Lecture-10: Conditional Expectation

1 Conditional Distribution

Consider the 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 event B was defined as

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

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 $A_x = X^{-1}(-\infty, x]$ and $Y = Y^{-1}(-\infty, y]$, 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.

1.1 Conditioning on simple random variables

Consider the probability space $(\Omega,\mathcal{F},P)$ and random variables X, Y on this probability space. If the random variable $Y : \Omega \to \mathcal{Y}$ is simple, then the events $(E_y : y \in \mathcal{Y})$ partition the sample space, where $E_y \triangleq Y^{-1}\{y\}$. Further the simple random variable Y has a PMF $P_Y \triangleq (P(E_y) : y \in \mathcal{Y})$. If $y \in \mathcal{Y}$ such that $P_Y(y) > 0$, then we can define a function $F_{X \mid E_y} : \mathbb{R} \to [0,1]$ such that

$$F_{X \mid E_y}(x) \triangleq P\left(\{X \leqslant x\} \mid E_y\right) = \frac{P\{X \leqslant x, Y = y\}}{P_Y(y)}, \text{ for all } x \in \mathbb{R}.$$

Exercise 1.1. For simple random variable $Y : \Omega \to \mathcal{Y}$, show that the function $F_{X \mid E_y}$ conditioned on the event $E_y = \{Y = y\}$ is a distribution.

Definition 1.2. For a simple random variable $Y \in \mathcal{Y}$, the distribution $F_{X \mid E_y}$ is called the **conditional distribution of X given event E_y**. The **conditional distribution of X given Y** denoted by $F_{X|Y} : \Omega \to [0,1]^{\mathbb{R}}$ is a measurable function of the random variable Y, and hence it is a random variable such that

$$F_{X \mid Y} : \omega \mapsto F_{X \mid Y = Y(\omega)}.$$

We can write the conditional distribution $F_{X|Y} = \sum_{y \in \mathcal{Y}} F_{X|E_y} \mathbb{1}_{E_y}$.

Example 1.3 (Conditional distribution). Consider the zero-mean Gaussian random variable N 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

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

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

1.2 Conditional densities

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 $f_{X \mid Y=y} : \mathbb{R} \to \mathbb{R}_+$ such that

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

Exercise 1.4. For continuous random variables X, Y, show that the function $f_{X \mid Y=y}$ is a density of continuous random variable X for each $y \in \mathbb{R}$.

Definition 1.5. The **conditional density of** X **given** Y for continuous random variables X, Y is defined as a measurable function of the random variable Y, and hence it is random variable such that

$$f_{X \mid Y} : \omega \mapsto f_{X \mid Y=Y(\omega)}.$$

2 Conditional expectation in terms of conditional distribution

Since we have defined the conditional distribution and densities, we can define the conditional expectation given an event as an integration with respect to the conditional distribution given that event. In the following, we will assume the random variables X, Y are defined on the same probability space $(\Omega, \mathcal{F}, P)$ and $\mathbb{E}|X| < \infty$ such that $\mathbb{E}X$ exists and is finite.

2.1 Simple random variables

When Y is a simple random variable, we can define the conditional expectation of X given the event $E_y = \{Y = y\}$ for $P_Y(y) > 0$ as

$$\mathbb{E}[X \mid E_y] \triangleq \int_{x\in\mathbb{R}} x dF_{X|E_y}(x) = \int_{(x,t)\in\mathbb{R}^2} x \frac{d_x F_{X,Y}(x,t)}{P_Y(y)} \mathbb{1}_{\{t=y\}} = \frac{\mathbb{E}[X \mathbb{1}_{E_y}]}{P_Y(y)}.$$

Definition 2.1 (Conditional expectation for conditioning on simple random variables). When Y is a simple random variable, we can define the conditional expectation of X given the random variable Y is a measurable function of random variable Y, denoted by $\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}$ such that

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

Hence, we can write

$$\mathbb{E}[X \mid Y] = \sum_{y\in\mathcal{Y}} \mathbb{E}[X \mid E_y] \mathbb{1}_{E_y} = \sum_{y\in\mathcal{Y}} \frac{\mathbb{E}[X \mathbb{1}_{E_y}]}{P_Y(y)} \mathbb{1}_{E_y}.$$

Remark 1. The random variable $\mathbb{E}[X \mid Y]$ takes value $\mathbb{E}[X \mid E_y]$ with probability $P_Y(y)$ for all $y \in \mathcal{Y}$.

Lemma 2.2. *For simple random variable $Y : \Omega \to \mathcal{Y}$, the mean of random variable $\mathbb{E}[X \mid Y]$ is $\mathbb{E}[X]$.*

Proof. Since $\mathbb{E}[X|Y]$ is a function of the random variable Y, we have

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \sum_{y \in \mathcal{Y}} P_Y(y)\mathbb{E}[X \mid E_y] = \sum_{y \in \mathcal{Y}} \mathbb{E}[X\mathbb{1}_{E_y}] = \mathbb{E}[X].$$

□

Example 2.3 (Conditional expectation). Consider a fair die being thrown and the random variable X takes the value of the outcome of the experiment. That is, $X \in \{1, \cdots, 6\}$ with $P[X = i] = 1/6$ for $i \in \{1, \cdots, 6\}$. Define another random variable $Y = \mathbb{1}_{\{X \le 3\}}$. Then the conditional expectation of X given Y is a random variable given by

$$\mathbb{E}[X|Y] = \begin{cases} \mathbb{E}[X|Y=1] = 2 & w.p \;\; 0.5 \\ \mathbb{E}[X|Y=0] = 5 & w.p \;\; 0.5. \end{cases}$$

It is easy to see that $\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X] = 3.5$.

2.2 Continuous random variables

When X, Y are continuous random variables, we can define the conditional expectation of X given Y as a continuous random variable $\mathbb{E}[X \mid Y] : \Omega \to \mathbb{R}$ such that

$$\mathbb{E}[X \mid Y] : \omega \mapsto \int_{x \in \mathbb{R}} x f_{X|Y=Y(\omega)}(x)dx,$$

with the density $f_Y(y)$ for all $y \in \mathbb{R}$.

Lemma 2.4. *For continuous random variables X, Y, the mean of random variable $\mathbb{E}[X \mid Y]$ is $\mathbb{E}[X]$.*

Proof. Since $\mathbb{E}[X|Y]$ is a function of the random variable Y and its density is $f_Y(y)$, we get

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \int_{y \in \mathbb{R}} dy f_Y(y)\mathbb{E}[X \mid Y = y] = \int_{y \in \mathbb{R}} dy f_Y(y) \int_{x \in \mathbb{R}} x f_{X \mid Y=y} dx.$$

From the definition of conditional density of X given $Y = y$, we get $f_{X \mid Y=y}(x) f_Y(y) = f_{X,Y}(x, y)$. Interchanging integrations from Fubini's theorem, and from the law of total probability, we get

$$\mathbb{E}[\mathbb{E}[X \mid Y]] = \int_{x \in \mathbb{R}} x dx \int_{y \in \mathbb{R}} f_{X,Y}(x, y) dy = \int_{x \in \mathbb{R}} x f_X(x) dx = \mathbb{E}[X].$$

□

3 Conditional expectation conditioned on an event space

Definition 3.1 (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

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

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

(iii) 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 3.2 (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 $Y \triangleq \mathbb{E}[X \mid \mathcal{G}]$, is an almost sure unique 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 $\mathbb{E}[X\mathbb{1}_A] = \mathbb{E}[Y\mathbb{1}_A]$,

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

3.1 Conditional expectation conditioned on a random variable

Remark 2. 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, this implies that $\mathbb{E}[X|Y]$ is a random variable, that takes value $\mathbb{E}[X|E_y]$ when $\omega \in E_y$. When Y is simple, $\mathbb{E}[X|Y]$ is a simple random variable with PMF P_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 avergaging 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 $\mathbb{E}[X|Y] \in L^1$.

Example 3.3 (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 the event space $\sigma(Y) = (\cup_{y\in I} E_y : I \subseteq \mathcal{Y})$. For a random variable $X : \Omega \to \mathbb{R}$ defined on the same probability space, the random variable $Z \triangleq \mathbb{E}[X \mid Y] \in \sigma(Y)$. Therefore,

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

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 $\alpha_y = \frac{\mathbb{E}[X\mathbb{1}_{E_y}]}{P_Y(y)}$ for any $y \in \mathcal{Y}$.