

Recall

- Covariance

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

- X, Y uncorrelated $\Leftrightarrow \text{Cov}(X, Y) = 0$

- X, Y uncorrelated $\Rightarrow \text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y)$

"Conditional Expectation Revisited"

Given X, Y define a new rv $\mathbb{E}(X|Y)$ as follows

- For each Y , compute $\mathbb{E}(X|Y=y)$

- this defines a function $g(y)$

- Define $\mathbb{E}(X|Y) = g(Y)$ i.e.

Did a few examples...

Law of Iterated Expectations: $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$

For any function $h(Y)$,

- $\mathbb{E}(h(Y)|Y) = h(Y)$

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

Can think of $\mathbb{E}(X|Y)$ as an estimator of X given Y .

In what sense does it "act like an estimator?"

- $\mathbb{E}(X|Y) = \mathbb{E}(X)$ by law of iterated expectations

- The estimation error $X - \mathbb{E}(X|Y)$ is uncorrelated w/ the estimate $\mathbb{E}(X|Y)$ - in fact, $X - \mathbb{E}(X|Y)$ is uncorrelated with Y - More generally, w/ any function $h(Y)$

Note: in class we went straight from (1) \rightarrow (2) \rightarrow (3). The work in between was done by me in an attempt to organize things/understand; so if it is incorrect I apologize.

To see this,

$$(1) \text{Cov}(\underbrace{x - \mathbb{E}(x|y)}_{\text{a rv-A}}, \underbrace{\mathbb{E}(x|y)}_{\text{a rv-B}}) = \mathbb{E}[(A - \mathbb{E}(A))(B - \mathbb{E}(B))]$$

$$= \mathbb{E}\left[\left(x - \mathbb{E}(x|y) - \underbrace{\mathbb{E}[x - \mathbb{E}(x|y)]}_{0}\right)\left(\mathbb{E}(x|y) - \underbrace{\mathbb{E}(\mathbb{E}(x|y))}_x\right)\right]$$

$$(2) \quad = \mathbb{E}\left[\left(\underbrace{x - \mathbb{E}(x|y)}_{\text{has zero mean}}\right)(\mathbb{E}(x|y) - \underbrace{\mathbb{E}(x)}_{\text{mean } \mathbb{E}(x)})\right]$$

$$= \mathbb{E}(x) - \mathbb{E}(\mathbb{E}(x|y))$$

$$= 0$$

$$= \mathbb{E}\left[(x - \mathbb{E}(x|y))\mathbb{E}(x|y) - (x - \mathbb{E}(x|y))\mathbb{E}(x)\right]$$

$$(3) \quad = \underbrace{\mathbb{E}\left[(x - \mathbb{E}(x|y))\mathbb{E}(x|y)\right]}_{=0 \text{ (shown below)}} - \underbrace{\mathbb{E}\left[(x - \mathbb{E}(x|y))\mathbb{E}(x)\right]}_{=0 \text{ (similar proof as green)}} = 0$$

To see this, observe

$$\begin{aligned} \mathbb{E}\left[(x - \mathbb{E}(x|y))\mathbb{E}(x|y)\right] &= \mathbb{E}\left[\mathbb{E}\left((x - \mathbb{E}(x|y))\mathbb{E}(x|y) | y\right)\right] \quad \text{law of iterated expectation} \\ &= \mathbb{E}\left[\mathbb{E}\left((x - \mathbb{E}(x|y))g(y) | y\right)\right] \\ &= g(y)\mathbb{E}\left[(x - \underbrace{\mathbb{E}(x|y)}_{}) | y\right] \quad (\text{next page}) \end{aligned}$$

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

$$= g(Y) \left(\mathbb{E}[X|Y] - \mathbb{E}[g(Y)|Y] \right)$$

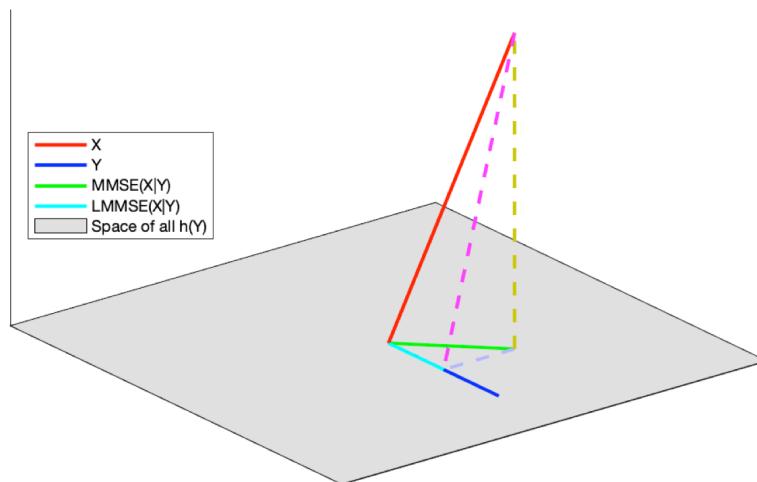
$$= g(Y) (g(Y) - g(Y)) = 0$$

Bottom Line: $X - \mathbb{E}(X|Y)$ is uncorrelated with $\mathbb{E}(X|Y)$

Visualizing this:

Random Variable Space

(Diagram Courtesy of Alex Cay)



(Sorry colors don't
match diagram
-read legend)

Think of Covariance \approx a dot product

Think of Uncorrelated \approx orthogonal

Taking $\mathbb{E}(X|Y) \approx$ orthogonally projecting X onto "space" of functions of Y

Further justification of this geometric world-view:

- orthogonal projections of X onto space of $h(Y)$ -functions should be the thing in that space "closest to X "

Standard notion of "closeness": mean-squared difference

Fact (Major): $\text{IE}(X|Y)$ is the function of Y that minimizes $\text{IE}((X-h(Y))^2)$ over ALL functions $h(Y)$

I.e.

$\text{IE}(X|Y)$ is the minimum mean-square estimator (MSE) of X given Y

Idea/Proof: Given ANY $h(Y)$

you'd be surprised how many proofs become easier by adding and subtracting "1"

$$\text{IE}[(X-h(Y))^2] = \text{IE}\left[((X-\text{IE}(X|Y)) + (\text{IE}(X|Y) - h(Y)))^2\right]$$

$$= \text{IE}[(X-\text{IE}(X|Y))^2] + 2 \underbrace{\text{IE}[(X-\text{IE}(X|Y))(\text{IE}(X|Y) - h(Y))]}_{=0} + \text{IE}[(\text{IE}(X|Y) - h(Y))^2]$$

$$\text{Middle part} = 2 \text{IE}\left[\text{E}(((X-\text{IE}(X|Y))(\text{IE}(X|Y) - h(Y))) | Y)\right]$$

$$= (\text{IE}(X|Y) - h(Y)) \text{IE}(X-\text{IE}(X|Y) | Y) \Rightarrow \text{middle term} = 0$$

Bottom Line: for any function $h(Y)$

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

"obvious" choice of $h(Y)$ to minimize LHS is $h(Y) = \text{IE}(X|Y)$

Next topic,

Conditional Variance

Given X, Y conditional variance of X given Y is the random variable

$$\text{Var}(X|Y) = \mathbb{E}((X - \mathbb{E}(X|Y))^2 | Y)$$

A recipe similar to the "g-thing" for computing $\text{Var}(X|Y)$

- Given y , compute

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

-Do this by finding conditional pmf $p_{X|Y}(x|y)$ or pdf $f_{X|Y}(x|y)$ and then computing variance of it

- This yields a function of y - $\gamma(y)$ - plug Y in for y ; that yields

$$\text{Var}(X|Y) = \gamma(Y)$$

Example - See This In Action

$$Y \sim \text{Uniform}[0, 1]$$

$X = \#\text{(Heads)}$ in n flips of coin w/ $P(\{\text{H}\}) = Y$

Saw last time

$$\mathbb{E}(X|Y=y) = ny \Rightarrow \mathbb{E}(X|Y) = nY$$

$$\begin{aligned}\text{Var}(X|Y=y) &= \text{sum of variances of } n\text{-independent Bernoulli}(y) \text{ trials} \\ &= ny(1-y)\end{aligned}$$

Thus

$$\text{Var}(X|Y) = \gamma(Y) = nY(1-Y)$$

$$\text{Since } X = \underbrace{(X - \mathbb{E}(X|Y))}_{\text{uncorrelated} \Rightarrow \text{variances add}} + \underbrace{\mathbb{E}(X|Y)}_{\dots \text{finish next time}}$$