) = \frac{1}{P_Y(y)} \int_{x\in\mathbb{R}} x dP(A_X(x) \cap E_y) = \frac{1}{P_Y(y)} \mathbb{E}[X\mathbb{1}_{E_y}] = \alpha_y.$$

Remark 6. There are three main takeaways from this definition. For a random variable Y, the event space generated by Y is $\sigma(Y)$.

1. The conditional expectation $\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$ and is $\sigma(Y)$ measurable. That is, $\mathbb{E}[X|Y]$ is a Borel measurable function of Y. In particular when Y is discrete, this implies that $\mathbb{E}[X|Y]$ is a simple random variable that takes value $\mathbb{E}[X|E_y]$ when $\omega \in E_y$, and the probability of this event is $P_Y(y)$. When Y is continuous, $\mathbb{E}[X|Y]$ is a continuous random variable with density f_Y.
2. Expectation is averaging. Conditional expectation is averaging over event spaces. We can observe that the coarsest averaging is $\mathbb{E}[X|\{\emptyset, \Omega\}] = \mathbb{E}X$ and the finest averaging is $\mathbb{E}[X|\sigma(X)] = X$. Further, $\mathbb{E}[X|\sigma(Y)]$ is averaging of X over events generated by Y. If we take any event $A \in \sigma(Y)$ generated by Y, then the conditional expectation of X given Y is fine enough to find the averaging of X when this event occurs. That is, $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[\mathbb{E}[X|Y]\mathbb{1}_A]$.
3. If $X \in L^1$, then the conditional expectation $\mathbb{E}[X|Y] \in L^1$.

3 Conditional distribution given an event space

Definition 3.1. The **conditional probability of an event** $A \in \mathcal{F}$ **given event space** $\mathcal{G}$ is defined as $P(A \mid \mathcal{G}) \triangleq \mathbb{E}[\mathbb{1}_A \mid \mathcal{G}]$.

Remark 7. From the definition of conditional expectation, it follows that $P(A \mid \mathcal{G}) : \Omega \to [0,1]$ is a $\mathcal{G}$ measurable random variable, such that $\mathbb{E}[\mathbb{1}_G P(A|\mathcal{G})] = P(A \cap G)$ for all $G \in \mathcal{G}$, and is uniquely defined up to sets of probability zero.

Example 3.2. For the trivial sigma algebra $\mathcal{G} = \{\emptyset, \Omega\}$, the conditional probability is the constant function $P(A \mid \{\emptyset, \Omega\}) = P(A)$.

Example 3.3. If $A \in \mathcal{G}$, then $P(A \mid \mathcal{G}) = \mathbb{1}_A$.

Definition 3.4. The **conditional distribution of random variable** X **given sub event space** $\mathcal{G}$ is defined as $F_{X|\mathcal{G}}(x) \triangleq P(A_X(x) \mid \mathcal{G})$ for all $x \in \mathbb{R}$.

Remark 8. Recall that $F_{X|\mathcal{G}}(x) : \Omega \to [0,1]$ a random variable, for each $x \in \mathbb{R}$. Further, we observe that $F_{X|\mathcal{G}}$ is monotone nondecreasing in $x \in \mathbb{R}$, right continuous at all $x \in \mathbb{R}$, and has limits $\lim_{x\downarrow-\infty} F_{X|\mathcal{G}}(x) = 0$ and $\lim_{x\uparrow\infty} F_{X|\mathcal{G}}(x) = 1$. It follows that $F_{X|\mathcal{G}} : \Omega \to [0,1]^{\mathbb{R}}$ is a random distribution.

Theorem 3.5. *Let $g : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function and $\mathcal{G}$ be a sub-event space. Then, the conditional expectation $\mathbb{E}[g(X) \mid \mathcal{G}] = \int_{x\in\mathbb{R}} g(x) dF_{X|\mathcal{G}}(x)$.*

Proof. It suffices to show this for simple random variables $X : \Omega \to \mathcal{X}$. Since g is Borel measurable, then $g(X)$ is a random variable. We will show that $\mathbb{E}[g(X) \mid \mathcal{G}] = \sum_{x\in\mathcal{X}} g(x) P_{X|\mathcal{G}}(x)$ by showing that it satisfies three properties of conditional expectation. For part (i), we observe that from the definition of conditional probability $P_{X|\mathcal{G}}(x)$ is a $\mathcal{G}$-measurable random variable for all $x \in \mathcal{X}$, and so is the linear combination

$\sum_{x\in\mathcal{X}} g(x)P_{X|\mathcal{G}}(x)$. For part (ii), we let $G \in \mathcal{G}$. Then, it follows from the linearity of expectation and the definition of conditional probability, that

$$\mathbb{E}[\sum_{x\in\mathcal{X}} g(x)P_{X|\mathcal{G}}(x)\mathbb{1}_G] = \sum_{x\in\mathcal{X}} g(x)\mathbb{E}[P_{X|\mathcal{G}}(x)\mathbb{1}_G] = \sum_{x\in\mathcal{X}} g(x)\mathbb{E}[\mathbb{1}_{E_x\cap G}] = \mathbb{E}[X\mathbb{1}_G].$$

For part (iii), it follows from the triangle inequality, the linearity of expectation, and the definition of conditional probability that $\mathbb{E}\left|\sum_{x\in\mathcal{X}} g(x)P_{X|\mathcal{G}}(x)\right| \leqslant \sum_{x\in\mathcal{X}} |g(x)|\,\mathbb{E}P_{X|\mathcal{G}}(x) = \sum_{x\in\mathcal{X}} |g(x)|\,P_X(x) = \mathbb{E}\,|X| < \infty$. □

Remark 9. The conditional characteristic function is given by $\Phi_{X|\mathcal{G}}(u) = \mathbb{E}[e^{juX} \mid \mathcal{G}] = \int_{x\in\mathbb{R}} e^{jux}dF_{X|\mathcal{G}}(x)$.

Definition 3.6. The **conditional distribution of random variable** X **given random variable** Y is defined as $F_{X|Y}(x) \triangleq P(A_X(x) \mid \sigma(Y)]$ for all $x \in \mathbb{R}$.

Example 3.7 (Conditional distribution given simple random variables). Consider a random variable $X : \Omega \to \mathbb{R}$ and a simple random variable $Y : \Omega \to \mathcal{Y}$ defined on the same probability space. Since random variables $F_{X|Y}(x) = \mathbb{E}[\mathbb{1}_{A_X(x)} \mid Y]$ are $\sigma(Y)$ measurable, they can can be written as $F_{X|Y}(x) = \sum_{y\in\mathcal{Y}} \beta_{x,y}\mathbb{1}_{E_y}$ for some $\beta_x \in \mathbb{R}^{\mathcal{Y}}$ and $E_y = Y^{-1}\{y\}$ for all $y \in \mathcal{Y}$. Further, we have $\mathbb{E}[F_{X|Y}(x)\mathbb{1}_{E_y}] = \mathbb{E}[\mathbb{1}_{A_X(x)}\mathbb{1}_{E_y}]$ for any $y \in \mathcal{Y}$, which implies that $\beta_{x,y} = \frac{P(A_X(x)\cap E_y)}{P_Y(y)} = F_{X|E_y}(x)$ for any $y \in \mathcal{Y}$. It follows that $F_{X|Y}$ is a $\sigma(Y)$ measurable simple random variable.

Example 3.8 (Conditional expectation). Consider a random experiment of a fair die being thrown and a random variable $X : \Omega \to \mathbb{R}$ taking the value of the outcome of the experiment. That is, for outcome space $\Omega = [6]$ and event space $\mathcal{F} = \mathcal{P}(\Omega)$, we have $X(\omega) = \omega$ with $P_X(x) = 1/6$ for $x \in [6]$. Define another random variable $Y = \mathbb{1}_{\{X\leqslant 3\}}$. Then the conditional expectation of X given Y is a random variable given by

$$\mathbb{E}[X|Y] = \mathbb{E}[X \mid \{Y = 1\}]\mathbb{1}_{\{Y=1\}} + \mathbb{E}[X \mid \{Y = 1\}]\mathbb{1}_{\{Y=0\}} = 2\mathbb{1}_{Y^{-1}(1)} + 5\mathbb{1}_{Y^{-1}(0)}.$$

Since $P\{Y = 1\} = P\{Y = 0\} = 0.5$, it follows that that $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X] = 3.5$.

Example 3.9 (Conditional distribution). Consider the zero-mean Gaussian random variable $N : \Omega \to \mathbb{R}$ with variance σ^2, and another independent random variable $Y \in \{-1,1\}$ with PMF $(1 - p, p)$ for some $p \in [0,1]$. Let $X = Y + N$, then the conditional distribution of X given simple random variable Y is

$$F_{X|Y} = F_{X|Y^{-1}(-1)}\mathbb{1}_{Y^{-1}(-1)} + F_{X|Y^{-1}(1)}\mathbb{1}_{Y^{-1}(1)},$$

where $F_{X|Y^{-1}(\mu)}$ is $\int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{\sigma^2}}dt$.

Definition 3.10. When X, Y are both continuous random variables, there exists a joint density $f_{X,Y}(x,y)$ for all $(x,y) \in \mathbb{R}^2$. For each $y \in \mathcal{Y}$ such that $f_Y(y) > 0$, we can define a function $f_{X|Y} : \mathbb{R}^2 \to \mathbb{R}_+$ such that

$$f_{X|Y}(x,y) \triangleq \frac{f_{X,Y}(x,y)}{f_Y(y)}, \text{ for all } x \in \mathbb{R}.$$

Exercise 3.11. For continuous random variables X, Y, show that the function $f_{X|Y^{-1}(y)}$ is a density of continuous random variable X for each $y \in \mathbb{R}$.