* Conditional Mean Function

· Conditional Density : $f(y|x) = \dfrac{f(y,x)}{f(x)}$ where $\underbrace{f(x) = \int_{-\infty}^{\infty} f(y,x)\,dy}_{\text{marginal density}}$

· Conditional Mean Function.

$$\underline{E(y|x) = \int_{-\infty}^{\infty} y \cdot f(y|x)\,dy} \quad (y: \text{scalar},\ x: \text{vector})$$

↳ ※ 이게 f(y|x) 자체를 나타내는 것은 아니므로 주의!

Holding variables as constant, we need to check an environment.

So we use conditional expectation to compare states fairly.

ex) Group A $x=100$

Group B $x=200$

Experimental data

vs

Observational data. $\left.\begin{array}{l} x=100,\ z=\cdots \\ x=200,\ z=\cdots \end{array}\right] \Rightarrow$ Those other conditions can be different.

⇓

Therefore, we <u>use conditioning argument</u> to make the groups as "<u>comparable as possible</u>"

⇒ Conditional Mean Function is one aspect of the conditional distribution.

There are many others : conditional median

conditional percentiles : 10%. 25%. 50%. 75%. 90%.

50% = median

conditional density : "

Note

$E(Y \mid X = x) = m(x)$: Conditional mean function

Y, X ↑ random variable; x ↑ actual value

(In class, he will use them like this)

· $m(X)$: a random variable, $m(X) = E(Y|X)$

· properties of $m(x)$

$E(aY_1 + bY_2 \mid X=x) = a\,E(Y_1|X=x) + b\,E(Y_2|X=x)$

$E(\,a(X)Y \mid X=x) = E(a(x)Y|X=x) = a(x) \cdot E(Y|X=x)$

Note 2) law of iterated expectations $E(Y) = E(E(Y|X))$

- $Var(Y|X=x) \overset{def}{=} E\{[Y - E(Y|X=x)]^2 \mid X=x\}$

conditional variance

$$= E\{Y^2 - 2Y\cdot E(Y|X=x) + E(Y|X=x)^2 \mid X=x\}$$

$$= E\{Y^2|X=x\} - 2E\{Y\cdot E(Y|X=x) \mid X=x\} + E\{E(Y|X=x)^2 \mid X=x\}$$

$$= E\{Y^2|X=x\} - 2E(Y|X=x)\cdot E(Y|X=x) + E(Y|X=x)^2\cdot \underbrace{E(1|X=x)}_{=1}$$

$$= E(Y^2|X=x) - E(Y|X=x)^2$$

- $Var\, Y \overset{def}{=} E(Y-E(Y))^2 = E(\boxed{Y - E(Y|X)} + \boxed{E(Y|X) - E(Y)})^2$

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

$$= \underbrace{E\{[Y-E(Y|X)]^2\}}_{①} + \underbrace{E\{[E(Y|X)-E(Y)]^2\}}_{②} + \underbrace{2E\{(Y-E(Y|X))(E(Y|X)-E(Y))\}}_{③}$$

① $E\{[Y-E(Y|X)]^2\} \underset{\text{L.I.E}}{=} E[E\{[Y-E(Y|X)]^2|X\}] = E[Var(Y|X)]$

② $E\{[E(Y|X)-E(Y)]^2\} = E\{[\underline{E(Y|X)} - E(\underline{E(Y|X)})]^2 = Var(\underline{E(Y|X)})$

③ $2E\{[Y-E(Y|X)][E(Y|X)-E(Y)]\} = 2E[E\{[Y-E(Y|X)]\underbrace{[E(Y|X)-E(Y)]}_{\text{"constant"}}|X\}]$

$$= 2E[\underbrace{E\{[Y-E(Y|X)]|X\}}_{=0}\cdot\{E(Y|X)-E(Y)\}] = 0$$

(Because $E\{Y-E(Y|X)|X\} = E(Y|X) - E(E(Y|X)|X)$

$= E(Y|X) - E(Y|X)\cdot \underbrace{E(1|X)}_{=1} = 0$

$$= E(Var(Y|X)) + Var(E(Y|X))$$

- When $E(Y|X=x)$ is well-defined for each x in the support of X,

We can always write :
$$\boxed{\begin{cases} Y = m(X) + \varepsilon \\ E(\varepsilon|X) = 0 \end{cases}}$$

proof)

$$m(X) = E(Y|X=x)$$

$$\varepsilon = Y - m(X) \quad \Rightarrow \quad Y = m(X) + \varepsilon.$$

$$E(\varepsilon|X) = E(Y - m(X)|X) = E(Y|X) - E(m(X)|X)$$
$$= E(Y|X) - m(X) = E(Y|X) - E(Y|X) = 0. \quad ▨$$

- Causal Effect

$Y = Y(X, w)$: Y depends on X and $\underbrace{\text{something else}}_{=w}$.

w is an additional randomness in Y given X.

Note "Parameters are not random variables."

probability model

- Suppose the case "no-conditioning"

$$\underset{\text{random v.}}{Y} \sim \underset{\text{parameter.}}{f(Y)} \quad \text{(non-parametric case)}$$

$Y = Y_1, \ldots, Y_n$

$$Y \sim N(\underset{\text{parameter}}{\mu}, \sigma^2) \quad \text{(parametric case)}$$

- Suppose conditional distribution case

$$Y|X=x \sim \underset{\text{parameter}}{f(Y|X=x)}$$

$$Y|X=x \sim N(\underbrace{\mu(x), \sigma^2(x)}_{\text{parameter}})$$

- Average Treatment Effect of changing X from x_2 to x_1

When X is a scalar random variable,

$$E(Y(x_1,W)-Y(x_2,W)) = E(Y(x_1,W)) - E(Y(x_2,W))$$

Note that $\underbrace{E(Y(x_1,W))}_{\text{unconditional}} \neq \underbrace{E(Y(X,W)\mid X=x_1)}_{\text{conditional}} = E(Y\mid X=x_1) = E(Y(x_1,W)\mid X=x_1)$

(It is equal if W & X are independent.)

- Let $X=(X_1,X_2)$. X_1: Scalar, X_2: Vector

Then, Average Treatment Effect of changing X_1 from x_1 to x_1' holding $X_2=x_2$ is

$$\overset{\text{def}}{=} E\{Y(x_1',x_2,W) - Y(x_1,x_2,W)\} = E(Y(x_1',x_2,W) - Y(x_1,x_2,W))$$

Note that $E[Y(X_1,X_2,W) \mid X_1=x_1', X_2=x_2]$

$$= E[Y(x_1',x_2,W) \mid X_1=x_1', X_2=x_2]$$

If W & X_1 are independent given X_2, ⊛ $= E[Y(x_1',x_2,W) \mid X_2=x_2]$

Analogously, under ⊛, $E[Y(X_1,X_2,W) \mid X_1=x_1, X_2=x_2] = E[Y(x_1,x_2,W) \mid X_2=x_2]$

- Conditional Average Treatment Effect, given $X_2=x_2$, of changing X_1 from x_1 to x_1'

Under ⊛,

$$E(Y\mid X_1=x_1', X_2=x_2) - E(Y\mid X_1=x_1, X_2=x_2)$$

$$\overset{*}{=} E(Y(x_1',x_2,W)\mid X_2=x_2) - E(Y(x_1,x_2,W)\mid X_2=x_2)$$

$$= E(Y(x_1',x_2,W) - Y(x_1,x_2,W)\mid X_2=x_2)$$

More specifically.

Conditional Average Treatment Effect given $X_2 = x_2$ under ⊛

$$E(Y \mid X_1 = x_1', X_2 = x_2) - E(Y \mid X_1 = x_1, X_2 = x_2)$$

By note, $E(Y \mid X_1 = x_1', X_2 = x_2) = E(Y(X_1, X_2, W) \mid X_1 = x_1', X_2 = x_2)$

$= E(Y(x_1', x_2, W) \mid X_1 = x_1', X_2 = x_2)$

Also, $E(Y \mid X_1 = x_1, X_2 = x_2) = E(Y(X_1, X_2, W) \mid X_1 = x_1, X_2 = x_2)$

$= E(Y(x_1, x_2, W) \mid X_1 = x_1, X_2 = x_2)$

$$= E(Y(x_1', x_2, W) \mid \underline{X_1 = x_1'}, X_2 = x_2) - E(Y(x_1, x_2, W) \mid \underline{X_1 = x_1}, X_2 = x_2)$$

Under ⊛, we can drop this part.

$$= E(Y(x_1', x_2, W) \mid X_2 = x_2) - E(Y(x_1, x_2, W) \mid X_2 = x_2)$$

$$= E(Y(x_1', x_2, W) - Y(x_1, x_2, W) \mid X_2 = x_2)$$

※ Two roles of conditioning

① To guarantee conditional independence assumption ⊛ to hold.

② To isolate the group of interest.