Lecture 12 – continued

Example 4.7.6 (with additional calculations).

Recall that we have derived the distribution of Y, denoted by $f_2(y)$, in "Lectures 7-8 continued", pages 63-64:

$$f_2(y) = \begin{cases} -\ln(1-y), & \text{for } 0 < y < 1. \\ 0, & \text{otherwise}. \end{cases}$$

- A direct calculation of $E(Y)$ gives:

$$E(Y) = -\int_0^1 y \ln(1-y)\, dy \;\overline{\overline{}}$$

$1-y=t \Leftrightarrow y=1-t$

$dy=-dt$

$y=0 \Rightarrow t=1$

$y=1 \Rightarrow t=0$

$$= \int_1^0 (1-t)\ln t \, dt = -\int_0^1 (1-t)\ln t \, dt$$

$$= -\int_0^1 \ln t + \int_0^1 t \ln t \, dt = -\int_0^1 \ln t + \int_0^1 \left(\frac{t^2}{2}\right)' \ln t \, dt$$

$$= -\left[t \ln t \Big|_{t=0}^{t=1} - t\Big|_{t=0}^{1} \right] + \left\{ \frac{t^2}{2} \ln t \Big|_0^1 - \int_0^1 \left(\frac{t^2}{2} \cdot \frac{1}{t} \right) dt \right\}$$

↑ Using integration by parts

- $\lim\limits_{t \to 0} t \ln t = \lim\limits_{t \to 0} \dfrac{\ln t}{\frac{1}{t}} = \lim\limits_{t \to 0} \dfrac{\frac{1}{t}}{-\frac{1}{t^2}} = \lim\limits_{t \to 0} (-t) = 0$

- $\lim\limits_{t \to 0} \dfrac{t^2}{2} \ln t = \dfrac{1}{2} \lim\limits_{t \to 0} \dfrac{\ln t}{\frac{1}{t^2}} = -\dfrac{2}{2} \lim\limits_{t \to 0} \dfrac{\frac{1}{t}}{t^3} = 0$

<u>Therefore</u>: $E(Y) = 0 + 1 + 0 - \dfrac{1}{2} \int_0^1 t \, dt$

$$= 1 - \frac{1}{2} \cdot \frac{t^2}{2} \Big|_0^1 = 1 - \frac{1}{4} = \frac{3}{4}$$

- A <u>much simpler</u> calculation of $E(Y)$, using the fact that $E(Y) = E(E(Y|X))$:

- We have derived $g_2(y|x) = \begin{cases} \frac{1}{1-x}, & \text{for } x < y < 1 \\ 0, & \text{otherwise.} \end{cases}$

For each <u>fixed</u> $x \in (0,1)$ we can therefore calculate

$$E(Y|X=x) = \int_x^1 y \cdot \frac{1}{1-x} \cdot dy = \frac{1}{1-x} \cdot \frac{y^2}{2}\Big|_{y=x}^{y=1}$$

$$= \frac{1}{2} \cdot \frac{1-x^2}{1-x} = \frac{1}{2} \cdot \frac{\cancel{(1-x)}(1+x)}{\cancel{1-x}} = \frac{1+x}{2},$$

<u>Therefore</u>: $E(Y|X) = \frac{1+X}{2}$ and so:

$$\left.\begin{array}{l} E(Y) = E(E(Y|X)) = \frac{1}{2}E(1+X), \\ \text{. Recall that } X \sim \text{Uniform}[0,1], \text{ so:} \\ E(X) = \int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2} \end{array}\right\} \Rightarrow$$

$$E(Y) = \frac{1}{2}\left(1 + \frac{1}{2}\right) = \frac{1}{2} \cdot \frac{3}{2} = \frac{3}{4}$$ //

Predicting Y after we observe $X = x$.

Problem 4.9.25.

Assume $(X, Y) \sim f(x, y)$, where

$$f(x,y) = \begin{cases} 8xy, & \text{for } 0 < y < x < 1. \\ 0, & \text{otherwise.} \end{cases}$$

Suppose we observe $X = 0.2$. What predicted value of Y has the smallest MSE?

Solution The smallest MSE is given by $E(Y \mid X = 0.2)$.

- To calculate it we need to determine the conditional distribution of Y given that $X = 0.2$. This is given by the formula: $g_1(y \mid x = 0.2) = \dfrac{f(0.2, y)}{f_X(0.2)}$, where $f_X(0.2)$ is the marginal distribution of X evaluated at $x = 0.2$, and $0 < y < 0.2$.

- To find the marginal of X we integrate out y from $f(x,y)$:

$$f_X(x) = \int_0^x 8xy\,dy = 8x \cdot \frac{y^2}{2}\Big|_0^x$$

$$= 4x^3, \text{ for } 0 < x < 1.$$

<u>Therefore</u> :

$$g_1(y \mid x = 0.2) = \frac{\cancel{8}^{2} \cdot 0.2 \cdot y}{\cancel{4} \cdot (0.2)^3} = \frac{2y}{(0.2)^2} = 50y,$$

for $0 < y < 0.2$.

<u>Thus</u> $$E(Y \mid X = 0.2) = \int_0^{0.2} y \cdot 50y\,dy$$

$$= 50 \cdot \frac{y^3}{3}\Big|_0^{0.2}$$

$$= 50 \cdot \frac{(0.2)^3}{3} = \frac{2}{15}$$

//.

Summary regarding prediction

(I) One variable, call it Y

- The best prediction of Y with respect to the MSE is a number s.t.: $E(Y-c)^2$ is the smallest.

Result: The value c that minimizes $E(Y-c)^2$ over all c is $c = E(Y)$.

- The MSE of this prediction is therefore $E(Y - E(Y))^2 = \text{var}(Y)$. (1)

(II) Two variables, X and Y

(a) The best prediction of Y, given X, is the function of X that minimizes the MSE, $E(X-d(X))^2$, over all possible functions d.

Result: $E[Y - E(Y|X)]^2 \leq E[Y-d(X)]^2$, for all d.

Thus: $d(X) =: E(Y|X)$ is the best prediction of Y w.r.t. the MSE.

- The MSE of this prediction is therefore: $E[Y - E(Y|X)]^2$; this is typically called the overall MSE.

- To draw a comparison with ① above, recall that we defined:

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

Then $E(var(Y|X)) = E[Y - E(Y|X)]^2$ is the overall (averaged over all values of X) MSE of predicting Y by $E(Y|X)$.

(b). If $X = x$ has been observed, then:

(1) Best prediction of Y given $X = x$ is $E(Y|X = x)$; "best" is with respect to MSE

(2) The value of the MSE is

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

NOTE

- $var(Y|X)$ is a random variable
- $var(Y|X=x)$ is a number, depending on x.

(III) In order to assess whether knowing X helped us predict Y better, we need to compare the MSEs:

- No info on X: $MSE = var(Y)$ ✓
- Info on X: (overall) $MSE = E(var(Y|X))$.

Thus, the possible gain in prediction is measured by

$$var(Y) - E(var(Y|X)),$$

which should be positive, if we are to be making a smaller error by knowing X.

- Recall from Lecture 11, page 12, that $var(Y) = E(var(Y|X)) + var(E(Y|X))$. Therefore $var(Y) - E(var(Y|X)) = var(E(Y|X)) \geq 0$, since variances are always non-negative. <u>Therefore</u>: Info on a variable X connected to Y helps us predict Y better.