<u>Theory of Probability</u> <u>Nov 25, 2020</u>

<u>Conditional Expectation</u>

$$E[X \mid Y=y] = \int x \, f_{X|Y}(x|y) \, dx$$

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}$$

continuous

$$= \sum_i x_i \, p_{X|Y}(x_i|y)$$

$$= \sum_i x_i \, P[X=x_i \mid Y=y]$$

discrete

$$E[g(X) \mid Y=y] = \int g(x) \, f_{X|Y}(x|y) \, dx.$$

Denote by $\underbrace{E[X \mid Y] = g(Y)}_{\text{this is itself a random variable.}}$ (function of the random variable Y)

Whereas: $E[X \mid Y=y]$ is <u>not</u> a random variable.

$$\Rightarrow \quad E\big[E[X \mid Y]\big] = \int E[X \mid Y=y] \, f_Y(y) \, dy$$

$$= \int_y \int_x x \, f_{X|Y}(x|y) \, dx \; f_Y(y) \, dy$$

$$= \iint x \, f(x,y) \, dx \, dy = E[X]$$

this calculation shows that

$$\underbrace{E\left[\, E[X \mid Y] \,\right]}_{\text{iterated expectation}} = \underline{\underline{E[X]}}$$

Conditional Variance

$$\underbrace{\mathrm{Var}(X \mid Y)}_{\substack{\text{a random variable,} \\ \text{function of } Y}} = E\left[\left(X - E[X \mid Y]\right)^2 \mid Y \right]$$

$$= E\left[X^2 \mid Y\right] - \left(E[X \mid Y]\right)^2$$

$$\Rightarrow \quad E\left[\mathrm{Var}(X \mid Y)\right] = E\left[E\left[X^2 \mid Y\right] - \left(E[X \mid Y]\right)^2 \right]$$

$$= E[X^2] - E\left(E[X \mid Y]^2 \right) \qquad (*)$$

Also we know that $E\left(E[X \mid Y]\right) = E[X]$, so

$$\mathrm{Var}\left[E[X \mid Y]\right] = E\left[\left(E[X \mid Y]\right)^2\right] - \Big(\underbrace{E\left[E[X \mid Y]\right]}_{E[X].} \Big)^2$$

$$= E\left[\left(E[X \mid Y]\right)^2\right] - \left(E[X]\right)^2 \qquad (**)$$

$$(*) + (**) = E\left[\mathrm{Var}(X \mid Y)\right] + \mathrm{Var}\left[E(X \mid Y)\right]$$

$$= E[X^2] - \left(E[X]\right)^2 = \mathrm{Var}(X).$$

$\Rightarrow$ We have the following result:

$$\mathrm{Var}[X] = \mathrm{Var}(E[X|Y]) + E[\mathrm{Var}[X|Y]].$$

$\boxed{\text{Conditional Variance Formula}}$

Ex: Let $X_1, \dots, X_n$ be IID r.v., and

$N > 0$ is an integer valued random variable.

What is $\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right)$?

Conditional Expectation & Prediction

Consider two random variables X, Y.

Observe X, predict what value Y will take.

Let $g(x)$ be our predictor:

I.e. observe $X = x$, predict $y = g(x)$.

One choice for a "good" predictor is to minimize the mean square error:

$$\min_g \underbrace{E\left[(Y - g(X))^2\right]}_{\iint (y-g(x))^2 f(x,y)\, dx\, dy.}$$

Proposition: $E\left[(Y - g(X))^2\right] \geq E\left[(Y - E[Y|X])^2\right]$

Example the best linear predictor of Y

i.e., Find the best a, b to minimize

$$E[(Y - \underbrace{(a+bX)}_{g(X)})^2].$$

To do this: - expand this expression
- take $\frac{\partial}{\partial a}, \frac{\partial}{\partial b}$, set to zero
- solve.

The solution to this optimization problem is:

$$a = \mu_Y - \frac{\rho \sigma_Y \mu_X}{\sigma_X}$$

$$b = \rho \frac{\sigma_Y}{\sigma_X}$$

$$\Rightarrow g(x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

and using these vals of a, b, we can compute the mean squared error:

$$E[(Y - (a+bX))^2] = \sigma_Y^2 (1 - \rho^2)$$

this means that if $\rho = \pm 1$ then the mean squared error is zero.