Recall

- Covariance

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

  - X, Y uncorrelated $\Leftrightarrow$ $\mathrm{Cov}(X,Y) = 0$
  - X, Y uncorrelated $\Rightarrow$ $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

- "Conditional Expectation Revisited"

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

- For each Y, compute $\mathbb{E}(X \mid Y = y)$
  - this defines a function $g(y)$
- Define $\mathbb{E}(X \mid Y) = g(Y)$ $\xrightarrow{\text{i.e}}$ $Y \rightarrow \boxed{g} \rightarrow \mathbb{E}(X \mid Y)$

Did a few examples...

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

For any function $h(Y)$,

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

Can think of $\mathbb{E}(X \mid Y)$ as an <u>estimator of X</u> given Y.

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

- $\mathbb{E}(X \mid Y) = \mathbb{E}(X)$ by law of iterated expectations
- The estimation error $X - \mathbb{E}(X \mid Y)$ is uncorrelated w/ the estimate $\mathbb{E}(X \mid Y)$ – in fact, $X - \mathbb{E}(X \mid Y)$ is uncorrelated with Y – More generally, w/ any function $h(Y)$

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

To see this,

(1) $$\mathrm{Cov}\big(\underbrace{X-\mathbb{E}(X|Y)}_{\text{a rv-}A},\ \underbrace{\mathbb{E}(X|Y)}_{\text{a rv-}B}\big) = \mathbb{E}\Big[\big(A-\mathbb{E}(A)\big)\big(B-\mathbb{E}(B)\big)\Big]$$

$$= \mathbb{E}\Big[\big(X-\mathbb{E}(X|Y)-\underbrace{\mathbb{E}[X-\mathbb{E}(X|Y)]}_{0}\big)\big(\mathbb{E}(X|Y)-\mathbb{E}(\underbrace{\mathbb{E}(X|Y)}_{X})\big)\Big]$$

(2) $$= \mathbb{E}\Big[\big(\overbrace{\underbrace{X-\mathbb{E}(X|Y)}_{=\mathbb{E}(X)-\mathbb{E}(\mathbb{E}(X|Y))=0}}^{\text{has zero mean}}\big)\big(\mathbb{E}(X|Y)-\underbrace{\mathbb{E}(X)}_{\text{mean } \mathbb{E}(X)}\big)\Big]$$

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

(3) $$= \underbrace{\mathbb{E}\Big[\big(X-\mathbb{E}(X|Y)\big)\mathbb{E}(X|Y)\Big]}_{=0 \text{ (shown below)}} - \underbrace{\mathbb{E}\Big[\big(X-\mathbb{E}(X|Y)\big)\mathbb{E}(X)\Big]}_{=0 \text{ (similar proof as green)}} = 0$$

To see this, observe

$$\mathbb{E}\Big[\big(X-\mathbb{E}(X|Y)\big)\mathbb{E}(X|Y)\Big] = \mathbb{E}\Big[\mathbb{E}\Big(\big(X-\mathbb{E}(X|Y)\big)\underline{\mathbb{E}(X|Y)}\ \Big|\ Y\Big)\Big]$$

law of iterated expectation

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

$$= g(Y)\,\mathbb{E}\Big[\big(X-\underline{\mathbb{E}(X|Y)}\big)\ \Big|\ Y\Big]$$ (next page)

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

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

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

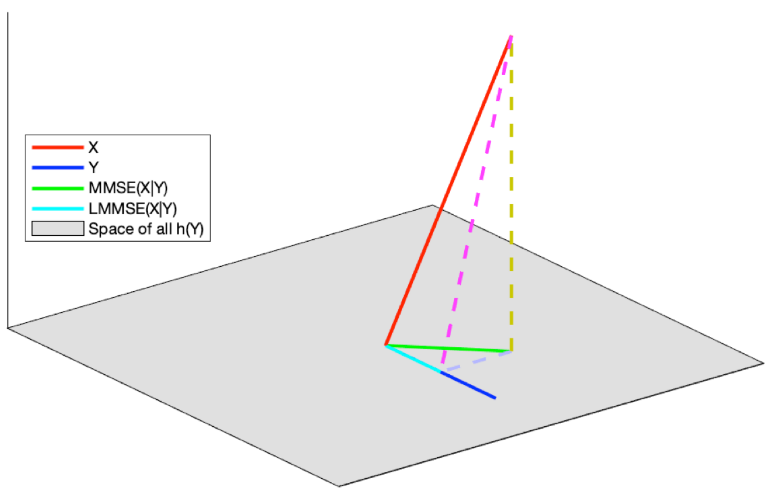
Bottom Line: $X - \mathbb{E}(X|Y)$ is <u>uncorrelated</u> with $\mathbb{E}(X|Y)$

Visualizing this:

Random Variable Space

(Diagram Courtesy of Alex Coy)

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



Think of Covariance ≈ a dot product

Think of Uncorrelated ≈ 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): $\mathbb{E}(X \mid Y)$ is the function of Y that minimizes $\mathbb{E}((X-h(Y))^2)$ over ALL functions $h(Y)$

I.e.

$\mathbb{E}(X \mid Y)$ is the minimum mean-square estimator (MMSE) of X given Y

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

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

$$\mathbb{E}\left[(X-h(Y))^2\right] = \mathbb{E}\left[\Big((X-\mathbb{E}(X\mid Y)) + (\mathbb{E}(X\mid Y)-h(Y))\Big)^2\right]$$

$$= \mathbb{E}\left[(X-\mathbb{E}(X\mid Y))^2\right] + \underbrace{2\,\mathbb{E}\left[(X-\mathbb{E}(X\mid Y))(\mathbb{E}(X\mid Y)-h(Y))\right]}_{=0} + \mathbb{E}\left[(\mathbb{E}(X\mid Y)-h(Y))^2\right]$$

Middle part $= 2\mathbb{E}\left[\underline{\mathbb{E}\Big(\big((X-\mathbb{E}(X\mid Y))(\mathbb{E}(X\mid Y)-h(Y))\big)\,\Big|\, Y}\Big)\right]$

$= \Big((\mathbb{E}(X\mid Y)-h(Y))\,\mathbb{E}(X-\mathbb{E}(X\mid Y))\mid Y\Big) \Rightarrow$ middle term $=0$

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

"obvious" choice of $h(Y)$ to minimize LHS is $h(Y)=\mathbb{E}(X\mid Y)$

$$\mathbb{E}\left[(X-h(Y))^2\right] = \mathbb{E}\left[(X-\mathbb{E}(X\mid Y))^2\right] + \mathbb{E}\left[(\mathbb{E}(X\mid Y)-h(Y))^2\right]$$

Next topic.

Conditional Variance

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

$$\mathrm{Var}(X|Y) = \mathbb{E}\left((X - \mathbb{E}(X|Y))^2 \,|\, Y\right)$$

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

- Given y, compute

$$\mathrm{Var}(X|Y=y) = \mathbb{E}\left[(X - \mathbb{E}(X|Y=y))^2 \,|\, Y=y\right]$$

  - 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

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

Example - See This In Action

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

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

Saw last time

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

$\mathrm{Var}(X|Y=y) =$ sum of variances of n-independent Bernoulli(y) trials

$= ny(1-y)$

Thus

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

Since $X = \overbrace{\overbrace{(X - \mathbb{E}(X|Y))}^{} + \overbrace{\mathbb{E}(X|Y)}^{}}^{\text{uncorrelated} \Rightarrow \text{variances add}}$

... finish next time