CHAPTER 9:

CONDITIONAL EXPECTATION I: DISCRETE CASE.

Goal: • contruction of $E(X|Y)$ when Y discrete

• intuition / motivation for abstract definition.

Setup. • $(\Omega, \mathcal{F}, P)$ fixed probability space.

• real random variables: $X : \Omega \to \mathbb{R}$ measurable.

I GENERAL IDEA.

• X real r.v. : output of a random experiment

Expected value:

Before the experiment: $E(X)$ "no information about the experiment"

? During experiment: $E\left(X \,\middle|\, \begin{smallmatrix}\text{partial}\\ \text{infor}\end{smallmatrix}\right)$ "partial information is known about X"

After I finished the experiment: X "all information is revealed"

2 WHAT ARE WE SEEKING?

$\underbrace{\text{Let } X \text{ be a r.v. in } \mathbb{R}}_{\text{The final output of our experiment}}$, $\underbrace{Y \text{ another r.v.}}_{\text{Some partial information about the experiment.}}$

We would like to define:

$E(X|Y) \to$ expected value of X given Y.

Example: Y_1, Y_2, Y_3 iid $Ber(\frac{1}{2})$ "3 coin flips."

$$\boxed{X = Y_2 + Y_3}$$

- Y_1 does not bring any information about X.

  We want $E(X|Y_1) = E(X)$.

- Y_2 brings partial information about X.

  $\to$ if $Y_2 = 0$ then $X = \begin{cases} 0 & \text{with proba } \frac{1}{2} \\ 1 & \text{with proba } \frac{1}{2} \end{cases}$

  $\hookrightarrow E(X|Y_2) = \frac{1}{2}$

  $\to$ if $Y_2 = 1$ then $X = \begin{cases} 1 & \text{with proba } \frac{1}{2} \\ 2 & \text{with proba } \frac{1}{2} \end{cases}$

  $\hookrightarrow E(X|Y_2) = \frac{3}{2}$

  We want $E(X|Y_2) = Y_2 + \frac{1}{2}$.

- $Y_2 + Y_3$ gives all the information about X:

We want $E(X \mid Y_1 + Y_2) = X$.

3 CONDITIONAL EXPECTATION GIVEN A > 0 PROBABILITY EVENT

Goal: def. $E(X \mid A)$

Fix $A \in \mathcal{F}$ s.t. $P(A) > 0$.

Def: Let X be a real r.v. $X \geq 0$ or $X \in L^1$. Define

$$\boxed{E(X \mid A) := \frac{1}{P(A)} E(X \mathbb{1}_A)}.$$

Reminder: $P_A : \mathcal{F} \longrightarrow [0,1]$ is a proba measure on $(\Omega, \mathcal{F})$.
$\qquad B \longmapsto P(B \mid A)$

Rk: $E(X \mathbb{1}_A)$ well def.

Prop. For every X real r.v. $X \geq 0$ or $X \in L^1$, we have

$$E(X \mid A) = \int_\Omega X \, dP_A$$

"$E(X \mid A)$ is the expectation with respect to $P_A = P(\cdot \mid A)$"

Proof: ① $X = 1_B$, $B \in \mathcal{F}$.

$$E(X \mid A) = E(1_B \mid A)$$
$$= \frac{1}{P(A)} E(1_B 1_A)$$
$$= \frac{P(A \cap B)}{P(A)}$$
$$= P_A(B) = \int_\Omega 1_B \, dP_A .$$

② X simple : $X = \sum_{i=1}^n \lambda_i 1_{B_i}$ $\lambda_i \in \mathbb{R}$ $B_i \in \mathcal{F}$.

$$E(X \mid A) = \sum_{i=1}^n \lambda_i E(1_{B_i} \mid A)$$
$$\overset{①}{=} \sum_{i=1}^n \lambda_i \int_\Omega 1_{B_i} \, dP_A$$
$$= \int_\Omega X \, dP_A \qquad \text{(linearity of } \int_\Omega \cdot \, dP_A\text{)}$$

③ X measurable ≥ 0. Let $X_n \uparrow X$ where X_n simple.

$\forall n$ we have $\dfrac{E(X_n 1_A)}{P(A)} = \int X_n \, dP_A$

By monotone convergence we get $\dfrac{E(X 1_A)}{P(A)} = \int X \, dP_A$

④ $X \in L^1$: we $X = X_+ - X_-$ where $X_+ , X_- \geq 0$.

Example !

$$\Omega = \{1, 2, 3, 4, 5, 6\} \qquad P = \frac{1 \cdot 1}{6}.$$

$$X(\omega) = \omega$$

$$E(X) = \sum_{k=1}^{6} k \, P(X=k) = \frac{1 + \cdots + 6}{6} = 3.5$$

$$A = \{2, 4, 6\} \qquad \text{“die is even”}$$

$$E(X \mid A) = \frac{1}{P(A)} E(X 1_A)$$

$$= \frac{1}{P(A)} E(X 1_{X \in \{2,4,6\}})$$

$$= 2 \sum_{k=1}^{6} k \, 1_{k \in \{2,4,6\}} \, P(X=k)$$

$$= 2 \times \frac{1}{6} (2+4+6) = 4.$$

Exercise: $A = \{4, 5, 6\}$ $\qquad E(X \mid A) = 5$

$A = \{3\}$ $\qquad E(X \mid A) = 3$

$A = \{1, 2, 3, 4, 5, 6\}$ $\qquad E(X \mid A) = 3.5.$

Example 2 Y_1, Y_2, Y_3 iid $Ber(\frac{1}{2})$

$$X = Y_2 + Y_3.$$

- $A = \{Y_1 = y\} \qquad y \in \{0,1\}.$

$$E(X \mid Y_1 = y) = \frac{1}{P(Y_1 = y)} E\left((Y_2 + Y_3) 1_{Y_1 = y}\right)$$

$$\overset{\text{indep.}}{=} \frac{1}{\cancel{P(Y_1 = y)}} E(Y_2 + Y_3)\, \cancel{P(Y_1 = y)}$$

$$= E(X) = \underline{1}.$$

- $A = \{Y_2 = y\} \qquad y \in \{0,1\}$

$$E(X \mid Y_2 = y) = \frac{1}{P(Y_2 = y)} E\left((Y_2 + Y_3) 1_{Y_2 = y}\right)$$

$$= \frac{1}{P(Y_2 = y)} E\left(y\, 1_{Y_2 = y}\right) + \frac{1}{P(Y_2 = y)} E\left(Y_3\, 1_{Y_2 = y}\right)$$

$$\overset{\text{indep.}}{=} y + E(Y_3) = \underline{y + \frac{1}{2}}.$$

- $A = \{Y_2 + Y_3 = y\} \qquad y \in \{0,1,2\}$

$$E(X \mid Y_2 + Y_3 = y) = \frac{1}{P(Y_2 + Y_3 = y)} E\left((Y_2 + Y_3) 1_{Y_2 + Y_3 = y}\right)$$

$$= \underline{y}$$

4. CONDITIONAL EXPECTATION WRT TO A DISCRETE R.V.

Let D be finite or countable set (fixed until end of chapter). (equipped with $\mathcal{P}(D)$)

Def. Let Y be a r.v. with values in D.

X be a real r.v. $X \in L^1$ or $X \geq 0$ a.s.

For $y \in D$, set

$$\boxed{\varphi(y) := \begin{cases} E(X \mid Y = y) & \text{if } P(Y=y) > 0 \\ 0 & \text{if } P(Y=y) = 0 \end{cases}}$$

and define the r.v.

$$\boxed{E(X \mid Y) := \varphi(Y)}$$

Example: Y_1, Y_2, Y_3 iid Ber$(\frac{1}{2})$ $X = Y_2 + Y_3$.

- $E(X \mid Y_1) = E(X)$ a.s
- $E(X \mid Y_2) = Y_2 + \frac{1}{2}$ a.s.
- $E(X \mid Y_2 + Y_3) = X$. a.s.

In $E(X \mid Y)$, Y is treated as a fixed element not any more as a random element..

Properties:

Let X, Y as above.

- if X indep. of Y then $E(X|Y) = E(X)$ a.s.
- if X is meas. wrt to $\sigma(Y)$, $E(X|Y) = X$ a.s.

Proof. Let $y \in D$ s.t $P(Y=y) > 0$

- If X indep. of Y,

$$E(X|Y=y) = \frac{1}{P(Y=y)} E(X 1_{Y=y}) = E(X).$$

- If $X = \Psi(Y)$ a.s. where $\Psi : D \to \mathbb{R}$ meas.

$$E(X|Y=y) = E(\Psi(y) | Y=y) = \Psi(y).$$ ∎

5 CHARACTERISTIC PROPERTY.

Motivation: How to define $E(X|Y)$ if Y not discrete?
How to define $E(X|Y_1, Y_2, \dots)$?

Prop. Let X, Y as above..

$$X' = E(X|Y) \text{ a.s.} \iff \begin{cases} X' \text{ is } \sigma(Y)\text{-measurable} \\ \forall A \in \sigma(Y) \; E(X 1_A) = E(X' 1_A). \end{cases}$$

Proof. $\boxed{\Rightarrow}$ $E(X|Y) = \varphi(Y)$ hence it is $\sigma(Y)$-measurable.

Let $A \in \sigma(Y)$, let $\psi : D \to \mathbb{R}$ s.t. $1_A = \psi(Y)$ a.s.

$$E\left(E(X|Y)\, 1_A \right) = E\left(\varphi(Y)\, \psi(Y)\right)$$

$$= \sum_{y \in D'} E\left(\varphi(y)\, \psi(y)\, 1_{Y=y}\right)$$

$(D' = \{y : P(Y=y) > 0\})$

$$= \sum_{y \in D'} \psi(y) \underbrace{\varphi(y)\, P(Y=y)}_{E(X 1_{Y=y})}$$

$$= \sum_{y \in D'} E\left(X\, \psi(y)\, 1_{Y=y}\right).$$

$$= E\left(X\, \psi(Y)\right)$$

$$= E\left(X\, 1_A\right)$$

$\boxed{\Leftarrow}$ Since $\{Y=y\} \in \sigma(Y)$, we have for $y \in D'$

$$\begin{cases} E(X' 1_{Y=y}) \overset{\text{Hyp}}{=} E(X 1_{Y=y}) \\ E\left(E(X|Y)\, 1_{Y=y}\right) \underset{\text{by } \boxed{\Rightarrow}}{=} E\left(X 1_{Y=y}\right) \end{cases}$$

Hence, for all $y \in D'$, writing $X' = \psi(Y)$

$$0 = E\left(\left(X' - E(X|Y)\right) 1_{Y=y} \right)$$

$$= E\left(\psi(y) - \varphi(y)\; 1_{Y=y} \right)$$

$$= \left(\psi(y) - \varphi(y)\right) P(Y=y) .$$

ccl $\forall y \in D'$ $\psi(y) = \varphi(y)$

Since $P(Y \in D') = \sum_{y \in D'} P(Y = y) = \sum_{y \in D} P(Y=y) = 1$,

we deduce
$$\underbrace{\psi(Y)}_{X'} = \underbrace{\varphi(Y)}_{E(X|Y)} \text{ a.s.}$$

6 CONCLUDING REMARKS.

The def. of $E(X|Y)$ suggests that it depends strongly on the r.v. Y. How much?

→ Ex: Y_1, Y_2, Y_3 iid $Ber(1/2)$ $\quad X = Y_2 + Y_3$.

Take $Y' = \begin{cases} 1 & \text{if } Y_2 = 0 \\ e & \text{if } Y_2 = 1 \end{cases}$ $\quad (Y' = e^{Y_2},\ Y_2 = \log Y')$

$$\varphi(y) = E(X \mid Y' = y) = \begin{cases} \frac{1}{2} & \text{if } Y' = 1 \\ \frac{3}{2} & \text{if } Y' = e \end{cases}$$

$$E(X|Y') = \frac{1}{2} + \log(Y') = \frac{1}{2} + Y_2 \quad \text{a.s.}$$

Hence $E(X|Y') = E(X|Y_2)$ a.s.

Rk: Let Y, Y' be two r.v.s on D.

$X \geq 0$ or $X \in L^1$.

$$\boxed{\sigma(Y) = \sigma(Y') \implies E(X \mid Y) = E(X \mid Y') \text{ a.s.}}$$

(indeed, by the charact. property.

$$X' = E(X \mid Y) \iff \begin{cases} X' \ \underbrace{\sigma(Y)}_{=\sigma(Y')}\text{-meas.} \\ \forall A \in \underbrace{\sigma(Y)}_{=\sigma(Y')} \ E(X' 1_A) = E(X 1_A) \end{cases}$$

$$\iff X' = E(X \mid Y) \quad)$$

↳ The "good" notion of conditional expectation is with respect to a σ-algebra.

<u>Idea:</u> ($\mathcal{G} \subset \mathcal{F}$ σ-algebra) Define $E(X \mid \mathcal{G})$ as
(abstract def)

a r.v. X' satisfying $\begin{cases} X' \ \mathcal{G}\text{-measurable.} \\ \forall A \in \mathcal{G} \quad E(X' 1_A) = E(X 1_A). \end{cases}$

↳ In general existence and uniqueness need to be justified.

advantage: generality → $\mathcal{G} = \sigma(Y)$ for Y non discrete.

$\mathcal{G} = \sigma(Y_1, Y_2, \ldots)$. are possible.

difficulty: non constructive approach: $E(X \mid \mathcal{G})$ may be hard to compute / manipulate.