Lecture 9

$\boxed{\text{Def: } E[X|Y] \text{ is a function of } Y \text{ s.t. } E\{(X-E[X|Y])H(Y)\}=0 \text{ for any } H. \quad (*)}$

Thm (1) Definition (*) is consistent in the discrete case.
(2) $E[X|Y]$ minimizes the mean squared prediction error of X.
i.e. $E[X|Y]$ is the function $\psi(Y)$ such that $E[X|\psi(Y)]^2$ is minimized among all functions of Y.

Proof: For any function $g(Y)$, observe $E[X-g(Y)]^2 = E[(\underbrace{X-E(X|Y)}_{A} + \underbrace{E[X|Y]-g(Y)}_{B})^2]$

$= E[(A+B)^2] = EA^2 + EB^2 + 2E(AB)$

However, $E(AB) = E[(X-E(X|Y)] \cdot \underbrace{(E[X|Y]-g(Y))}_{\text{function of } y} = 0$

$\Rightarrow E[(X-g(Y))^2]$
$= E[(X-E[X|Y])^2] + E[(E[X|Y]-g(Y))^2]$

$\Rightarrow E[(X-g(Y))^2]$ is minimized iff $g(Y)=E[X|Y]$ a.s.

Notation. An event A is said to occur almost surely or a.s. if $P(A)=1$.

(3). (Uniqueness)
If $g_1(Y)$ & $g_2(Y)$ are two different versions of $E[X|Y]$ then $P[g_1(Y)=g_2(Y)]=1$
i.e. $g_1(Y)=g_2(Y)$ a.s.

Proof: Note that, by the proof of (2)
$E[(X-g_1(Y))^2] = E[(X-g_1(Y))^2] + E[(g_1(Y)-g_2(Y))^2]$ (essentially from)
$\Rightarrow E[(g_1(Y)-g_2(Y))^2] = 0$
$\Rightarrow P[g_1(Y)=g_2(Y)]=1$
$\Rightarrow \exists$ only 1 version of $E[X|Y]$.

Properties

① $E[E(X|Y)] = E(X)$ ☆ !!!
Proof: Let $H(Y)\equiv 1$ in (*)

② $E[\,\cdot\,|Y]$ when treated as an operator satisfies all five axioms of expectations

i. If $X \geq 0$, then $E[X|Y] \geq 0$ a.s.

Proof:
Let $g(Y)=E[X|Y]$
Write $g(Y)=g^+(Y)-g^-(Y)$
where $g^+(Y)=\max(g(Y),0)$
$g^-(Y)=\max(-g(Y),0)$
Observe that $g^+(Y), g^-(Y) \geq 0$

By (*) Let $H(Y)=g^-(Y)$
$\Rightarrow E[(X-g^+(Y)+g^-(Y))\, g^-(Y)]=0$
$\Rightarrow E[(X-g^+(Y))\, g^-(Y)] + E[(g^-(Y))^2]=0$

On the other hand, note that
$g^+(Y)\, g^-(Y)=0$
$\Rightarrow E[X\cdot g^-(Y)] = -E[(g^-(Y))^2]$
$\because X g^-(Y) \geq 0 \Rightarrow E[g^-(Y)X] \geq 0 \Rightarrow E[(g^-(Y))^2] \leq 0$
$\Rightarrow E[(g^-(Y))^2]=0 \Rightarrow g^-(Y)=0$ a.s. $\Rightarrow E[X|Y] \geq 0$ a.s.

HW: ii, iii, iv proof of conditional expectation (much easier)

v. If $X_1, \cdots, X_n$ monotonically converge to $X \Rightarrow E(X_i|Y) \to E(X|Y)$ as $i \to \infty$ a.s.

every conditional expectation only holds in this almost sure condition

proof: To save time, assume $X_i \uparrow X$ according to i $E[X_i|Y]$ also $\uparrow$

$\boxed{\text{Corollary: if } X \geq Z \Rightarrow E[X|Y] \geq E[Z|Y] \text{ a.s.}}$

and $E[X_i|Y] \leq E[X|Y]$ for any i.

Let $L = \lim_{n\to\infty} E[X_n|Y]$ & let $\Delta = X - L$

Now, by ⊛ $E[(X_i - E[X_i|Y])H(Y)] = 0$ & $E[(X - E[X|Y])H(Y)] = 0$

Let $H(Y) = \Delta$. Note that $\Delta \geq 0$ a.s.

We have

$E[X_i \Delta] = E[E[X_i|Y]\Delta]$ ①

$E[X\Delta] = E[E[X|Y]\Delta]$ ②

② − ① we observe LHS $= E[X\Delta] - E[X_i\Delta] \xrightarrow{X_i \uparrow X} E[X\Delta] = \lim_{i\to\infty} E[X_i\Delta]$

and let $i \to \infty$ then LHS $\to 0$

RHS $\to E(\Delta^2) \Rightarrow E\Delta^2 = 0$

$\Rightarrow \Delta = 0$ a.s.

more properties

③ Law of iterated conditional expectations. / "the Tower Law"

$E[E[X|Y_1, Y_2]|Y_1] = E[X|Y_1]$

(Ontario's average height)

⊛ $E[X|Y_1, Y_2]$ is the function of Y_1, Y_2 s.t.

$E[(X - E[X|Y_1, Y_2])H(Y_1, Y_2)] = 0$ for any H

proof: let $g_1(Y_1) = E[X|Y_1]$

$g_2(Y_2) = E[E[X|Y_1, Y_2]|Y_1]$

$h(Y_1, Y_2) = E[X|Y_1, Y_2]$

Task: to show $g_1(Y_1) = g_2(Y_1)$ a.s.

Note that

$E[(X - g_1(Y_1))H(Y_1)] = 0$ ✓

$E[(X - h(Y_1, Y_2))H(Y_1, Y_2)] = 0$ ✓✓

$E[(h(Y_1, Y_2) - g_2(Y_1))H(Y_1)] = 0$ ✓✓✓

Let $H(Y_1) = H(Y_1, Y_2) = g_1(Y_1) - g_2(Y_1)$ in ✓ ✓✓ & ✓✓✓

Consider ✓✓✓ + ✓✓ − ✓

HW

$\Rightarrow E[(g_1(Y_1) - g_2(Y_1))^2] = 0 \Rightarrow g_1(Y_1) = g_2(Y_1)$ a.s. ∎

④ If $E[X|Y_1, Y_2, Y_3]$ turns out to be a function of Y_1 only. let $\psi(Y_1)$ be the function, then $\psi(Y_1)$ is also $E[X|Y_1]$ and $E[X|Y_1, Y_2]$

Proof: By definition & assumption: $E[(X - \psi(Y_1))H(Y_1, Y_2, Y_3)] = 0$ for any H.

However, $H(Y_1, Y_2, Y_3)$ can be chosen as any function of Y_1 or any function of (Y_1, Y_2) or (Y_1, Y_3) ∎

⑤ $V[X] = E(V[X|Y]) + V[E[X|Y]]$

Proof: By definition, $V[X|Y] = E[X^2|Y] - E[X|Y]^2$

Therefore, $E[V[X|Y]] = E[E[X^2|Y]] - E[(E[X|Y])^2]$

$= E[X^2] - E[(E(X|Y))^2]$

$= E(X^2) - (EX)^2 + (EX)^2 - E[(E(X|Y))^2]$

$= V[X] - [E[(E(X|Y))^2] - (EX)^2]$

Note: $V[E(X|Y)] = E[(E(X|Y))^2] - (E[E[X|Y]])^2$

$= E[(E[X|Y])^2] - (EX)^2$

∎

Ex: Select n items out of finished products. Each item has probability p of being defect, suppose that p is random and changes from day to day. In detail, $p \sim U(0, \frac{1}{2})$ distribution. Let X be # of defect items for any single day. Find $E(X)$ & $V[X]$

Solution: Conditioning on p

$$E[X] = E[E[X|p]] = E[np] = nE[p] = \frac{n}{4}$$

$$V[X] = V[E[X|p]] + E[V[X|p]] = V[np] + E[np(1-p)] = n^2 V[p] + nE[p(1-p)]$$

$$E[p^2] = \int_0^{\frac{1}{2}} p^2 \underbrace{(\text{density})}_{\shortparallel\ 2} dp = \int_0^{\frac{1}{2}} 2p^2\, dp = \frac{2}{3}p^3 \Big|_0^{\frac{1}{2}} = \frac{2}{24} = \frac{1}{12}$$

$$V[p] = \frac{1}{12} - \left(\frac{1}{4}\right)^2 = \frac{4}{48} - \frac{3}{48} = \frac{1}{48}$$

$$E[p(1-p)] = Ep - Ep^2 = \frac{1}{4} - \frac{1}{12} = \frac{1}{6}$$

$$V[X] = n^2 \cdot \frac{1}{48} + n \cdot \frac{1}{6} = \frac{n^2}{48} + \frac{n}{6}$$

More on INDEPENDENCE

Thm: If X, Y are independent, and $E[H^2(X)] < \infty$ for some $H(\ )$

Then $E[H[X]|Y] = E[H(X)]$ a.s.

Proof: $E[\{H(X) - E(H(X))\} G(Y)] = 0$ for any G.

Skip the following chapters: § 5.4 (Sigma field), § 5.6, § 5.7, § 5.8

CHAPTER 8

CONTINUOUS R.V. and Their Transfor-mations

$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{pmatrix}$ is a continuous random vector of dimension m if there exists a function $f(x_1, \cdots, x_m) \geq 0$ s.t. $P(\vec{X} \in A) = \iint_A f(x_1, \cdots, x_m)\, dx_1 \cdots dx_m$ for any set $A \subset \mathbb{R}^m$

In this case $f(\cdot)$ is called the joint density function of $\vec{X}$.

CDF (cumulative density function)

$F(x_1, x_2, \cdots, x_m) = P[X_1 \leq x_1, X_2 \leq x_2, \cdots, X_m \leq x_m] = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_m} f(y_1, \cdots, y_m)\, dy_1 \cdots dy_m$

is called the CDF of $\vec{X}$.

$$\boxed{\text{Corollary: } f(x_1, x_2, \cdots, x_m) = \frac{\partial^m F(x_1, \cdots, x_m)}{\partial x_1 \partial x_2 \cdots \partial x_m}}$$

Thm: (Proof not required)

For any $H: \mathbb{R}^m \to \mathbb{R}^d$

$$E[H(\vec{X})] = \iint \cdots \int H(x_1, \cdots, x_m) f(x_1, \cdots, x_m)\, dx_1 \cdots dx_m$$

Thm: For $r \leq m$, the density of $\begin{pmatrix} X_1 \\ \vdots \\ X_r \end{pmatrix}$ is $\iint \cdots \int f(x_1, \cdots, x_m)\, dx_{r+1}\, dx_{r+2} \cdots dx_m$

remove the affects from the rest parts

HW: proof of this thm

Cor: For any i, $1 \leq i \leq m$, the marginal density function of X_i is

$$f_{X_i}(x_i) = \iint \cdots \int f(x_1, \cdots, x_m)\, dx_1 dx_2 \cdots dx_{i-1}\, dx_{i+1}\, dx_{i+2} \cdots dx_m$$