Conditional Mean Function

Conditional Density $f(y|x) = \frac{f(y,x)}{f(x)}$ $f(x) = \int_{-\infty}^{\infty} f(y,x)\,dy$

Conditional Mean

$$E(y|x) = \int_{-\infty}^{\infty} y\, f(y|x)\,dy$$

y scalar x vector

Observational Data vs. Experimental. Econ mostly the former.

In experimental, can control design. In observational, must control using conditional. Use conditioning argument to make groups as comparable as possible.

- Conditional Mean tells less info than con. density. Doesn't say much about $dist^n$ itself.
- Conditional Mean is one aspect of the conditional $dist^n$. There are others
  - Conditional median/percentiles.
  - Conditional density

Note: Capital letters denote matrix. lowercase vector.

Here, not so for now

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

lc, up case

$m(X)$ is a random variable.

$$\overset{\shortparallel}{E}(Y|X)$$

Properties of $m(x) = \mathbb{E}(Y \mid X=x)$

① Linear $\mathbb{E}(aY_1 + bY_2 \mid X=x) = a\mathbb{E}(Y_1 \mid X=x) + b\mathbb{E}(Y_2 \mid X=x)$

② $\mathbb{E}(a(X)Y \mid X=x) = \mathbb{E}(a(x)Y \mid X=x) = a(X)\mathbb{E}(Y \mid X=x)$

③ Law Iterated Exp $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|X)]$

$$\begin{aligned}
Var(Y \mid X=x) &= \mathbb{E}[(Y - \mathbb{E}(Y|X=x))^2 \mid X=x] \\
&= \mathbb{E}[Y^2 - 2Y\mathbb{E}(Y|X=x) + \mathbb{E}(Y|X=x)^2 \mid X=x] \\
&= \mathbb{E}[Y^2|X=x] - 2\mathbb{E}[Y\,\mathbb{E}(Y|X=x) \mid X=x] + \mathbb{E}(\mathbb{E}(Y|X=x)^2 \mid X=x) \\
&= \mathbb{E}[Y^2 \mid X=x] - 2\mathbb{E}[Y|X=x] \overset{\text{Prop}(2)}{\mathbb{E}}[Y|X=x] + \mathbb{E}(Y|X=x)^2 \underbrace{\mathbb{E}(1|X=x)}_{1} \\
&= \mathbb{E}(Y^2 \mid X=x) - \mathbb{E}(Y|X=x)^2
\end{aligned}$$

$$\begin{aligned}
Var(Y) &\overset{\text{def}}{=} \mathbb{E}([Y - \mathbb{E}(Y)]^2) \\
&= \mathbb{E}\{[Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\} \\
&= \mathbb{E}\{[Y - \mathbb{E}(Y|X)]^2 + [\mathbb{E}(Y|X) - \mathbb{E}(Y)]^2 + 2[Y - \mathbb{E}(Y|X)][\mathbb{E}(Y|X) - \mathbb{E}(Y)]\} \\
&= \mathbb{E}\{[Y - \mathbb{E}(Y|X)]^2\} + \mathbb{E}\{[\mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\} + 2\mathbb{E}\{[Y - \mathbb{E}(Y|X)][\mathbb{E}(Y|X) - \mathbb{E}(Y)]\} \\
&= \mathbb{E}[\underbrace{\mathbb{E}((Y - \mathbb{E}(Y|X))^2 \mid X)}_{Var(Y|X)}] + \mathbb{E}\{[\mathbb{E}(Y|X) - \mathbb{E}(Y)]^2\} + \underbrace{2\mathbb{E}\{\mathbb{E}\{[Y - \mathbb{E}(Y|X)][\mathbb{E}(Y|X) - \mathbb{E}(Y)] \mid X\}\}}_{2\mathbb{E}\{\underbrace{\mathbb{E}([Y - \mathbb{E}(Y|X)] \mid X)}_{0}(\mathbb{E}(Y|X) - \mathbb{E}(Y))\}} \\
&= \mathbb{E}(Var(Y|X)) + Var(\mathbb{E}(Y|X))
\end{aligned}$$

$=0$ because $\mathbb{E}[Y - \mathbb{E}(Y|X)|X] = \mathbb{E}(Y|X) - \mathbb{E}(\mathbb{E}(Y|X)|X)$
$= \mathbb{E}(Y|X) - \mathbb{E}(Y|X)\mathbb{E}(1|X) = 0$ $\square$

When $\mathbb{E}[Y|X=x]$ is well defined for each x for support of X
we can always write:

$$Y = m(\overset{cap}{X}) + \varepsilon$$

$$\mathbb{E}(\varepsilon|X) = 0$$

Proof: $m(x) = \mathbb{E}(Y|X=x)$

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

$$\mathbb{E}(\varepsilon|X) = \mathbb{E}(Y - m(X)|X) = \mathbb{E}(Y|X) - m(X)$$
$$= \mathbb{E}(Y|X) - \mathbb{E}(Y|X) = 0$$

Causal Effect:

$Y = Y(X, \omega)$ assume y is a r.v. after conditioning on X, ω is additional randomness in Y given X

Aside: Probability model $Y \sim f(Y)$

We have $Y_1, \dots, Y_n$. f is parameter in nonparametric model

$Y \sim N(\mu, \sigma^2)$ here, μ, σ^2 is parameter.

In above, ω is not parameter, but r.v.

$$Y|X=x \sim f(Y|X=x) \quad \text{or} \quad Y|X=x \sim N(\mu(x), \sigma^2(x))$$

Average Treatment Effect

Consider a case when X is a scalar random variable.

of changing X from x_2 to x_1

$$\mathbb{E}(Y(x_1, \omega) - Y(x_2, \omega)) = \mathbb{E}(Y(x_1, \omega)) - \mathbb{E}(Y(x_2, \omega))$$

Note: $\mathbb{E}(Y(x_1, \omega)) \neq \mathbb{E}(Y(X, \omega) \mid X = x_1)$

$\mathbb{E}(Y(x_1, \omega))$: unconditional; $\mathbb{E}(Y(X, \omega) \mid X = x_1)$: conditioning on x_1

It is equal if $\mathbb{E}(Y \mid X = x_1) = \mathbb{E}(Y(x_1, \omega) \mid X = x_1)$

ω and X are independent

$X = (X_1, X_2)$. X_1 is a scalar, X_2 is a vector.

Average treatment effect of changing X_1 from x_1 to x_1' holding $X_2 = x_2$ constant is:

$$\mathbb{E}(Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega))$$

$$= \mathbb{E}(Y(x_1', x_2, \omega)) - \mathbb{E}(Y(x_1, x_2, \omega))$$

Note that $\mathbb{E}(Y(X_1, X_2, \omega) \mid X_1 = x_1', X_2 = x_2)$

$$= \mathbb{E}(Y(x_1', x_2, \omega) \mid X_1 = x_1', X_2 = x_2)$$

If ω and X_1 are independent given X_2 (*)

$$= \mathbb{E}(Y(x_1', x_2, \omega) \mid X_2 = x_2)$$

Analogously, if w and X_1 are independent given X_2

$$\mathbb{E}(Y(X_1, X_2, w) \mid X_1 = x_1, X_2 = x_2) = \mathbb{E}(Y(X_1, X_2, w) \mid X_2 = x_2)$$

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

under assump. of indepence from above! (*)

$$= \mathbb{E}(Y(x_1', x_2, w) \mid X_2 = x_2) - \mathbb{E}(Y(x_1, x_2, w) \mid X_2 = x_2)$$

Conditional Avg treatment effect: (given $X_2 = x_2$)

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

by assuming X_1 is independent of w given x_2, allows us to say different is not between groups, but rather the outcome

cannot interpret something as causal effect w/out this condition independence assumption.

Two roles of conditioning:

① Guarantee conditional independence assumption

② Isolate the group of interest.

Sect. 2.1.4 sect. 2.30

Recall: $Y(\overbrace{X_1}^{\text{indexed}}, X_2, \overset{\text{random}}{\underset{\uparrow}{\omega}})$ and X_2 are independent given $X_2 = x_2$ for any x_1, x_2 in the support of (X_1, X_2). Then $\mathbb{E}(Y \mid X_1 = x_1', X_2 = x_2') - \mathbb{E}(Y \mid X_1 = x_1, Y = x_2)$ is equivalent to:

$$\mathbb{E}\left(Y(x_1', x_2, \omega) - Y(x_1, x_2, \omega) \mid X_2 = x_2\right)$$

Y being driven by ω. So, ω & x_1 independent means same thing above.

Role of Conditioning

① Make $Y(x_1, x_2, \omega)$ & X_1 independent given X_2

② To define parameter of interest

Example: JPE article Rosenzweig & Wolpin (in syllabus) (one of first papers to use twins in econ)

Y(number of children, x_2, ω)

↑

labor force participation decision: $\underset{\text{not}}{0}$ or $\underset{\text{participate}}{1}$

$\mathbb{E}(Y \mid X_1 = 1, X_2 = x_2) - \mathbb{E}(Y \mid X_1 = 0, X_2 = x_2)$ $X_2 = \{\text{age, family income}\}$

Now:

$\mathbb{E}(Y \mid \underset{\text{twin}}{X_1 = 1}, X_2 = x_2, \text{first time child} = 1) - \mathbb{E}(Y \mid \underset{\text{1 child}}{X_1 = 0}, X_2 = x_2, \text{first time} = 1)$

Holding X_2, first time constant, can look at having one child vs 2 at once, use variation.

But need to have race in X_2 as well (original paper did not)

*Conditioning on age and race is important for reason (1)
Need to carefully consider conditioning for assumption to hold.

$X_2 = (\underset{\text{group of interest}}{X_{21}}, \underset{\text{conditioning variables to secure (2)}}{X_{22}})$

$\mathbb{E}(Y(x_1', X_2, \omega) - Y(x_1, X_2, \omega) \mid X_{21} = x_{21})$

$= \mathbb{E}(\mathbb{E}(Y(x_1', X_2, \omega) - Y(x_1, X_2, \omega) \mid X_2) \mid X_{21} = x_{21})$

Use distribution of $X_{22} \mid X_{21} = x_{21}$ to integrate out X_{22}

Linear Regression Model

$$m(x) = \beta_0 + \beta_1 r_1(x) + \beta_2 r_2(x) + \dots + \beta_k r_k(x) = r(x)'\beta$$

$r_j(x)$ $j = 1, \dots, k$ are known functions.

$\beta_0, \beta_1, \dots, \beta_k$ are unknown constants.

$$r(x) = \begin{bmatrix} 1 \\ r_1(x) \\ \vdots \\ r_k(x) \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

Two regressor case: $m(x_1, x_2) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1 x_2$

$r_0(X) = 1$

$r_1(X) = X_1$

$r_2(X) = X_2$

$r_3(X) = X_1^2$

$r_4(X) = X_2^2$

$r_5(X) = X_1 X_2$

$$m(x_1, x_2) = \mathbb{E}(Y \mid X_1 = x_1, X_2 = x_2) = [r_0(X), \dots, r_5(X)] \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

often $\mathbb{E}(Y \mid X = x) = x'\beta$.

$$m(x) = x'\beta \qquad X = \begin{bmatrix} 1 \\ X_1 \\ \vdots \\ X_k \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_k \end{bmatrix}$$

We allow x to include a constant term and keep $\beta \in R^k$.

$$Y = X'\beta + U \qquad x'\beta = \mathbb{E}(Y \mid X = x)$$

$$\Leftrightarrow \mathbb{E}(U \mid X = x) = 0$$

Once $Y = X'\beta + u$ is maintained, $X'\beta \Leftrightarrow \mathbb{E}(U \mid X = x) = 0$

Assumption: (*) $\mathbb{E}(Y|X=x) = x'\beta \Leftrightarrow \mathbb{E}(u|X=x) = 0$

① Which x to use

② How to put together $r(x)$

How to estimate β?

$\min_{g(\cdot)} \mathbb{E}((Y - g(X))^2)$ Want to minimize estimator of Y

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

$$\Rightarrow g(X) = \mathbb{E}(Y|X).$$

Let $g(x) = x'b$. Still true that this relationship holds.

$$\mathbb{E}\{[Y - X'b]^2\} = \mathbb{E}\{[Y - \mathbb{E}(X|Y)]^2\} + \mathbb{E}\{\underbrace{[\overset{x'\beta}{\overset{\|}{\mathbb{E}(Y|X)}} - x'\beta]^2}_{[X'(\beta-b)]^2}\}$$

We use sample analog: $\min_{b \in \mathbb{R}^k} \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'b)^2$

Sampling of (X, Y) results in (x_i, y_i) for $i = 1, \dots, N$

The solution to this problem is called OLS estimate.

$$\mathbb{Y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} \qquad \mathbb{X} = \begin{bmatrix} x_1' \\ x_2' \\ \vdots \\ x_N' \end{bmatrix}$$

$n \times k$ matrix

x_i' row vector

Objective function

$$(\mathbb{Y} - \mathbb{X}b)'(\mathbb{Y} - \mathbb{X}b)$$

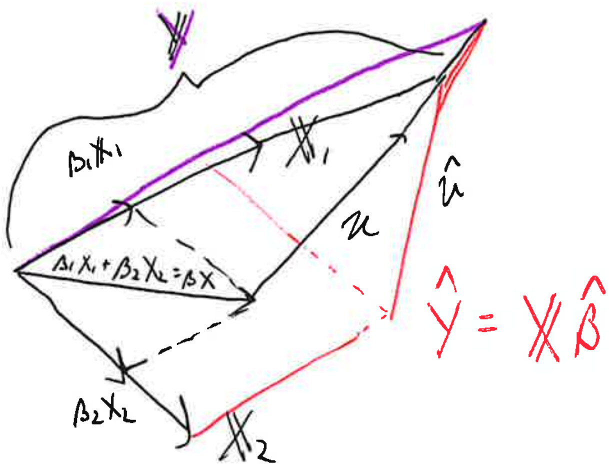
First order Condition: $-\mathbb{X}'(\mathbb{Y} - \mathbb{X}b) = 0$

Solution $\hat{\beta} = (\mathbb{X}'\mathbb{X})^{-1}\mathbb{X}\mathbb{Y}$

When $\mathbb{X}$ has full rank $\Leftrightarrow$ $\mathbb{X}'\mathbb{X}$ is invertible

Geometry of OLS $\mathbb{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_3 \end{bmatrix}$ $k=2$ $\mathbb{X}_1 = \begin{bmatrix} X_{11} \\ X_{12} \\ X_{13} \end{bmatrix}$ $\mathbb{X}_2 = \begin{bmatrix} X_{21} \\ X_{22} \\ X_{23} \end{bmatrix}$

$\mathbb{Y} = \mathbb{X}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \mathcal{U}$ $\quad \mathcal{U} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$



Look at various linear combos of $\mathbb{X}_1$ & $\mathbb{X}_2$ to minimize distance of y

This decomposition can't be done if X_1 X_2 are co-linear

$\hat{u}$ and $\mathbb{X}_1$ are ortogonal (and $\hat{u}$ and $\mathbb{X}_2$)

$\mathbb{X}_1'\hat{u} = 0 \quad \mathbb{X}_2'\hat{u} = 0$ by construction

$$\mathbb{X}'\underbrace{(\mathbb{Y} - \mathbb{X}\hat{\beta})}_{\hat{u}} = 0$$

$$\mathbb{X} = [\mathbb{X}_1, \dots, \mathbb{X}_k] \Rightarrow \mathbb{X}' = \begin{bmatrix} \mathbb{X}_1' \\ \vdots \\ X_k' \end{bmatrix} \quad \mathbb{X}\hat{u} = \begin{bmatrix} \mathbb{X}_1'\hat{u} \\ \vdots \\ \mathbb{X}_k'\hat{u} \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Alternative motivation for the OLS

$\mathbb{E}(u|X) = 0 \quad \mathbb{E}(Xu) = 0 \Leftrightarrow \mathbb{E}(X, u) = 0$

(**)

$\Leftrightarrow \frac{1}{N}\sum_{i=1}^{N} x_{1,i}(y_i - x_i'b) = 0$

Method of Moments for OLS

$\vdots$

(first was exploiting properties of conditional mean)

$\frac{1}{N}\sum x_{k,i}(y_i - x_i'b) = 0$

Sampling analog of the moment condition

$$\underset{N\times 1}{\mathbb{Y}} = \underset{N\times 1}{\mathbb{X}_1}\hat{\beta}_1 + \underset{N\times 1}{\mathbb{X}_2}\hat{\beta}_2 + \dots + \underset{N\times 1}{\mathbb{X}_k}\hat{\beta}_k + \underset{N\times 1}{\hat{\mathbb{U}}}$$