Lecture 31

4.7. Joint pdf of functions of random variables (not covered in the final)

For given f_{XY} of X and Y,

$$W = g_1(X,Y) \text{ and } Z = g_2(X,Y).$$

$\Rightarrow$ What is the joint pdf f_{WZ} of W and Z?

Consider three cases

1. g_1 & g_2 are linear.
2. $(g_1(u,v), g_2(u,v))$ is one-to-one mapping of (u,v)
3. $(g_1(u,v), g_2(u,v))$ is many-to-one mapping of (u,v)

4.7.1 Transformation under a linear mapping.

Say $W = aX + bY$, $Y = cX + dY$

$$\Rightarrow \begin{pmatrix} W \\ Z \end{pmatrix} = A \begin{pmatrix} X \\ Y \end{pmatrix} \quad \text{where } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Suppose (X,Y) in u-v plane and (W,Z) in α-β plain.

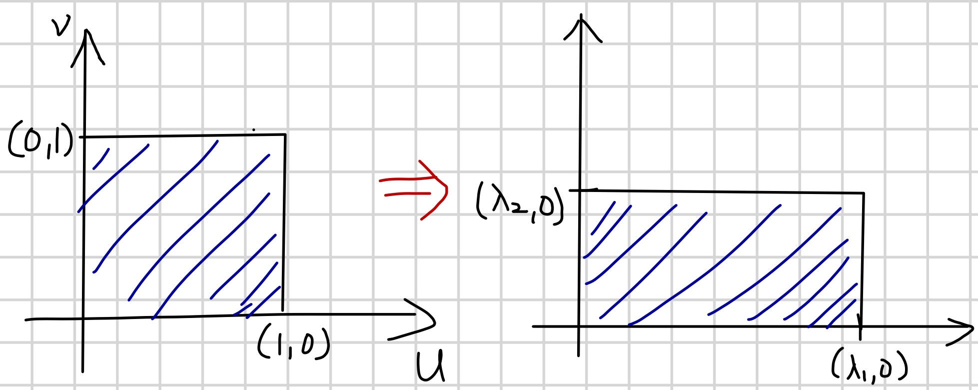
$\Rightarrow (W,Z)$ is the image of (X,Y) under the linear mapping

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = A \begin{pmatrix} u \\ v \end{pmatrix} \Leftrightarrow \begin{pmatrix} u \\ v \end{pmatrix} = A^{-1} \begin{pmatrix} u \\ v \end{pmatrix} \quad \text{(if } A \text{ is invertible)}$$

* Several linear transformations.

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix}$$

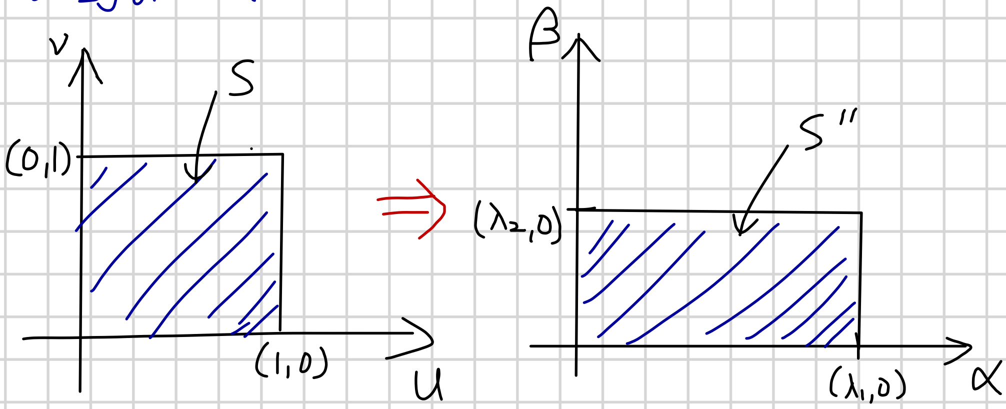
Area $1 \to \lambda_1 \lambda_2$



* Several linear transformations.

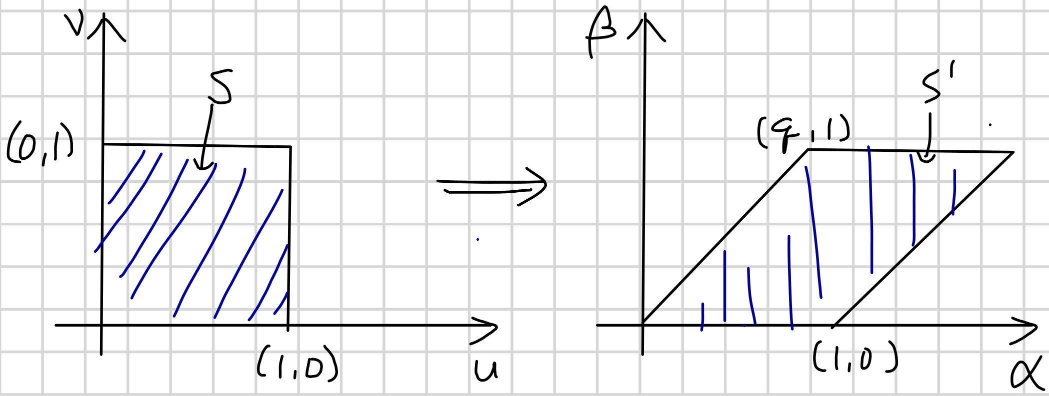
$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}$$

$\text{Area}(S)\cdot\lambda_1\lambda_2=\text{Area}(S')$



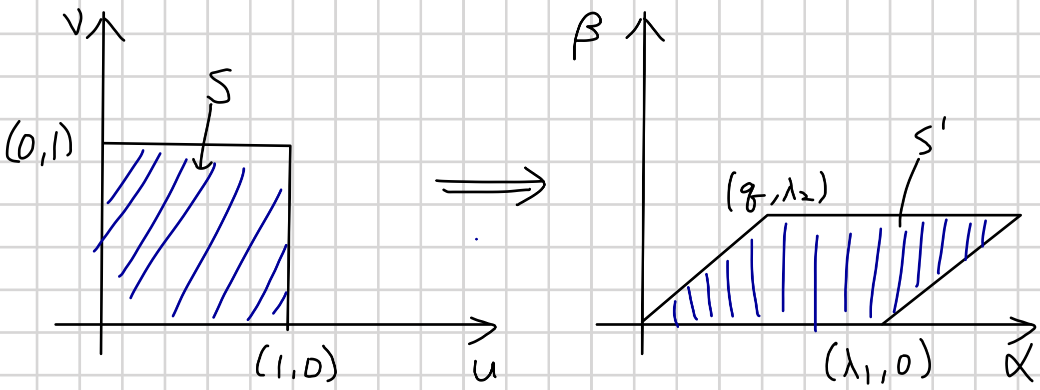
$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}1 & q\\ 0 & 1\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}$$

$\text{Area}(S)=\text{Area}(S')$



$$\begin{pmatrix}\alpha\\ \beta\end{pmatrix}=\begin{pmatrix}\lambda_1 & q\\ 0 & \lambda_2\end{pmatrix}\begin{pmatrix}u\\ v\end{pmatrix}$$

$\Rightarrow \text{Area}(S)\cdot\lambda_1\lambda_2=\text{Area}(S')$



For a general transformation $A=\begin{pmatrix}a & b\\ c & d\end{pmatrix}$

$$\begin{pmatrix}a & b\\ c & d\end{pmatrix}=\underbrace{\begin{pmatrix}1 & b/d\\ 0 & 1\end{pmatrix}}_{\text{No area change}}\underbrace{\begin{pmatrix}a-\frac{bc}{d} & 0\\ c & d\end{pmatrix}}_{\text{change area by } |ad-bc| = \det A}$$

* After transformation under A

$\text{Area}(S)\cdot|\det A|=\text{Area}(S')$

Proposition 4.7.1:

$$f_{W,Z}(\alpha,\beta) = \frac{1}{|\det A|} f_{XY}\left(A^{-1}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)$$

Recall that if $Y = aX \Rightarrow f_Y(v) = \frac{1}{|a|} f_X\left(\frac{v}{a}\right)$.

Ex) Let $W = X - Y$, $Z = X + Y$. Express f_{WZ} with f_{XY}.

$$\begin{pmatrix}W\\ Z\end{pmatrix} = A\begin{pmatrix}X\\ Y\end{pmatrix} \quad \text{where } A = \begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$$

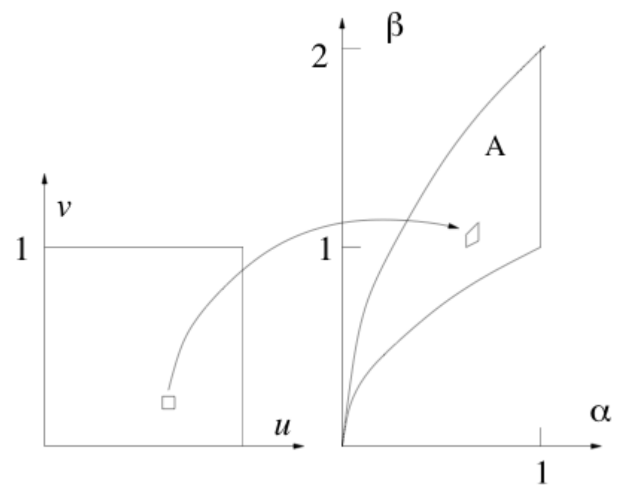
$$\Rightarrow \det A = 2, \quad A^{-1} = \frac{1}{2}\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$$

$$f_{WZ}(\alpha,\beta) = \frac{1}{|\det A|} f_{XY}\left(A^{-1}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right) = \frac{1}{2} f_{XY}\left(\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}\begin{pmatrix}\alpha\\ \beta\end{pmatrix}\right)$$

$$= \frac{1}{2} f_{XY}\left(\frac{\alpha+\beta}{2}, \frac{-\alpha+\beta}{2}\right)$$

4.7.2 Transformation under a one-to-one mapping

Suppose $\begin{pmatrix}W\\ Z\end{pmatrix} = g\left(\begin{pmatrix}X\\ Y\end{pmatrix}\right)$ where g is one-to-one mapping.



Region S in u-v plane $\xrightarrow{g}$ Region S' in α-β plain

$$\text{Area}(S) \cdot |\det J| = \text{Area}(S')$$

where $J = \begin{pmatrix} \frac{\partial g_1(u,v)}{\partial u} & \frac{\partial g_1(u,v)}{\partial v} \\ \frac{\partial g_2(u,v)}{\partial u} & \frac{\partial g_2(u,v)}{\partial v} \end{pmatrix}$

For a small square S, $(u,v) \in S$, and $(\alpha,\beta) = g((u,v))$,

$$P((X,Y) \in S) = P((W,Z) \in S')$$

$\Rightarrow f_{XY}(u,v) \cdot \text{Area}(S) \approx f_{WZ}(\alpha,\beta) \cdot \text{Area}(S')$, for $(u,v) \in S$, (α,β)

$\Rightarrow f_{WZ}(\alpha,\beta) = \frac{\text{Area}(S)}{\text{Area}(S')} f_{XY}(u,v)$

Proposition 4.7.4:

$$f_{WZ}(\alpha,\beta) = \frac{1}{|\det J|} f_{XY}\left(g^{-1}\begin{pmatrix}\alpha \\ \beta\end{pmatrix}\right) \text{ for the support of } f_{WZ}.$$

Ex) $f_{XY}(u,v) = \begin{cases} u+v & \text{if } (u,v) \in [0,1] \\ 0 & \text{else} \end{cases}$

Let $W = X^2$, $Z = X(1+Y)$. Find joint pdf f_{WZ}.

$\Rightarrow g_1(u,v) = u^2$, $g_2(u,v) = u(1+v)$. Let $\alpha = u^2$, $\beta = u(1+v)$.

$J = \begin{pmatrix} 2u & 0 \\ 1+v & u \end{pmatrix}$, $\quad \det J = 2u^2 = 2\alpha$

Let A is the image of $[0,1]^2$ in u-v plane after transformation g.

$\Rightarrow f_{WZ}(\alpha,\beta) = \frac{1}{2\alpha} f_{XY}\left(g^{-1}\begin{pmatrix}\alpha \\ \beta\end{pmatrix}\right) = \frac{1}{2\alpha} f_{XY}\left(\sqrt{\alpha}, \frac{\beta}{\sqrt{\alpha}} - 1\right)$ if $(\alpha,\beta) \in A$.

$= \frac{1}{2\alpha}\left(\sqrt{\alpha} + \frac{\beta}{\sqrt{\alpha}} - 1\right)$ if $(\alpha,\beta) \in A$

$f_{WZ}(\alpha,\beta) = 0$ else.

Lecture 33

Q) What is the variance of the joint distribution X and Y?

Three important metrics
* Correlation : $E[XY]$
* Covariance : $Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$
* Correlation coefficient $\rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

Covariance : the variance of two random variables X and Y jointly.

$Cov(X,Y) = 0 \Rightarrow$ X and Y are uncorrelated

$> 0 \Rightarrow$ positively correlated

$< 0 \Rightarrow$ negatively correlated.

"Uncorrelated" is the same as "independent"? No!

i) If X and Y are independent, they are correlated!

proof) $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} uv f_{XY}(u,v)\,du\,dv = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v\underbrace{u f_X(u)}_{\text{constant}} f_Y(v)\,du\,dv$

$= \int_{-\infty}^{\infty} u f_X(u) \Big[\underbrace{\int_{-\infty}^{\infty} v f_Y(v)\,dv}_{\text{constant}}\Big]\,du$

$= \int_{-\infty}^{\infty} u f_X(u)\,du \int_{-\infty}^{\infty} v f_Y(v)\,dv = E[X]\,E[Y]$

Thus, $E[XY] = E[X]E[Y]$ if X & Y are independent

$Cov(X,Y) = E[\underbrace{(X-\mu_X)(Y-\mu_Y)}_{\text{independent}}] = E[X-\mu_X]\,E[Y-\mu_Y] = 0$ "uncorrelated"

if X and Y are independent

ii) If X and Y are uncorrelated, X and Y could be dependent!

Counter example)

$$X = \begin{cases} 1 & \text{w.p. } \frac{1}{2} \\ -1 & \text{w.p. } \frac{1}{2} \end{cases}$$

If $X = 1$, $Y = 0$

If $X = -1$, Y is either 1 or -1 with prob. $\frac{1}{2}$ for each.

$\Rightarrow$ X and Y are dependent

$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E[XY] \quad (\because \mu_X = 0, \mu_Y = 0) \\ &= P_{XY}(1,0) \times 0 + \underbrace{P_{XY}(-1,1)}_{1/4} \cdot (-1)(1) + \underbrace{P_{XY}(-1,-1)}_{1/4}(-1)(-1) \\ &= 0 \end{aligned}$$

$$\boxed{\text{Thus, uncorrelated} \underset{\Longleftarrow}{\not\Longrightarrow} \text{Independent}}$$

Properties of Covariance

i) $Cov(X+Y, U+V) = Cov(XU) + Cov(XV) + Cov(YU) + Cov(YV)$

ii) $Cov(aX+b, cY+d) = ac\,Cov(X,Y)$

proof)
$$\begin{aligned} Cov(aX+b, cX+d) &= E[(aX+b-(a\mu_X+b))(cY+d-(c\mu_Y+d))] \\ &= E[(aX-a\mu_X)(cY-c\mu_Y)] = acE[(X-\mu_X)(Y-\mu_Y)] \\ &= ac\,Cov(X,Y) \end{aligned}$$

iii) $Cov(X,X) = E[(X-\mu_X)(X-\mu_X)] = E[(X-\mu_X)^2] = Var(X)$

iv) $Cov(X,Y) = E[XY] - E[X]E[Y]$

Why?
$$\begin{aligned} Cov(X,Y) &= E[(X-\mu_X)(Y-\mu_Y)] = E[X(Y-\mu_Y)] - \overset{0}{\cancel{E[\mu_X(Y-\mu_Y)]}} \\ &= E[XY] - E[X\mu_Y] = E[XY] - \mu_X\mu_Y. \end{aligned}$$

v) If $X_1, X_2, \cdots, X_n$ are pairwise uncorrelated, i.e., $Cov(X_i, X_j) = 0$ if $i \neq j$

Let $S_n = X_1 + X_2 + \cdots + X_n$. Then

$$Var(S_n) = \sum_{i=1}^{n} Var(X_i)$$

vb) If $X_1, \cdots, X_n$ are independent $\Rightarrow$ pairwise uncorrelated

$$\Rightarrow Var(S_n) = \sum_{i=1}^{n} Var(X_i)$$

proof of v)

$$Var(S_n) \overset{\because (iii)}{=} Cov(S_n, S_n) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} X_j\right)$$

$$\text{by (i)} = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) = \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \underbrace{Cov(X_i, X_j)}_{=0 \text{ since uncorrelated}} + \sum_{i=1}^{n} Cov(X_i, X_i)$$

$$= \sum_{i=1}^{n} Cov(X_i, X_i) = \sum_{i=1}^{n} Var(X_i)$$

Ex) Mean & Variance of $X \sim Binomial(n, p)$

Binomial (n,p) = Sum of n independent Bernoulli r.v.s (say $X_1, X_2, \cdots, X_n$) with parameter p.

Note that $E[X_i] = p$, $Var(X_i) = p(1-p)$ for $i = 1, 2, \cdots, n$.

$$E[X] = \sum_{i=1}^{n} E[X_i] = n E[X_i] = np$$

$$Var(X) = \sum_{i=1}^{n} Var(X_i) = n \cdot Var(X_i) = np(1-p).$$

EX) Express $Cov(X+2, 10Y-3X)$ with $Var(X)$, $Var(Y)$, or $Cov(X,Y)$

$$= Cov(X+2, 10Y-3X) = Cov(X, 10Y) + Cov(X, -3X)$$

$$= 10\,Cov(X,Y) - 3\,Cov(X,X) = 10\,Cov(X,Y) - 3\,Var(X).$$

Correlation Coefficient ρ_{XY}

Suppose X : Temperature in °C
Y : # of ice creams sold.
Unit of $Cov(X,Y)$: $(°C) \times (\text{unit})$

$X \times \frac{9}{5} + 32$: temperature in F
$10Y$: Sales of 3-dollar ice creams
Unit of $Cov(\frac{9}{5}X+32, 10Y)$: $(F) \times (\$)$
$Cov(\frac{9}{5}X+32, 10Y) = 18\, Cov(X,Y)$

Covariance is different depending on unit!

To consider unit-free covariance, we define

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

Eg) if $X \to \frac{9}{5}X+32$, then $\sigma_X \to \frac{9}{5}\sigma_X$
$Y \to 10Y$, then $\sigma_Y \to 10\sigma_Y$
$\Rightarrow$ correlation coefficient remains same.

$$\Rightarrow \rho_{XY} = \rho_{\frac{9}{5}X+32,\, 10Y} = \rho_{aX+b,\, cY+d} \text{ for } a, c > 0.$$

Properties of ρ_{XY}

i) $-1 \le \rho_{XY} \le 1$
ii) $\rho_{XY} = 1 \iff Y = aX+b$ for some a, b with $a > 0$.
iii) $\rho_{XY} = -1 \iff Y = aX+b$ for some a, b with $a < 0$.

Ex) Roll a die n times.

$$X_i = \begin{cases} 1 & \text{if one shows in the } i\text{-th roll} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_i = \begin{cases} 1 & \text{if two shows in the } i\text{-th roll} \\ 0 & \text{otherwise} \end{cases}$$

Let $X = \sum_{i=1}^{n} X_i$, $Y = \sum_{i=1}^{n} Y_i$

(a) $E[X_1]$, $Var(X_1)$? $E[X_1] = \frac{1}{6}$ $Var(X_1) = \frac{1}{6}\frac{5}{6} = \frac{5}{36}$

(b) $E[X]$, $Var(X)$? $E[X] = nE[X_1] = \frac{n}{6}$, $Var(X) = nVar(X_1) = \frac{5n}{36}$

(c) Find $Cov(X_i, Y_j)$

If $i \neq j$, X_i and Y_j are independent $\Rightarrow$ uncorrelated $\Rightarrow Cov(X_i, Y_j) = 0$

If $i = j$,

$p_{X_iY_i}(1,0) = \frac{1}{6}$, $p_{X_iY_i}(0,1) = \frac{1}{6}$, $p_{X_iY_i}(0,0) = \frac{4}{6}$, $p_{X_iY_i}(u,v) = 0$ else

$\Rightarrow Cov(X_i, Y_i) = E[X_iY_i] - E[X_i]E[Y_i] = 0 - \frac{1}{6}\frac{1}{6} = -\frac{1}{36}$ (negatively correlated)

(d) $Cov(X,Y)$?

$$Cov(X,Y) = Cov\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, Y_j)$$

$$= \sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n} \underbrace{\cancel{Cov(X_i, Y_j)}}_{0} + \sum_{i=1}^{n} \underbrace{Cov(X_i, Y_i)}_{= -\frac{1}{36}} = -\frac{n}{36}$$

(e) ρ_{XY}?

$\sigma_X^2 = Var(X) = \frac{5n}{36}$, $\sigma_Y^2 = Var(Y) = \frac{5n}{36}$

$$\Rightarrow \rho_{XY} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{-\frac{n}{36}}{\frac{5n}{36}} = -\frac{1}{5}$$

Recall

* Maximum Likelihood (ML) parameter estimation

a R.V. X is known to follow pdf $f_\theta(u)$ with unknown θ.
For given observation $X=j$, find

$$\hat{\theta}_{ML} = \arg\max_{\theta} f_\theta(j)$$

The parameter that maximizes the likelihood of $X=j$.

* ML Hypothesis testing

Two hypothesis H_0 & H_1

$$X \sim \begin{cases} f_1(u) & \text{if } H_1 \text{ is true.} \\ f_0(u) & \text{if } H_0 \text{ is true.} \end{cases}$$

For given $X=j$, estimate underlying hypothesis H_{i^*}

$$\Rightarrow \quad i^* = \arg\max_{i \in \{0,1\}} f_i(j)$$

* Minimum Mean Square Error Estimation.

1) Constant Estimator

i) $f_Y(u)$ is known, but do not know Y.

ii) Estimate Y by a constant δ.

Error $e = Y - \delta$.

constant

Q) Among possible choices of δ, which will minimize the mean square error?

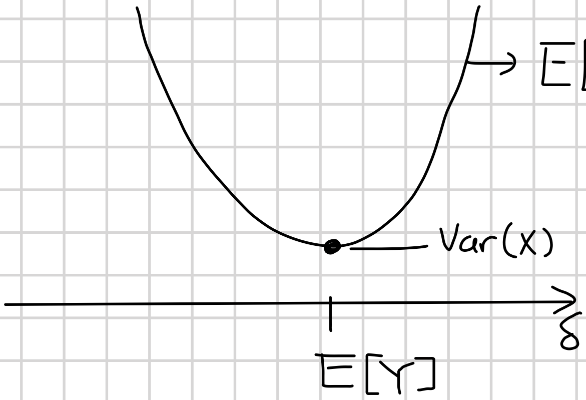
In other words, find δ^* s.t.

$$\delta^* = \arg\min_{\delta} E[(Y-\delta)^2]$$

Ans) $E[(Y-\delta)^2] = E[Y^2] - 2\delta E[Y] + \delta^2 - E[Y]^2 + E[Y]^2$

$$= E[Y^2] - E[Y]^2 + (\delta - E[Y])^2$$

$$= Var(Y) + (\delta - E[Y])^2$$



→ $E[(Y-\delta)^2]$ is minimized at $\delta^* = E[Y]$.

Thus, the minimum mean square error estimator of Y is $\delta^* = E[Y]$, and the minimum mean square error is $Var(Y)$.

2) Unconstrained Estimators

i) Don't know X and Y, but know f_{XY}.

ii) Observe X only and estimate Y with the observation.

⇒ Estimation of Y must be a function of observation X. Say $g(X)$.

Error $e = Y - g(X)$.

Among possible choices of $g(\cdot)$, which choice will minimize the mean square error?

i.e., find g^* s.t. $g^* = \arg\min_g E[(Y - g(X))^2]$.

Ans) Suppose $X = 10$. ⇒ pdf of Y conditioned on $X = 10$: $f_{Y|X}(v|10)$

⇒ We know the distribution of Y

⇒ Best estimation of $Y = E[Y \mid X = 10]$ (By Constant Est.)

For observation $X = u$,

$\therefore g^*(u) = E[Y \mid X = u] = \int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,dv$

MSE cond. on $X=u$ → $MSE(u) = Var(Y \mid X = u)$

MSE for random observation X (MSE before you observe X but if you will use $g^*(X)$ for any X to be observed)

MSE? $MSE = E[(Y - g^*(X))^2]$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty}(v - g^*(u))^2 f_{XY}(u,v)\,dv\right)du$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} v^2 + g^{*2}(u) - 2v\,g(u) f_{XY}(u,v)\,dv\right)du$$

$$= \int_{-\infty}^{\infty} v^2 f_{XY}(u,v)\,dv + g^*(u)^2 \underbrace{\int_{-\infty}^{\infty} f_{XY}(u,v)\,dv}_{f_X(u)} - 2g(u)\int_{-\infty}^{\infty} v f_{XY}(u,v)\,dv$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} v^2 f_{XY}(u,v)\,dv\,du + \int_{-\infty}^{\infty} g^{*2}(u) f_X(u)\,du + \int_{-\infty}^{\infty} 2g^*(u)\int_{-\infty}^{\infty} v f_{XY}(u,v)\,dv\,du$$

In the last term, $\int_{-\infty}^{\infty} v f_{XY}(u,v)\,dv = \int_{-\infty}^{\infty} v f_{Y|X}(v|u)\,du\, f_X(u)$

$= E[Y \mid X=u] f_X(u) = g^*(u) f_X(u)$.

$$= E[Y^2] + E[g^*(X)^2] - 2\int_{-\infty}^{\infty} g^{*2}(u) f_X(u)\,du$$

$$= E[Y^2] - E[g^*(X)^2] = E[Y^2] - E[E(Y|X)^2]$$

$$= E[Y^2] - E[Y]^2 + E[Y]^2 - E[E(Y|X)^2] = Var(X) + E[E[Y|X]]^2 - E[E[Y|X]^2]$$

$$= Var(Y^2) - Var(\underbrace{E[Y|X]}_{\text{Random Variable}})$$

Both are the same

Quick Quiz!

X, Y ~ Random

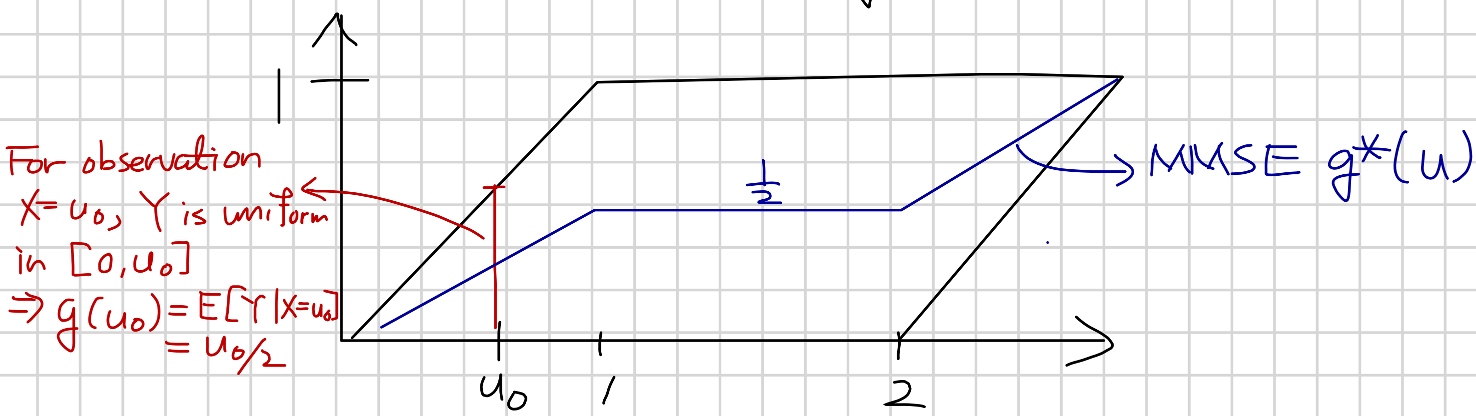
$E[X], E[Y], E[XY], E[\frac{X}{Y}]$ ~ all deterministic

$E[Y|X=u]$: Deterministic value affected by u

$\Rightarrow$ function of u.

$E[Y|X]$: Function of X $\Rightarrow$ Random Variable.

Ex) (X,Y) is uniform over the region below



Find and Sketch $g^*(u)$ for observation $X=u$

$$g^*(u) = \begin{cases} \frac{1}{2}u & \text{if } u\in[0,1] \\ \frac{1}{2} & \text{if } u\in[1,2] \\ \frac{1}{2}u-\frac{1}{2} & \text{if } u\in[2,3] \\ 0 & \text{else} \end{cases}$$

Find MSE for observation $X=u$.

Condition on $X=u$

Y is uniform over $\begin{cases} [0,u] & \text{if } u\in[0,1] \\ [0,1] & \text{if } u\in[1,2] \\ [u-2,1] & \text{if } u\in[3,4] \end{cases}$

Thus, $MSE(u) = Var(Y|X=u) = \begin{cases} \frac{1}{12}u^2 & \text{if } u\in[0,1] \\ \frac{1}{12} & \text{if } u\in[1,2] \\ \frac{1}{12}(3-u)^2 & \text{if } u\in[2,3] \end{cases}$

For observation X, MSE?

$$MSE = E[MSE(X)] = \int_{-\infty}^{\infty} MSE(u) f_X(u)\, du$$

$$f_X(u) = \begin{cases} \frac{1}{2}u & \text{if } u \in [0,1] \\ \frac{1}{2} & \text{if } u \in [1,2] \\ \frac{1}{2}(3-u) & \text{if } u \in [2,3] \\ 0 & \text{else} \end{cases}$$

$$\Rightarrow MSE = \int_0^1 \frac{u^2}{12} \frac{u}{2} du + \int_1^2 \frac{1}{12} \cdot \frac{1}{2} du + \int_2^3 \frac{1}{12}(3-u)^2 \frac{1}{2}(3-u) du$$

$$= \int_0^1 \frac{u^3}{24} du + \int_1^2 \frac{1}{24} du + \underbrace{\int_2^3 \frac{1}{24}(3-u)^3 du}_{= \int_0^1 \frac{u^3}{24} du} = 2\int_0^1 \frac{u^3}{24} du + \frac{1}{24}$$

$$= \frac{1}{3} + \frac{1}{24} = \frac{9}{24} = \frac{3}{8}$$

Ex) $X \sim N(0, \sigma^2)$, $W \sim Exp(1)$, X and Y are independent

$Y = 3X^2 + W^4$

For given observation $X = u$, the estimator $g^*(u)$ of Y?

$$\Rightarrow g^*(u) = E[Y \mid X=u] = E[3X^2 + W^4 \mid X=u] = E[3u^2 + W^4 \mid X=u]$$

$$= 3u^2 + \underbrace{E[W^4 \mid X=u]}_{\text{independent}} = 3u^2 + E[W^4]$$

$$= 3u^2 + \underbrace{\int_0^\infty v^4 e^{-v} dv}_{4!} = 3u^2 + 24$$

Note that $\int_0^\infty \frac{u^k e^{-u}}{k!} du = -\frac{e^{-u} \cdot u^k}{k!}\Big|_0^\infty + \int_0^\infty \frac{e^{-u} \cdot k \cdot u^{k-1}}{k!} du$

$$= \int_0^\infty \frac{u^{k-1} e^{-u}}{(k-1)!} du = \cdots = \int_0^1 e^{-u} du = 1$$

$$MMSE = Var(Y \mid X=u) = Var(3X^2 + W^4 \mid X=u)$$

$$= Var(3u^2 + W^4 \mid X=u) = Var(W^4)$$

$$= E[W^8] - E[W^4]^2 = 8! - (4!)^2$$

3) Linear MMSE

What if $E[Y|X=u]$ is hard to find in a closed form?

$\Rightarrow f_{Y|X}(v|u)$ is complecate or its integral is not easy.

We use an approximation $aX+b$.

How to choose a and b to minimize MSE ?

$$MSE = E[(Y-(aX+b))^2]$$
$$= E[(\underbrace{(Y-aX)}_{\text{random variable}} - \underbrace{b}_{\text{constant estimator of } Y-aX})^2]$$

$$\Rightarrow b^* = E[Y-aX] = E[Y] - aE[X].$$

$\Rightarrow$ optimal estimator : $aX+b^* = aX + E[Y] - aE[X] = a(X-E(X)) + E[Y]$
$$\triangleq L(x)$$

$$MSE = Var(Y-aX) = Cov(Y-aX, Y-aX)$$
$$= Var(Y) + a^2 Var(X) - 2a\,Cov(X,Y)$$
$$\frac{d\,Var(Y-aX)}{da} = 2a\,Var(X) - 2Cov(X,Y) = 0$$
$$\Rightarrow a = \frac{Cov(X,Y)}{Var(X)}$$

$\Rightarrow$ Optimal Linear Estimator
$$L^*(X) = \mu_Y + \frac{Cov(X,Y)}{Var(X)}(X-\mu_X)$$

or
$$\frac{L^*(X)-\mu_Y}{\sigma_Y} = \rho_{XY}\left(\frac{X-\mu_X}{\sigma_X}\right)$$

$$MSE = \sigma_Y^2 - 2\frac{Cov(X,Y)}{Var(X)}Cov(X,Y) + \left(\frac{Cov(X,Y)}{Var(X)}\right)^2 Var(X) = \sigma_Y^2 - \frac{Cov(X,Y)^2}{Var(X)}\left(2 - 1\right)$$
$$= \sigma_Y^2 - \rho_{XY}^2 \times \sigma_X^2 \times \sigma_Y^2 / \sigma_X^2 = \sigma_Y^2(1-\rho_{XY}^2) = Var(Y)(1-\rho_{XY}^2)$$

Comparison of MMSE. (smaller is better)

$$\underset{\text{MSE for } \delta^* = E[Y]}{Var(Y)} \overset{①}{\geq} \underset{\text{MSE for } L^*(X)}{Var(Y) - \frac{Cov^2(X,Y)}{Var(X)}} \overset{②}{\geq} \underset{\text{MSE for } g^*(X)}{Var(Y) - Var(E[Y|X])}$$

Estimation w/o observation X ⇐|⇒ Estimation with observation X (Additional information provides better estimation)

Estimation function is restricted to a linear (or a constant) function of X ⇐|⇒ Estimation function is not restricted

⇒ Freedom of the estimation function provides better estimation.

⇒ If $L^*(X)$ is a constant function, then ① becomes "="
Why? δ^* is the best constant estimator ⇒ No constant estimator can reduces MSE lower than $Var(Y)$ ⇒ MSE for $L^*(X)$ cannot be lower than $Var(Y)$

Q) When $L^*(X)$ becomes a constant estimator?

$L^*(X) = E[Y] - \frac{Cov(X,Y)}{\sigma_X \sigma_Y}(X - E[X]) \Rightarrow Cov(X,Y) = 0 \Leftrightarrow L^*(X)$ is constant

⇒ If $g^*(X)$ is a linear function, then ② becomes "="
Why? $L^*(X)$ is the best linear estimator ⇒ No linear estimator cannot reduce MSE lower than $Var(Y) - \frac{Cov^2(X,Y)}{Var(X)}$.
Thus, MSE for $g^*(X)$ is no smaller than $Var(Y) - \frac{Cov(X,Y)^2}{Var(X)}$.