Lecture-09: 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 $A = X^{-1}(-\infty, x]$ and $B = Y^{-1}(-\infty, y]$, 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 it has a PMF $((P_Y(y) = P(Y^{-1}\{y\}) : y \in \mathcal{Y})$. If $y \in \mathcal{Y}$ such that $P_Y(y) > 0$, then we can define a function $F_{X \mid \{Y=y\}} : \mathbb{R} \to [0,1]$ such that

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

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

Definition 1.2. For a simple random variable $Y \in \mathcal{Y}$, the distribution $F_{X \mid Y=y}$ is called the **conditional distribution of** X **given** $Y = y$. 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 $F_{X|Y} = \sum_{y \in \mathcal{Y}} F_{X|Y=y} \mathbb{1}_{\{Y=y\}}$.

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

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 $\{Y = y\}$ for $P_Y(y) > 0$ as

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

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

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

Hence, we can write

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

Remark 1. The random variable $\mathbb{E}[X \mid Y]$ takes value $\mathbb{E}[X \mid Y = y]$ 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

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

□

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 $Y = \mathbb{1}_{\{X \leq 3\}}$. 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

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

□

2.3 Event space generated by random variables

Definition 2.5 (Event space generated by a random variable). Let $X : \Omega \to \mathbb{R}$ be a random variable on the probability space $(\Omega, \mathcal{F}, P)$, then the smallest event space generated by the events of the form $X^{-1}(-\infty, x]$ for $x \in \mathbb{R}$ is called the **event space generated** by this random variable X, and denoted by $\sigma(X)$.

Example 2.6 (Indicator function). Let $A \in \mathcal{F}$ be an event, then $X = \mathbb{1}_A$ is a random variable and

$$X^{-1}(-\infty, x] = \begin{cases} \Omega, & x \geqslant 1, \\ A^c, & x \in [0, 1), \\ \emptyset, & x < 0. \end{cases}$$

This implies that the smallest event space generated by this random variable is $\sigma(X) = \{\emptyset, A, A^c, \Omega\}$.

Example 2.7 (Simple random variables). Let X be a simple random variable, then $X = \sum_{x \in \mathcal{X}} x \mathbb{1}_{A_x}$ where $(A_x = X^{-1}\{x\} \in \mathcal{F} : x \in \mathcal{X})$ is a finite partition of the sample space Ω. Without loss of generality, we can denote $\mathcal{X} = \{x_1, \dots, x_n\}$ where $x_1 \leqslant \dots \leqslant x_n$. Then,

$$X^{-1}(-\infty, x] = \begin{cases} \Omega, & x \geqslant x_n, \\ \cup_{j=1}^{i} A_{x_i}, & x \in [x_i, x_{i+1}), i \in [n-1], \\ \emptyset, & x < x_1. \end{cases}$$

Then the smallest event space generated by the simple random variable X is $\{\cup_{x \in S} A_x : S \subseteq \mathcal{X}\}$.