5) Conditional Expectation and Projection

Vitor Kamada

January 2018

Law of Iterated Expectation (Total Expectation)

$$E[E(Y|X)] = E[g(X)]$$

$$\sum_{x} g(x)P(X = x)$$

$$\sum_{x}[\sum_{y} yP(Y = y|X = x)]P(X = x)$$

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

$$\sum_{y} yP(Y = y) = E(Y)$$

Example: Law of Iterated Expectation

$$E[E(Y|X)] = E(Y)$$

$$E(Wage) = 10.9 \textbf{ and } E(Educ) = 11.5$$

$$E(Wage|Educ) = 4 + .6Educ$$

$$E[E(Wage|Educ)] = E[4 + .6Educ]$$

$$= E(4) + .6E(Educ)$$

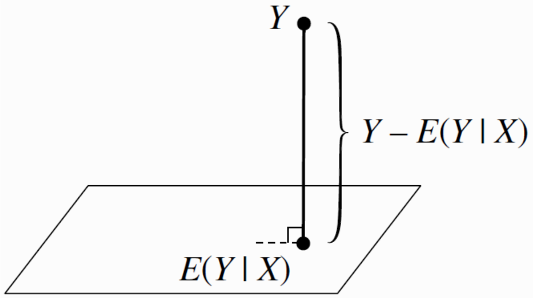
$$= 4 + .6 * 11.5 = 10.9$$

Projection Interpretation

$$E[(Y - E(Y|X))h(X)]$$

$$E[h(X)Y] - E[h(X)E(Y|X)]$$

$$E[h(X)Y] - E[E(h(X)Y|X)] = 0$$



Linear Regression: Assumptions

$$Y = \alpha + \beta X + \epsilon$$

$$E(\epsilon) = E[E(\epsilon|X)]$$
$$= E(0) = 0$$

$$E(\epsilon X) = E[E(\epsilon X|X)]$$
$$= E[XE(\epsilon|X)]$$
$$= E(0) = 0$$

Linear Regression: Estimation

$$Y = \alpha + \beta X + \epsilon$$

$$Cov(X, Y)$$

$$= Cov(X, \alpha) + \beta Cov(X, X) + Cov(X, \epsilon)$$

$$= \beta Var(X)$$

$$\beta = \frac{Cov(X,Y)}{Var(X)}$$

Conditional Variance

$$Var(Y) = E[Y - E(Y)]^2$$

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

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

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

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

Law of Total Variance

$$g(X) = E(Y|X) \textbf{ and } E[g(X)] = E(Y)$$

$$E[Var(Y|X)]$$

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

$$Var[E(Y|X)]$$

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

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

Homoskedasticity vs Heteroskedasticity

$$Y = \alpha + \beta X + \epsilon$$

$$Var(\epsilon|X) = E(\epsilon^2|X) - [E(\epsilon|X)]^2$$

$$E(\epsilon^2|X) = \sigma^2$$

$$E(\epsilon^2|X) = \sigma^2(X)$$

Conditional Independence Assumption (CIA)

$$E(\epsilon|x_1, x_2) = E(\epsilon|x_2)$$

$$E(W_J|C=0) = 10 \text{ and } E(W_J|C=1) = 20$$

$$E(W_G|C=0) = 8 \text{ and } E(W_G|C=1) = 12$$

$\frac{1}{2}$ of people is J, and other $\frac{1}{2}$ is G

$$ACE = \$7$$

Violation of CIA

$$P(C|H) = 3/4 \text{ and } P(C|L) = 1/4$$

$$P(H|J) = 3/4 \text{ and } P(H|G) = 1/4$$

$$P(C|J) = P(\tfrac{C}{H})P(\tfrac{H}{J}) + P(\tfrac{C}{L})P(\tfrac{L}{J}) = 62.5\%$$

$$P(C|G) = 37.5\%$$

	\$8	\$10	\$12	\$20	Mean
High-School	10	6	0	0	\$8.75
College	0	0	6	10	\$17.00

$$E(W|C) = 8.75 + 8.25C$$

Conditioning on the Test Score

	\$8	\$10	\$12	\$20	Mean
High-School + High Test Score	1	3	0	0	\$9.50
College + High Test Score	0	0	3	9	\$18.00
High-School + Low Test Score	9	3	0	0	\$8.50
College + Low Test Score	0	0	3	1	\$14.00

$$E(W|C,H) = 8.5 + 1H + 5.5C + 3HC$$

$$E(C|H) = 18 - 9.5 = 8.5$$

$$E(C|L) = 14 - 8.5 = 5.5$$

$$ACE = \$7$$

Selection Bias

<table>
<tr><th rowspan="3"></th><th colspan="2">Potential</th><th colspan="2">Causal</th><th colspan="2">Job</th></tr>
<tr><th colspan="2">Job</th><th colspan="2">Wage</th><th colspan="2">Avg Wage</th></tr>
<tr><th colspan="2">Degree</th><th colspan="2">Degree</th><th colspan="2">Degree</th></tr>
<tr><th>Type</th><th>No</th><th>Yes</th><th>No</th><th>Yes</th><th>No</th><th>Yes</th></tr>
<tr><td>Always Blue</td><td>Blue</td><td>Blue</td><td>1K</td><td>1.5K</td><td rowspan="2">1.5K</td><td>1.5K</td></tr>
<tr><td>Blue White</td><td>Blue</td><td>White</td><td>2K</td><td>2.5K</td><td rowspan="2">3K</td></tr>
<tr><td>Always White</td><td>White</td><td>White</td><td>3K</td><td>3.5K</td><td>3K</td></tr>
</table>