STAT0041: Stochastic Calculus

Lecture 2 - Conditional expectation

Lecturer: Weichen Zhao | Fall 2024

Key concepts:

- *Conditional expectation;*

2.1 Basic Definition

Estimate is an important topic in probability and statistic. We consider a random variable ξ on $(\Omega, \mathscr{F}, \mathrm{P})$ and sub event field $\mathscr{G}$ of $\mathscr{F}$. If ξ is $\mathscr{G}$-measurable, then the information in $\mathscr{G}$ is sufficient to determine the value of ξ. If ξ is independent of $\mathscr{G}$, then the information in $\mathscr{G}$ provides no help in determining the value of ξ. In the intermediate case, we can use the information in $\mathscr{G}$ to estimate but not precisely evaluate ξ. The conditional expectation of ξ given $\mathscr{G}$ is such an estimate.

First we give the basic definition in this lecture.

Definition 2.1 (Conditional expectation) *Let $(\Omega, \mathscr{F}, \mathrm{P})$ be a probability space, $\mathscr{G}$ be a sub event field of $\mathscr{F}$, X be a integrable random variable ($\mathbb{E}[|X|] < \infty$). The **conditional expectation** of X given $\mathscr{G}$, denoted $\mathbb{E}[X|\mathscr{G}]$, is any random variable that satisfies*

(1) $\mathbb{E}[X|\mathscr{G}]$ *is $\mathscr{G}$-measumble, and;*

(2)

$$\int_A \mathbb{E}[X|\mathscr{G}](\omega)\,\mathrm{dP}(\omega) = \int_A X(\omega)\,\mathrm{dP}(\omega), \quad \text{for all } A \in \mathscr{G}. \tag{2.1}$$

The second property ensures that $\mathbb{E}[X|\mathscr{G}]$ is indeed an estimate of X. It gives the same averages as X over all the sets in $\mathscr{G}$.

Connection to elementary probability. Considering a simple case, let X and Y be two random variables on $(\Omega, \mathscr{F}, \mathrm{P})$ taken values in $\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots, y_n\}$, $m, n \in \mathbb{N}^+$, respectively. In elementary probability, conditional probability is defined as

$$\mathrm{P}(X = x_i | Y = y_j) := \frac{\mathrm{P}(X = x_i; Y = y_j)}{\mathrm{P}(Y = y_j)}.$$

Conditional expectation is defined as

$$\mathbb{E}[X|Y = y_j] := \sum_{i=1}^{m} x_i \mathrm{P}(X = x_i; Y = y_j).$$

Using axiomatic language introduced in lecture 1,

$$\mathbb{E}[X|Y = y_j] = \int_\Omega X(\omega)\mathrm{dP}(\omega|Y = y_j) = \int_{\mathbb{R}} x\mathrm{dP}_X(x|Y = y_j),$$

where P_X is distribution of X. We define random variable

$$\mathbb{E}[X|Y](\omega) = \sum \mathbb{E}[X|Y = y_j]\mathbf{1}_{\{Y=y_j\}}(\omega)$$

as conditional expectation of X given Y.

Let $\mathscr{G} = \sigma(Y)$ is the event field generated by Y, we have

$$\sigma(Y) = \{\{\omega : Y(\omega) \in B\} : B \in \mathscr{B}(\mathbb{R})\} = Y^{-1}(\mathscr{B}(\mathbb{R})).$$

In discrete case, $\sigma(Y)$ is given by $\{y_1, y_2, \ldots, y_n\}$ of 2^n of all possible concatenation sets. Thus $\mathbb{E}[X|Y](\omega) = \sum \mathbb{E}[X|Y = y_j]\mathbf{1}_{\{Y=y_j\}}(\omega)$ is $\sigma(Y)$ measurable, which satisfies property (1) in definition 2.1.

Moreover, we have

$$\begin{aligned}\int_{\{Y=y_j\}} \mathbb{E}[X|Y]d\mathrm{P} &= \mathbb{E}[X|Y = y_j]\mathrm{P}(Y = y_j) = \sum_i x_i \mathrm{P}(X = x_i|Y = y_j)\mathrm{P}(Y = y_j) \\ &= \sum_i x_i \mathrm{P}(X = x_i; Y = y_j) = \int_{\{Y=y_j\}} X d\mathrm{P}.\end{aligned}$$

Denote $G_j = \{Y = y_j\}$, we have

$$\mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_{G_j}] = \mathbb{E}[X\mathbf{1}_{G_j}].$$

Since for all $G \in \sigma(Y)$, there exist finite $j_1, \ldots, j_k,\ k \leq n$, s.t. $G = G_{j_1} \cup \cdots \cup G_{j_k}$, that is $\mathbf{1}_G = \sum_{j_i} \mathbf{1}_{j_i}$. Then

$$\begin{aligned}\mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_G] &= \mathbb{E}[\mathbb{E}[X|Y]\sum_i \mathbf{1}_{G_{j_i}}] = \sum_i \mathbb{E}[\mathbb{E}[X|Y]\mathbf{1}_{G_{j_i}}] \\ &= \sum_i \mathbb{E}[X\mathbf{1}_{G_{j_i}}] = \mathbb{E}[X\sum_i \mathbf{1}_{G_{j_i}}] \\ &= \mathbb{E}[X\mathbf{1}_G].\end{aligned}$$

Thus

$$\int_G \mathbb{E}[X|Y]\mathrm{dP} = \int_G X\mathrm{dP} \quad \forall G \in \sigma(Y),$$

which implies property (2) in definition 2.1.

Mean squared error. Given two random variables X, Y, a key problem is predicting the value of X from observation values of Y. (Such as estimating one's height from foot length). That is finding function f, such that $f(Y)$ is closed to X. We usually consider using *mean squared error*:

$$\mathbb{E}[(X(\omega) - f(Y(\omega)))^2]$$

to measure the distance between X and $f(Y)$.

Claim 2.2 *Conditional expectation $\mathbb{E}[X|Y]$ is the estimate of X which minimizes the mean squared error, that is*

$$\mathbb{E}[(X - \mathbb{E}[X|Y])^2] = \inf_f \mathbb{E}[(X - f(Y))^2]$$

2.2 Geometric intuition

Random variable space with finite second-order moment. We often use two statistical characteristics, expectation and variance, to describe random phenomena. When a random variable has finite second-order moments, its expectation and variance must exist. Therefore, we will learn what kind of mathematical structure such a class of random variables has.

Denote all random variables on probability space $(\Omega, \mathscr{F}, \mathrm{P})$ with finite second-order moment as $L^2(\Omega, \mathscr{F}, \mathrm{P})$, satisfies:

(1) *linear space*: For all $\xi, \eta \in L^2(\Omega, \mathscr{F}, \mathrm{P})$, $a, b \in \mathbb{R}$,

$$\begin{aligned}\mathbb{E}(a\xi + b\eta)^2 &\leq a^2\mathbb{E}\xi^2 + b^2\mathbb{E}\eta^2 + 2|ab|\mathbb{E}(\xi\eta)\\ &\leq a^2\mathbb{E}\xi^2 + b^2\mathbb{E}\eta^2 + 2|ab|\sqrt{\mathbb{E}\xi^2\mathbb{E}\eta^2}\\ &< \infty \in L^2(\Omega, \mathscr{F}, \mathrm{P}).\end{aligned}$$

(2) *Inner product structure*: For all $\xi, \eta \in L^2(\Omega, \mathscr{F}, \mathrm{P})$, we define inner product as:

$$\langle \xi, \eta \rangle = \mathrm{E}(\xi\eta) \leq \sqrt{\mathrm{E}\xi^2\mathrm{E}\eta^2} < \infty.$$

Further we have Euclidean distance:

$$\|\xi - \eta\|_{L^2} := \sqrt{\langle \xi - \eta, \xi - \eta \rangle} = \sqrt{\mathrm{E}(\xi - \eta)^2},$$

which is exactly mean squared error of ξ and η.

Geometric intuition of conditional expectation. Let $\mathscr{G}$ be a sub event field of $\mathscr{F}$, $X \in L^2(\Omega, \mathscr{F}, \mathrm{P})$. It can be proved that $L^2(\Omega, \mathscr{G}, \mathrm{P})$ is a closed subspace of $L^2(\Omega, \mathscr{F}, \mathrm{P})$ (reflection question).

Let X be a random variable in $L^2(\Omega, \mathscr{F}, \mathrm{P})$, $\mathbb{E}[X|\mathscr{G}]$ is orthogonal projection of X to the space $L^2(\Omega, \mathscr{G}, \mathrm{P})$. That is, for all random variable $Y \in L^2(\Omega, \mathscr{G}, \mathrm{P})$, we have

$$\mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}]) \cdot Y] = 0. \tag{2.2}$$

In fact, consider $Y = \mathbf{1}_B$, $B \in \mathscr{G}$, for every $A \in \mathscr{G}$

$$\begin{aligned}\int_A \mathbb{E}[X|\mathscr{G}](\omega)Y(\omega)\mathrm{dP}(\omega) &= \int_{A\cap B} \mathbb{E}[X|\mathscr{G}](\omega)\mathrm{dP}(\omega)\\ &= \int_{A\cap B} X(\omega)\mathrm{dP}(\omega)\\ &= \int_A X(\omega)Y(\omega)\mathrm{dP}(\omega).\end{aligned}$$

Then follow the standard method in measure theory (Indicator function - simple function - non-negative measurable function - measurable function), Eq.(2.2) holds.

For every $Y \in L^2(\Omega, \mathscr{G}, \mathrm{P})$,

$$\begin{aligned}\|X - Y\|_{L^2}^2 &= \langle X - Y, X - Y \rangle\\ &= \langle X - \mathbb{E}[X|\mathscr{G}] + (\mathbb{E}[X|\mathscr{G}] - Y), X - \mathbb{E}[X|\mathscr{G}] + (\mathbb{E}[X|\mathscr{G}] - Y)\rangle\\ &= \|X - \mathbb{E}[X|\mathscr{G}]\|_{L^2}^2 + \|\mathbb{E}[X|\mathscr{G}] - Y\|_{L^2}^2\\ &\geq \|X - \mathbb{E}[X|\mathscr{G}]\|_{L^2}^2.\end{aligned}$$

That is

$$\mathbb{E}[(X - \mathbb{E}[X|\mathscr{G}])^2] = \inf_{Y\in L^2(\Omega,\mathscr{G},\mathrm{P})} \mathbb{E}[(X - Y)^2]$$

Remark 2.3 *For $X \in L^2(\Omega, \mathscr{G}, \mathrm{P})$, Hilbert projection theorem implies existence and uniqueness of $\mathbb{E}[X|\mathscr{G}]$.*

2.3 Properties of conditional expectation

Proposition 2.4 (Basic properties) *Let X and Y*

(1) For $a, b \in \mathbb{R}$, $\mathbb{E}[aX + bY|\mathscr{G}] = a\mathbb{E}[X|\mathscr{G}] + b\mathbb{E}[Y|\mathscr{G}]$;

(2) If $X \geq Y$, then $\mathbb{E}[X|\mathscr{G}] \geq \mathbb{E}[Y|\mathscr{G}]$;
If $X \geq 0$, then $\mathbb{E}[X|\mathscr{G}] \geq 0$;
$\mathbb{E}[|X||\mathscr{G}] \geq |\mathbb{E}[X|\mathscr{G}]|$;

(3) ξ is $\mathscr{G}$ measurable $\Longrightarrow E[\xi|\mathscr{G}] = \xi$.

(4) For every X is $\mathscr{G}$ measurable, expectation of X and XY are exist, then $\mathbb{E}[XY|\mathscr{G}] = X\mathbb{E}[Y|\mathscr{G}]$;

(5) $\mathbb{E}[\mathbb{E}[X|\mathscr{G}]] = \mathbb{E}[X]$;

(6) Let $\mathscr{G}_1 \subset \mathscr{G}_2 \subset \mathscr{F}$, then $\mathbb{E}[\mathbb{E}[X|\mathscr{G}_1]|\mathscr{G}_2] = \mathbb{E}[\mathbb{E}[X|\mathscr{G}_2]|\mathscr{G}_1] = \mathbb{E}[X|\mathscr{G}_1]$.