Lecture 7 / Week 5

OUTLINE

1) Further Properties of Conditional Expectation *Book* §3.2

2) Conditional Probability and its distribution. *Book* 3.1, 3.3

Conditional Expectation

Recall the definition, i.e suppose that Y is an *integrable* random variable defined on the probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{F}_0$ be a sub σ-algebra of $\mathcal{F}$. Then, the *conditional expectation* of Y given $\mathcal{F}_0$ is a random variable Z on same probability space $(\Omega, \mathcal{F}, P)$ such that

1. Z is $\mathcal{F}_0-$ measurable.
2. For every event $A \in \mathcal{F}_0 \to \int_A Z dP = \int_A Y dP$

We have also seen the following properties if Y is $\mathcal{F}_0$-measurable, then;

$$\begin{aligned} E(Y|\mathcal{F}_0) &= Y \ \ \textbf{a.s} \\ E(Y.Z|\mathcal{F}_0) &= Y.E(Z|\mathcal{F}_0) \ \textbf{a.s} \end{aligned}$$

Now suppose we have $\mathcal{F}_0, \mathcal{F}_1$ and the random variable Y. The question is whether the following equality holds in general;

$$E(Y|\mathcal{F}_0) \ is \text{ r.v} \to E(E(Y|\mathcal{F}_0)|\mathcal{F}_1) =^? E(E(Y|\mathcal{F}_1)|\mathcal{F}_0)$$

So in other words, does the order matter? The answer is in general yes, the order matter, i.e they are not equal.

Theorem Let $\mathcal{F}_0 \subseteq \mathcal{F}_1$, then

$$E(E(Y|\mathcal{F}_0)|\mathcal{F}_1) = E(E(Y|\mathcal{F}_1)|\mathcal{F}_0) = E(Y|\mathcal{F}_0)$$

Proof First we will proof the equality between first and third term; Given that $E(Y|\mathcal{F}_0)$ is $\mathcal{F}_0$-measurable

$$E(Y|\mathcal{F}_0)^{-1}(B) \in \mathcal{F}_0 \subseteq \mathcal{F}_1 \Rightarrow E(Y|\mathcal{F}_0) \text{ is also } \mathcal{F}_1\text{-measurable.}$$

But then

$$E(E(Y|\mathcal{F}_0)|\mathcal{F}_1) = E(Y|\mathcal{F}_0) \ \ a.s$$

It is left to prove that

$$E(E(Y|\mathcal{F}_1)|\mathcal{F}_0) = E(Y|\mathcal{F}_0).$$

Call $Z = E(Y|\mathcal{F}_0)$ and we'll check whether it satisfies the definition of the conditional expectation of $E(Y|\mathcal{F}_1)$ given $\mathcal{F}_0$.

Z is $\mathcal{F}_0$-measurable by definition.

Take an event $A \in \mathcal{F}_0$, so it is also true that $A \in \mathcal{F}_0 \subseteq \mathcal{F}_1$, but we want to show that the following holds

$$\int_A Z dP = \int_A E(Y|\mathcal{F}_1) dP$$

since the event belongs to both σ-algebras and following the definition of conditional expectation, the RHS above equals

$$\int_A E(Y|\mathcal{F}_0) dP = \int_A Y dP$$

On the other hand from the definition of conditional expectation Z, we know that it equals to the LHS below. Hence it follows that

$$\int_A Y dP = \int_A Y dP. \quad \text{QED.}$$

Theorem Suppose we have $E(Y|\mathcal{F}_0)$, Y and $\mathcal{F}_0$ are independent. This is same as saying $\sigma(Y)$ and $\mathcal{F}_0$ are independent. Then

$$E(Y|\mathcal{F}_0) = Y \ \ \textbf{a.s}$$

Proof $E(Y)$ is $\mathcal{F}_0$- measurable. (Because it behaves like a constant→ constant random variable)

$$c^{-1}(B) = \{\omega \in \Omega : c \in B\} = \begin{Bmatrix} \emptyset & c \notin B \\ \Omega & c \in B \end{Bmatrix}$$

since every σ-algebra contains $(\Omega, \emptyset)$. So, in general every constant variable is measurable w.r.t. every σ-algebra.

Definition $\sigma(Y)$ and $\mathcal{F}_0$ are independent. (or Y and $\mathcal{F}_0$ are independent.). For every $A \in \sigma(Y)$ and $B \in \mathcal{F}_0 \Rightarrow A, B$ are independent

$$P(A \cap B) = P(A).P(B)$$

So knowing that Y and $\mathcal{F}_0$ are independent,

$$\begin{aligned} A &\in \mathcal{F}_0 \\ \int_A E(Y)dP &= \int_A YdP = E(Y).P(A) \\ \int_A YdP &= \int \mathbf{1}_A.YdP = E(\mathbf{1}_A.Y) = E(\mathbf{1}_A).E(Y) = E(Y).P(A) \quad \text{(RHS)} \end{aligned}$$

Exercise We exploited the fact that $\mathbf{1}_A$ and Y are independent. Show that it's true.

Proof We know that $\sigma(Y)$ and $\mathcal{F}_0$ are independent. Take an event $A \in \mathcal{F}_0$. Notice that

$$\{\omega \in \Omega : \mathbf{1}_A(\omega) \in B_1\} = \left\{ \begin{array}{ccc} A & 1 \in B_1 & 0 \notin B_1 \\ A^c & 1 \notin B_1 & 0 \in B_1 \\ \emptyset & 1 \notin B_1 & 0 \notin B_1 \\ \Omega & 1 \in B_1 & 0 \in B_1 \end{array} \right\}$$

Insert here Figure 1

But then

$$B_1, B_2 \quad P(\mathbf{1}_A(\omega) \in B_1, Y \in B_2)$$

since the event $(\mathbf{1}_A(\omega) \in B_1)$ belongs $\mathcal{F}_0$, which follows from the fact that $A \in \mathcal{F}_0$ and $\mathcal{F}_0$ is a σ-algebra, then it also follows

$$P(\mathbf{1}_A(\omega) \in B_1, Y \in B_2) = P(\mathbf{1}_A(\omega) \in B_1).P(Y \in B_2)$$

since $\sigma(Y)$ and $\mathcal{F}_0$ are independent and this completed the proof that $\mathbf{1}_A$ and Y are independent.

Example Toss a coin 10 times. Let $Y =$ the number heads and $X =$ the number heads after 8 trials. We need to formalize $E(Y|\sigma(X))$. Call

$$\begin{aligned} Z &= Y - X := \text{ the number heads in the last 2 trials} \\ Y &= X + Z \end{aligned}$$

Note that X and Z are independent, while X and Y are not. Then

$$\begin{aligned} E(Y|\sigma(X)) &= E(X+Z|\sigma(X)) =^{lin \text{ of E.}} E(X|\sigma(X)) + E(Z|\sigma(X)) = \\ &= {}^{prev.theorem} X + E(Z) = X + 1 \end{aligned}$$

So in fact the conditional expectation $E(Y|X)$ is a function in X

$$\begin{aligned} E(Y|X) &= X + 1 = g(X) \\ where \;\; g(x) &= x + 1 \end{aligned}$$

Note also that

$$E(Y|\sigma(X)) = E(Y|X)$$

Theorem Let Y be a random variable and X be a random vector. If Y is measurable w.r.t $\sigma(X)$, then there exists a function g $s.t$

$$Y = g(X)$$

Proof Before proving the theorem, we will show that the converse of the theorem holds, i.e.

$$Y = g(X) \Rightarrow Y \text{ is measurable w.r.t } \sigma(X)$$

We take the inverse of the r.v

$$\begin{aligned} Y^{-1}(B) &= \{\omega \in \Omega : Y(\omega) \in B\} = \{\omega \in \Omega : g(X(\omega)) \in B\} = \\ &= \{\omega \in \Omega : X(\omega) \in g^{-1}(B)\} \in \sigma(X) = \\ &= \{\omega \in \Omega : X^{-1}(g^{-1}(B))\} \in \sigma(X) \\ &\quad i.e \text{ Y is measurable w.rt. } \sigma(X). \end{aligned}$$

On the other hand, the theorem says that

$$E(Y|\sigma(X)) = E(Y|X) \text{ is a function of } X.$$

Suppose $\mathcal{F}_0$ is σ-algebra. We have the triple $(\Omega, \mathcal{F}, P)$ and $\mathcal{F}_0 \subseteq \mathcal{F}$. Also suppose that the event $B \in \mathcal{F}$. We use the following definition

Definition

$$\begin{aligned} P(B|\mathcal{F}_0) &= E(\mathbf{1}_B|\mathcal{F}_0) \\ E(\mathbf{1}_B) &= P(B) \end{aligned}$$

We observe that

1. $P(B|\mathcal{F}_0)$ is $\mathcal{F}_0$-measurable random variable. (by definition of conditional expectation.)
2. $A \in \mathcal{F}_0$, $\int_A P(B|\mathcal{F}_0) dP =^{def.E(\mathbf{1}_B|\mathcal{F}_0)} \int_A \mathbf{1}_B dP = \int \mathbf{1}_A \mathbf{1}_B dP = \int \mathbf{1}_{A\cap B} dP = P(A \cap B)$. Hence

$$\begin{aligned} \forall A &\in \mathcal{F}_0, \\ \int_A P(B|\mathcal{F}_0) dP &= P(A \cap B) \end{aligned}$$

Since the conditional expectation is monotone we have

$$\begin{aligned} 0 &\leq \mathbf{1}_B \leq 1 \\ E(0|\mathcal{F}_0) &\leq E(\mathbf{1}_B|\mathcal{F}_0) \leq E(1|\mathcal{F}_0) \quad \text{a.s} \\ 0 &\leq P(B|\mathcal{F}_0) \leq 1 \\ \sin ce \ \ P(B|\mathcal{F}_0) &= E(\mathbf{1}_B|\mathcal{F}_0) \text{ and } E(0|\mathcal{F}_0) = 0,\ E(1|\mathcal{F}_0) = 1 \end{aligned}$$

Definition We can generalize the previous result. Let both X and Y be random vectors. Then

$$\begin{aligned} P(Y &\in B|X) = g(X) \ \ B \in \mathcal{B}(\mathbb{R}^K) \\ P(Y &\in B|X) = P(\{\omega \in \Omega : Y(\omega) \in B\}|X) \\ g(x) &= P(Y(\omega) \in B|X = x) \end{aligned}$$

this function is called **the conditional probability distribution** of Y given $X = x$. In fact, this is more general formulation of the wellknown conditional probability

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

which requires that $P(B) > 0$, whereas the new definition well defined over the whole domain.

We will introduce the conditional probability distribution function in two different cases, namely; discrete and absolutely continuous cases

1. (X, Y) *discrete random vector*

   $$P(X = x, Y = y) = p_{X,Y}(x, y)$$

   Then we can define the function as follows

   $$p_{Y|X}(y|x) = \left\{ \begin{array}{ll} \frac{p_{X,Y}(x,y)}{p_X(x)} & p_X(x) \neq 0 \\ 0 & p_X(x) = 0 \end{array} \right\}$$

   Then

   $$P(Y \in B|X = x) = \sum_{y \in B} p_{Y|X}(y|x)$$

   Then the **conditinal expectation** will be calculated by

   $$E(g(Y)|X = x) = \sum_{y} g(y) p_{Y|X}(y|x)$$

2. (X, Y) *absolutely continuous random vector*

In this case we will have the following density function

$$f_{Y|X}(y|x) = \left\{ \begin{array}{lr} \frac{f_{X,Y}(x,y)}{f_X(x)} & f_X(x) \neq 0 \\ 0 & f_X(x) = 0 \end{array} \right\}$$

And consequently

$$P(Y \in B|X = x) = \int_B f_{Y|X}(y|x)dy$$

Then the **conditinal expectation** will be calculated by

$$E(g(Y)|X = x) = \int g(y) f_{Y|X}(y|x)dy$$

Definition Let $A_1, A_2.....$ be events and let $\mathcal{F}_0$ be σ-algebra. $A_1, A_2...$ are said to be **conditionally independent** given $\mathcal{F}_0$ if for every n and $i_1, i_2, ...i_n$

$$P(\cap_{j=1}^n A_{i_j}|\mathcal{F}_0) = \prod_{j=1}^n P(A_{i_j}|\mathcal{F}_0) \ \ \textbf{a.s}$$

Proposition $Y_1, Y_2...$ are conditionally independent given $\mathcal{F}_0$ if for every $B_1, B_2....(Borel\ Sets)$ the events $(Y_1 \in B_1), (Y_2 \in B_2)$ are conditionally independent given $\mathcal{F}_0$.

Example Toss a coin 100 times. Let

X_1 = the number of heads after first 10 tosses
X_2 = the number of heads after first 50 tosses
X_3 = the number of heads after first 70 tosses

Note that these three variables are not independent, but X_1 and X_3 are conditionally independent given X_2. In other words, knowing X_2 (the number of heads after first 50 tosses), X_1 (the number of heads after first 10 tosses) does not give any information on X_3. (the number of heads after first 70 tosses.)

Proof We will show that

$$P(X_1 = x_1, X_3 = x_3|X_2 = x_2) = P(X_1 = x_1|X_2 = x_2).P(X_3 = x_3|X_2 = x_2)$$

First note that the random variables X_1, $X_2 - X_1$, and $X_3 - X_2$ are independent. Then

$$\frac{P(X_1 = x_1, X_2 = x_2, X_3 = x_3)}{P(X_2 = x_2)} =^{same \text{ event}} \frac{P(X_1 = x_1, X_2 - X_1 = x_2 - x_1, X_3 - X_2 = x_3 - x_2)}{P(X_2 = x_2)}$$

We multiply and divide by $P(X_2 = x_2)$ and using indepence

$$\begin{aligned} &= \quad \frac{P(X_1 = x_1, X_2 - X_1 = x_2 - x_1)}{P(X_2 = x_2)} . P(X_3 - X_2 = x_3 - x_2) = \\ &= \quad \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} . \frac{P(X_2 = x_2, X_3 - X_2 = x_3 - x_2)}{P(X_2 = x_2)} \end{aligned}$$

Also noting that $P(X_1 = x_1, X_2 - X_1 = x_2 - x_1)$ and $P(X_1 = x_1, X_2 = x_2)$ are the same events(strictly speaking probability of the same event), we can conclude that

$$P(X_1 = x_1 | X_2 = x_2) . P(X_3 = x_3 | X_2 = x_2)$$