Lecture 16

$$\boxed{E(XY) \Rightarrow \text{also known as Correlation between } X \text{ and } Y.}$$

CORRELATION COEFFICIENT

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)\,Var(Y)}} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

Two important properties of ρ_{XY}

(1) $-1 \le \rho_{XY} \le 1$

(2) $-\sigma_X \sigma_Y \le Cov(X,Y) \le \sigma_X \sigma_Y$

(2) implies (1)
(1) implies (2)

Proof: Let σ_X^2, σ_Y^2 be the variances of X and Y.
Let $W = X - aY$ for some constant a.

$$Var(W) = E\left[(X-aY)^2\right] - \left(E[X-aY]\right)^2$$
$$= E\left[X^2 + a^2Y^2 - 2aXY\right] - (\mu_X - a\mu_Y)^2$$
$$= \left(E(X^2) - \mu_X^2\right) - 2a\left(E(XY) - \mu_X\mu_Y\right) + a^2\left(E(Y^2) - \mu_Y^2\right)$$

$$\Rightarrow Var(W) = Var(X) - 2a\,Cov(X,Y) + a^2\,Var(Y)$$

[Alternatively] $a^2V(Y) - 2a\,Cov(X,Y) + V(X) \geq 0$ for all $\underline{\underline{a}}$

$\Rightarrow$ This is a positive quadratic $\Rightarrow$ No real roots

2

$\Rightarrow$ Discriminant ≤ 0

$\Rightarrow (2\,cov(X,Y))^2 - 4\,V(X)V(Y) \leq 0$

$\Rightarrow C_{XY}^2 \leq V(X)\,V(Y).$

$\Rightarrow -\sqrt{\sigma_X^2\sigma_Y^2} \leq Cov(X,Y) \leq \sqrt{\sigma_X^2\sigma_Y^2}$

Proof continued....

We know that Variance ≥ 0

$$\Rightarrow \quad Var(W) \geq 0$$

$$\Rightarrow \quad Var(X) - 2a\,Cov(X,Y) + a^2\,Var(Y) \geq 0$$

$$\Rightarrow \quad 2a\,Cov(X,Y) \leq Var(X) + a^2\,Var(Y)$$

and this inequality holds for <u>any</u> constant a.

Choose $a = \dfrac{\sigma_X}{\sigma_Y}$

$$2\frac{\sigma_X}{\sigma_Y}\cdot Cov(X,Y) \leq \sigma_X^2 + \frac{\sigma_X^2}{\sigma_Y^2}\cdot\sigma_Y^2$$

$$\Rightarrow 2\frac{\sigma_X}{\sigma_Y}\,Cov(X,Y) \leq 2\sigma_X^2$$

$$\Rightarrow \underbrace{\frac{Cov(X,Y)}{\sigma_X\,\sigma_Y}}_{\Downarrow} \leq 1$$

$$\rho_{XY} \leq 1$$

Choose $a = -\dfrac{\sigma_X}{\sigma_Y}$

$$\Downarrow$$

$$-2\frac{\sigma_X}{\sigma_Y}\,Cov(X,Y) \leq 2\sigma_X^2$$

$$\Rightarrow -Cov(X,Y) \leq \sigma_X\sigma_Y$$

$$\Rightarrow Cov(X,Y) \geq -\sigma_X\sigma_Y$$

$$\Rightarrow \underbrace{\frac{Cov(X,Y)}{\sigma_X\sigma_Y}}_{\Downarrow} \geq -1$$

$$\rho_{XY} \geq -1.$$

Interpretation of ρ_{XY}:

ρ_{XY} describes the information we gain about Y by observing X. For example, a positive correlation coefficient $\rho_{XY} > 0$ suggests that when X is high w.r.t. its expected value, Y also tends to be high. When X is low, Y is likely to be low.

A negative correlation coeff. $\rho_{XY} < 0$ suggests that a high value of X is likely to be accompanied by a low value of Y & vice versa.

A Linear relationship between X and Y produces the extreme values, $\rho_{XY} = \pm 1$.

Do this Yourself $\Rightarrow$ If X and Y are such that $Y = aX + b$

$$\rho_{XY} = \begin{cases} -1, & \text{if } a < 0 \\ 0, & \text{if } a = 0 \\ +1, & \text{if } a > 0 \end{cases}$$

Eg:

$X \to$ height, $Y \to$ weight

$0 < \rho_{XY} < 1$

$X \to$ temp in Kelvin, $Y \to$ temp in Celcius

$\rho_{XY} = 1$

$X \to$ Telephone number, $Y \to$ Social Security Number

$\rho_{XY} = 0$

$X \to$ distance of a cell phone from a Tower

$Y \to$ Power of received signal

$-1 < \rho_{XY} < 0$

~~UNCORRELATED RANDOM VARIABLES~~

UNCORRELATED RANDOM VARIABLES

2 r.v.'s X and Y are uncorrelated if

$$C_{XY} = \mathrm{Cov}(X,Y) = 0 \quad (\text{Covariance} = 0)$$

or $\Updownarrow$

$$\rho_{XY} = 0 \quad (\text{Correlation Coefficient} = 0)$$

$\Updownarrow$

$$E(XY) = E(X)E(Y)$$

INDEPENDENT RANDOM VARIABLES

2 r.v.'s X and Y are independent if

$$\underbrace{f_{X,Y}(x,y)}_{\text{Joint}} = \underbrace{f_X(x) f_Y(y)}_{\text{Product of marginals.}}$$

ALWAYS TRUE

$$\boxed{\begin{array}{ccc} \text{If X and Y are independent} & \Rightarrow & \text{X and Y are also uncorrelated} \\ \text{If X and Y are uncorrelated} & \not\Rightarrow & \text{X and Y are independent.} \end{array}}$$

(NOT TRUE IN General)

Proof of Fact $\rightarrow$ $\boxed{\text{If } X, Y \text{ are independent} \Rightarrow X \text{ and } Y \text{ are uncorrelated.}}$

$\because$ X, Y are independent, $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$

$$\Rightarrow E(XY) = \int\int xy\, f_{X,Y}(x,y)\,dx\,dy$$

$$= \int\int x\,y\, f_X(x) f_Y(y)\,dx\,dy$$

$$= \underbrace{\left(\int_x x f_X(x)\,dx\right)}\underbrace{\left(\int y\, f_Y(y)\,dy\right)}$$

$$= E(X) \cdot E(Y)$$

$\Rightarrow$ $E(XY) = E(X)E(Y)$

$\Rightarrow$ $E(XY) - E(X)E(Y) = 0$

$\Rightarrow$ $Cov(X,Y) = 0$

$\Rightarrow$ X, Y are uncorrelated.

Example. Let $X \sim unif(0,1)$, $Y \sim unif(0,1)$
and $Z = X + Y$, $W = X - Y$.

Show that Z and W are uncorrelated
but they are NOT independent.

$$E(WZ) = E\big((X-Y)(X+Y)\big) = E(X^2 - Y^2) = E(X^2) - E(Y^2) = 0$$

$$E(W) = E(X) - E(Y) = 0$$

$$\Rightarrow Cov(W,Z) = \overbrace{E(WZ)}^{=0} - \overbrace{E(W)}^{=0}E(Z) = 0 \Rightarrow$$

W and Z are uncorrelated

To show W and Z are NOT independent,

find the joint PDF of (Z, W)

$f_{Z,W}(z,w)$

find the product of marginals

$f_Z(z) f_W(w)$

$$f_{Z,W}(z,w) \neq f_Z(z) f_W(w)$$

Show they are NOT ~~the~~ Equal

α

Conditioning by an EVENT

Conditional Joint PMF

For discrete random variables X and Y, an an event B, with $P(B) > 0$, the conditional joint PMF of X and Y given B is

$$P_{X,Y|B}(x,y) = \begin{cases} \dfrac{P_{X,Y}(x,y)}{P[B]} & \text{if } (x,y) \in B \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Joint PDF

Given an event B, with $P(B) > 0$, the conditional joint PDF of X and Y is

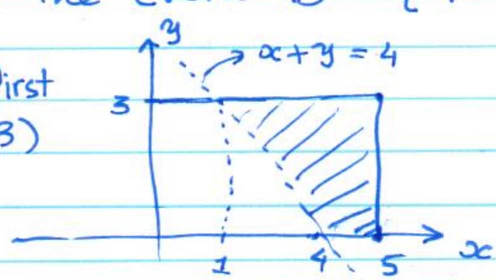
$$f_{X,Y|B}(x,y) = \begin{cases} \dfrac{f_{X,Y}(x,y)}{P(B)}, & \text{if } (x,y) \in B \\ 0, & \text{otherwise.} \end{cases}$$

Example: (X, Y) have the Joint PDF

$$f_{XY}(x,y) = \begin{cases} 1/15 & 0 \le x \le 5, 0 \le y \le 3 \\ 0, & \text{otherwise} \end{cases}$$

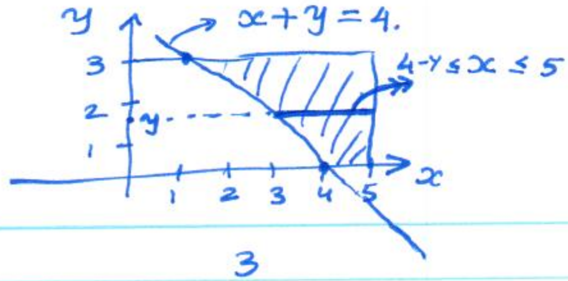
Find conditional joint PDF of (X, Y) given the event $B = \{X + Y \geq 4\}$.

Answer: We first find P(B)



$$P(B) = \int_{y=0}^{3} \int_{x=4-y}^{5} f_{X,Y}(x,y)\,dx\,dy$$

8



$$P(B) = P(X+Y \geq 4)$$

$$= \int_{y=0}^{3} \int_{x=4-y}^{5} \frac{1}{15}\, dx\, dy = \frac{1}{15} \int_{y=0}^{3} (1+y)\, dy$$

$$= \frac{1}{2}$$

$$\Rightarrow f_{X,Y|B}(x,y) = \begin{cases} \dfrac{2}{15} \left(\dfrac{f_{X,Y}(x,y)}{P[B]}\right) & 0 \le x \le 5,\ 0 \le y \le 3,\ x+y \geq 4 \\ 0 & \text{otherwise.} \end{cases}$$

Conditional Expected Value

For some function $W = g(X,Y)$ of r.v.'s X and, Y, and an event B, with $P(B) > 0$.

(if X, Y are discrete) $$E[W|B] = \sum_x \sum_y g(x,y) \cdot \underbrace{P_{X,Y|B}(x,y)}_{\text{conditional PMF.}}$$

(continuous) $$E[W|B] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y) \underbrace{f_{X,Y|B}}_{\text{conditional PDF.}}(x,y)\, dx\, dy$$

We can denote it as

$(\mu_{W|B}) \rightarrow$ Conditional expected value of the random variable W given the event B.

Conditional Variance

Conditional Variance of the random Variable $W = g(X,Y)$ given an event B is

$$\mathrm{Var}[W|B] = E\left[\left(W - \mu_{W|B} \right)^2 \,\middle|\, B \right]$$

Expand and show

$$= E[W^2|B] - (\mu_{W|B})^2$$

Example continued

In Previous Example, what is

$E[W|B]$, where $W = XY$, $B = \{X+Y \geq 4\}$

Recall, we already found $f_{X,Y|B}$ ie the Conditional PDF

$$\Rightarrow E[W|B] = E[XY|B]$$

$$= \int_{y=0}^{3} \int_{x=4-y}^{5} xy \underbrace{f_{X,Y|B}(x,y)}_{2/15} \, dx \, dy$$

(from previous page).

$$= \frac{123}{20}.$$ (Check yourself....)

CONDITIONING BY A RANDOM VARIABLE

Conditional PMF

For any event $Y = y$, such that $P_Y(y) > 0$,

Conditional PMF of X given $Y = y$ is

$$P_{X|Y}(x|y) = P[X = x \,|\, Y = y]$$
$$= \frac{P[X = x, Y = y]}{P[Y = y]}$$

$\Rightarrow$ the joint PMF can be written as

$$P_{X,Y}(x, y) = P_Y(y) \cdot P_{X|Y}(x|y).$$
$$= P_X(x) \cdot P_{Y|X}(y|x).$$

Conditional Expected Value

$$E[g(X,Y) \,|\, Y = y] = \sum_x g(x,y) \underbrace{P_{X|Y}(x|y)}_{\text{Conditional PMF.}}$$

as a special case,

$$E[X \mid Y=y] = \sum_{x} x \cdot P_{X|Y}(x|y).$$

Conditional PDF

For y such that $f_Y(y) > 0$, the conditional PDF of X given $\{Y=y\}$ is

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

This implies

$$\underbrace{f_{X,Y}(x,y)}_{\text{Joint}} = \underbrace{f_Y(y)}_{\text{marginal of } Y} \cdot \underbrace{f_{X|Y}(x|y)}_{\text{conditional of } X \text{ given } Y}$$

$$= \underbrace{f_X(x)}_{\text{marginal of } X} \cdot \underbrace{f_{Y|X}(y|x)}_{\text{conditional of } Y \text{ given } X}$$

$\Rightarrow$ X and Y are independent if

$$f_{Y|X}(y|x) = f_Y(y).$$

i.e. conditional of Y given X = marginal of Y

& $f_{X|Y}(x|y) = f_X(x)$, ...

V. Important!!

$\boxed{\text{Conditional Expected Value}}$

$$E[g(X,Y)\,|\,Y=y] = \int_{x=-\infty}^{\infty} g(x,y)\, \underbrace{f_{X|Y}(x|y)}_{\text{Conditional PDF}}\, dx.$$

$$E[X\,|\,Y=y] = \int_{x=-\infty}^{\infty} x\, f_{X|Y}(x|y)\, dx.$$

$E[X|Y=y]$ is a function of the observation y.

When we perform an experiment, and observe $Y=y$, $E[X|Y=y]$ is a function of the r.v. Y. We use the notation $E[X|Y]$ to denote this function of the random variable Y. Since a function of a random variable is also a r.v, hence $E[X|Y]$ is a random variable.

Example: $f_{X|Y}(x|y) = \begin{cases} \frac{1}{1-y} & y \le x \le 1. \\ 0 & \text{otherwise.} \end{cases}$

$$E[X|Y=y] = \int_{x=y}^{1} x \cdot \left(\frac{1}{1-y}\right) dx = \frac{1+y}{2}$$

and $E[X|Y] = \frac{1+Y}{2}$ → this is a r.v. too!!

(Iterated Expectation) **Theorem**

Recall: (this is a random variable. which takes the value $E[X|Y=y]$ when $Y=y$.)

$$\boxed{E\Big[\overbrace{E[X|Y]}\Big] = E[X]}$$

Proof:

$$E\Big[E[X|Y]\Big] = \int_{y=-\infty}^{\infty} E[X|Y=y]\cdot f_Y(y)\,dy$$

$$= \int_{y=-\infty}^{\infty}\left(\int_{x=-\infty}^{\infty} x\, f_{X|Y}(x|y)\,dx\right) f_Y(y)\,dy$$

$$= \int_{x=-\infty}^{\infty} x\left[\int_{y=-\infty}^{\infty} \underbrace{f_{X|Y}(x|y)\, f_Y(y)}_{f_{X,Y}(x,y)}\,dy\right] dx.$$

$$= \int_{x=-\infty}^{\infty} x\left[\underbrace{\int_{y=-\infty}^{\infty} f_{X,Y}(x,y)\,dy}_{\text{marginal of } X}\right] dx$$

$$= \int_{x=-\infty}^{\infty} x\, f_X(x)\,dx = E[X]$$

Similarly : $E\left[E[g(X) | Y] \right] = E[g(X)]$

Example: Consider jointly normal (X, Y)

$$f_{X,Y}(x,y) = \frac{\exp\left[- \dfrac{\left\{ \left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\frac{\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{y-\mu_2}{\sigma_2}\right)^2 \right\}}{2(1-\rho^2)} \right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$$

We can define

$$\tilde{\mu}_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

$$\tilde{\sigma}_2 = \sigma_2\sqrt{1-\rho^2}$$

and rewrite the joint PDF as

$$f_{X,Y}(x,y) = \underbrace{\frac{1}{\sqrt{2\pi}\,\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}}_{f_X(x)} \cdot \underbrace{\frac{1}{\sqrt{2\pi}\,\tilde{\sigma}_2} \cdot e^{-\frac{(y-\tilde{\mu}_2(x))^2}{2\tilde{\sigma}_2^2}}}_{f_{Y|X}(y|x)}$$

$\Rightarrow$ Conditional PDF of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi}\,\tilde{\sigma}_2} e^{-\frac{(y-\tilde{\mu}_2(x))^2}{2\tilde{\sigma}_2^2}}$$

$\hookrightarrow$ Gaussian with mean $\tilde{\mu}_2(x)$, and Variance $\tilde{\sigma}_2^2$.

$$\Rightarrow E[Y|X=x] = \mu_2(x) = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)$$

$$\Rightarrow \underbrace{E[Y|X]}_{\text{this is a random variable}} = \mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)$$

$$E[E[Y|X]] = E\left[\mu_2 + \rho\frac{\sigma_2}{\sigma_1}(X-\mu_1)\right]$$

$$= E[\mu_2] + 0 = \mu_2 = E[Y]$$

$$Var(Y|X=x) = \tilde{\sigma}_2^2 = \sigma_2^2(1-\rho^2) < \sigma_2^2$$

The above tells us that when $\rho \neq 0$, learning the value of X reduces the variance of Y, i.e reduces the uncertainty in Y.

For jointly Gaussian r.v.'s to be uncorrelated $\rho = 0$. However, if $\rho = 0$, then $\tilde{\sigma}_2^2 = \sigma_2^2$ and $\tilde{\mu}_2(x) = \mu_2$

$$\Rightarrow f_{Y|X}(y|x) = f_Y(y)$$

$$\Rightarrow f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

$\Rightarrow$ Uncorrelated $\Rightarrow$ Independence (for Jointly Gaussian)

Independence $\Rightarrow$ Uncorrelated (for any pair of r.v.'s)

Conditional Expectation and Minimum Mean Square Estimation

Conditional Expectetation plays an important role in the estimation of a random variable Y from another random variable X, when the optimality criterion is to minimize the mean squared value of the estimation error.

Suppose that we wish to estimate a random variable Y with a constant c. How should we choose this constant?

Error in estimation $= (Y-c)$

$\therefore$ Squared Estimation Error $= (Y-c)^2$

Mean Squared Estimation Error $= E(Y-c)^2$

So, we wish to find the constant c such that

$$MSE = E(Y-c)^2 \text{ is minimized.}$$

$$\Rightarrow \quad c^* = \arg\min E(Y-c)^2$$

$$MSE = E(Y-c)^2 = \int_{y=-\infty}^{\infty} (y-c)^2 f_Y(y)\,dy$$

$$\frac{d(MSE)}{dc} = \int_{y=-\infty}^{\infty} 2(y-c) f_Y(y)\,dy = 0$$

$$\Rightarrow \frac{d(MSE)}{dc} = \int_{y=-\infty}^{\infty} 2(y-c)\, f_Y(y)\, dy$$

Setting $\frac{d(MSE)}{dc} = 0$

$$\Rightarrow \int c\, f_Y(y)\, dy = \int y\, f_Y(y)\, dy$$

$$\Rightarrow c \underbrace{\int f_Y(y)\, dy}_{=1} = \underbrace{\int y\, f_Y(y)\, dy}_{E(Y)}$$

$$\Rightarrow \boxed{c = E(Y)}$$

this choice minimizes the MSE if the estimator is a constant.

Now, suppose that we wish to estimate Y not by a constant but by some function $c(X)$ of the random variable X. Hence, our goal is to minimize

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

What function $c(X)$ minimizes the MSE?

$$\boxed{c^*(X) = E[Y|X]}$$

minimizes the MSE

or the (Minimum Mean Squared Estimator)

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

$$= \int_x \int_y (y - c(x))^2 f_{X,Y}(x,y)\, dy\, dx$$

$$= \int_x \int_y (y - c(x))^2 f_X(x)\, f_{Y|X}(y|x)\, dy\, dx.$$

$$MSE = \int_{x=-\infty}^{\infty} f_X(x) \underbrace{\left[\int_{y=-\infty}^{\infty} (y - c(x))^2 f_{Y|X}(y|x)\, dy \right]}_{\text{this is positive for all } x.} dx$$

$\Rightarrow$ we can minimize

$$\int_{y=-\infty}^{\infty} (y - c(x))^2 f_{Y|X}(y|x)\, dy \quad \text{for every } x.$$

$\Rightarrow$ this is the Same minimization as before!!
(since for each fixed x, $c(x)$ is a constant).

$$\Rightarrow \quad c^*(x) = \int_{y=-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy = E\left[Y \mid X = x\right]$$

$$\Rightarrow \quad \boxed{c^*(X) = E\left[Y \mid X\right]}$$

this is the optimal MMSE estimator.