Chapter 6: Joint Distributions

Now we will discuss the joint probability distributions of two or more random variables.

joint pmf

x & y are logically related like say height & weight etc.

Discrete Case

Joint probability mass function of X and Y

$$p(x,y) = P(X = x, Y = y)$$

$0 \le p(x,y) \le 1$ for all x, y ↪ $p(X = x \cap Y = y)$

Both can be discrete, both cont or one discrete & one continuous

1) $\sum_x \sum_y p(x,y) = 1$

The support of X and Y $\quad S_{X,Y} = \{\, (x,y):\ \ p(x,y) > 0 \,\}$

Example 1: A grocery store has two types of checkout lines, regular and self. Let X be the number of customers in the regular checkout at a certain time of day, and Y be the number of customers in the self-checkout at the same time of day. The joint pmf of X and Y is given in the table on the right.

		Y			
		0	1	2	3
X	0	0.08	0.07	0.04	0.00
	1	0.06	0.15	0.05	0.04
	2	0.05	0.04	0.10	0.06
	3	0.00	0.03	0.04	0.07
	4	0.00	0.01	0.05	0.06

a) What is the probability that the lines have equal numbers of customers?
b) Find the probability that there are at least 2 more customers in one line than in the other.
c) Find the probability that the total number of customers in the two lines is exactly four.

a) $P(X=Y) = p(0,0) + p(1,1) + p(2,2) + p(3,3)$

$= 0.08 + 0.15 + 0.10 + 0.07$

$= \boxed{0.40}$

x	y
0	0
1	1
2	2
3	3

b) $P(|X-Y| \ge 2) = p(0,2) + p(0,3) + p(1,3) + \cdots p(4,2)$

$= 0.04 + 0 + \cdots + 0.05$

$= \boxed{0.22}$

x	y
0	2
0	3
1	3
2	0
3	0
3	1
4	0

c) $P(X+Y=4) = p(1,3) + p(2,2) + p(3,1) + p(4,0)$

$= \boxed{0.17}$

Additional) If at least 4 ⇒ $P(X+Y \geq 4) = \ldots = \boxed{0.56}$

4 1
4 2

2)

Marginal probability mass function of X

$p_X(x) = \sum_y p(x,y)$

$P(X=x) = \sum_y P(X=x, Y=y)$ (over all values of y)

$P(X \in A)$, $E(X)$, $var(X)$, $E(g(X))$

Marginal probability mass function of Y

$p_Y(y) = \sum_x p(x,y)$

$P(Y=y) = \sum_x P(X=x, Y=y)$

The random variables X and Y are independent if and only if

$\boxed{P(x,y) = P_X(x) \cdot P_Y(y) \quad \text{for all } x, y}$ ⇒ Joint = Marg X · Marg Y

$\boxed{P(X=x, Y=y) = P(X=x) \cdot P(Y=y)}$

Example 1 (cont.): The joint pmf of X and Y is given in the table.

d) What is the probability that there are 2 customers in the regular checkout line? Same for the self-checkout line?

e) What is the expected number of customers in the self-checkout line?

f) Are X and Y independent?

		Y (SC)				$P_X(x)$
		0	1	2	3	
X (RC)	0	0.08	0.07	0.04	0.00	0.19
	1	0.06	0.15	0.05	0.04	0.30
	2	0.05	0.04	0.10	0.06	0.25
	3	0.00	0.03	0.04	0.07	0.14
	4	0.00	0.01	0.05	0.06	0.12
$P_Y(y)$		0.19	0.30	0.28	0.23	

d) $P(X=2)$ ⇒ Question just about X

$= P_X(2) = \sum_y P(X=2, Y=y) = P(2,0)+P(2,1)+P(2,2)+P(2,3) = \boxed{0.25}$

$P(Y=2) = P_Y(2) = \sum_x P(X=x, 2) = \boxed{0.28}$

e) $E(Y) = \sum_y y \cdot P_Y(y) = 0 \cdot (0.19) + 1(0.30) + 2 \cdot (0.28) + 3(0.23) = \boxed{1.55}$

f) Just present one cell which violates, thats enough

$P(0,3) = 0$, but is it $= P_X(0) \cdot P_Y(3) = ?$

$= (0.19) \cdot (0.23)$

∴ X & Y are not independent $\neq \boxed{0.0437}$

Recall the conditional probability of ~~[illegible]~~ A given B

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \qquad \text{provided that } P(B) > 0$$

The conditional probability mass function of X, given $Y = y$ is

3)

$$p_{X|Y}(x|y) = \frac{P(x,y)}{P_y(y)} \iff \underbrace{P(X=x}_{A} \mid \underbrace{Y=y)}_{B} = \frac{P(X=x, Y=y)}{P(Y=y)}$$

The conditional probability mass function of Y, given $X = x$ is

$$p_{Y|X}(y|x) = \frac{P(x,y)}{P_x(x)} \iff P(\underbrace{Y=y}_{A} \mid \underbrace{X=x}_{B}) = \frac{P(X=x, Y=y)}{P(X=x)}$$

$\boxed{\text{A given B}}$

4)

⊛ Conditional pmf properties to be satisfied

i) $0 \le P_{X|Y}(x|y) \le 1$ for all x & y

ii) $\sum_x P_{X|Y}(x|y) = 1$

5)

The conditional probability

$$P(X \in A|Y = y) = \sum_{x \in A} p_{X|Y}(x|y)$$

The conditional mean of X, given $Y = y$

$$E(X|Y = y) = \mu_{X|Y=y} = \sum_x x\, p_{X|Y}(x|y) = \sum x \cdot \frac{P(x,y)}{P_y(y)}$$

The conditional variance of X, given $Y = y$

$$Var(X|Y = y) = \sigma^2_{X|Y=y} = E(X^2|Y = y) - [E(X|Y = y)]^2$$

where $E(X^2|Y = y) = \sum_x x^2\, p_{X|Y}(x|y)$

Each of these concept is conditional

⊛ $Var(X) = E(X^2) - (E(X))^2$

Example 1 (cont.): Find

g) the distribution of the number of customers in the self-checkout line when there 3 customers in the regular checkout line. y X=3

h) the mean number of customers in the self-checkout line when there 3 customers in the regular checkout line.

X=3 is given

g) $P_{Y|X}(y|3) = \frac{P(3,y)}{P_X(3)} = \frac{P(3,y)}{0.14}$

y	~~0~~	1	2	3
$P_{Y\|X}(y\|3)$	$\cancel{\frac{0}{0.14} = 0}$	$\frac{0.03}{0.14} = 0.21$	$\frac{0.04}{0.14} = 0.29$	$\frac{0.07}{0.14} = 0.5$

not needed ∵ not possible

h) $E(Y|X=3) = \sum_y y \cdot P_{Y|X}(y|3) = 0 + 1\cdot(0.21) + 2\cdot(0.29) + 3\cdot(0.5) = \boxed{2.29}$

We don't round!!

Extra

i) $E(Y^2|X=3) = \sum_y y^2 \cdot P_{Y|X}(y|3) = 1^2\cdot(0.21) + 2^2\cdot(0.29) + 3^2(0.5) = \boxed{5.87}$

Conditional second moment

j) Conditional variance $= Var(Y|X=3) = 5.87 - (2.29)^2 = \boxed{0.626}$

k) $\sigma_{Y|X=3} = \sqrt{0.626} = \boxed{0.79}$

 X & Y are independent iff $\boxed{P(x,y) = P_X(x) \cdot P_Y(y)}$ for all x & y

Proposition: Discrete random variables X and Y are independent if and only if

1) • $p_{X|Y}(x|y)$ doesn't depend on y
2) • $p_{X|Y}(x|y) = p_X(x)$
3) • $p_{Y|X}(y|x)$ doesn't depend on x
4) • $p_{Y|X}(y|x) = p_Y(y)$

for all x & y

5) • Definition approach $P(x,y) = P_X(x) \cdot P_Y(y)$

Using (4)

y	0	1	2	3
$P_{Y\|X}(y\|3)$	0	0.21	0.29	0.5
$P_y(y)$	0.19	0.30	0.28	0.23

∴ X & Y are not independent

Using (3)

y	0	1	2	3
$P_{Y\|X}(y\|3)$	0	0.21	0.29	0.5
$P_{Y\|X}(y\|0)$	0.42	0.37	0.21	0

6) ✱ Changing x from $3 \to 0$ should have given the same values but in this case they are different ∴ not independent

X and Y are discrete random variables with the joint pmf $p(x, y)$

7) Expectation of a function of X and Y

$$\boxed{E(g(X,Y)) = \sum_x \sum_y g(x,y) \cdot p(x,y)}$$

Q: Find the expected absolute difference b/w the number of customers in the two lines

$$E(\underbrace{|X-Y|}_{=g(X,Y)}) = E(g(X,Y)) = \sum_x \sum_y |x-y| \cdot p(x,y)$$

$\Rightarrow |0-0|(0.08) + |0-1|(0.07) + |0-2|(0.04) + |0-3|(0)$

$+ |1-0|(\ldots) + \ldots$

$+ |2-0|(\ldots) \ldots$

$+ |3-0|(\ldots) \ldots$

$+ |4-0|(\ldots) \ldots \quad = \boxed{0.83}$

$E(Y) = \sum_y y \cdot p_Y(y) = \boxed{1.55}$ OR

$$E(\underbrace{Y}_{=g(X,Y)}) = E(g(X,Y)) = \sum_x \sum_y y \cdot p(x,y) =$$

Marginal expectation of Y

$(0)(0.08) + (1)(0.07) + (2)(0.04) + (3)(0) \ldots$

$+ (0)(0.06) + \ldots$

$+ \ldots \quad = \boxed{1.55}$

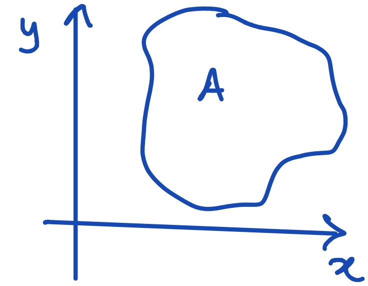
Continuous Case

The joint probability density function of continuous random variables X and Y is

$f(x,y) \geq 0$ for all x, y ⇒ always non negative

such that for any two-dimensional set A

$$\boxed{P(\,(X,Y) \in A\,) = \iint_A f(x,y)\,dx\,dy}$$



$f(x,y)$ is a valid joint pdf if and only iff

- $f(x,y) \geq 0$ for any x, y
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1$ $\boxed{\text{OR}}$ $dy\,dx = 1$

⊛ f is density so doesn't have to ≤ 1, $\int f$ is probability so that has to be ≤ 1

The support of continuous random variables X and Y

$$S_{X,Y} = \{\,(x,y): \quad f(x,y) > 0\,\}$$

Draw the support on X-Y plane

8)

Marginal probability density function of X

⇒ we integrate wrt other variable & get rid of it

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$$ and provide the support S_X

Marginal probability density function of Y

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$ and provide the support S_Y

⇒ use whichever is easier dy dx or dx dy

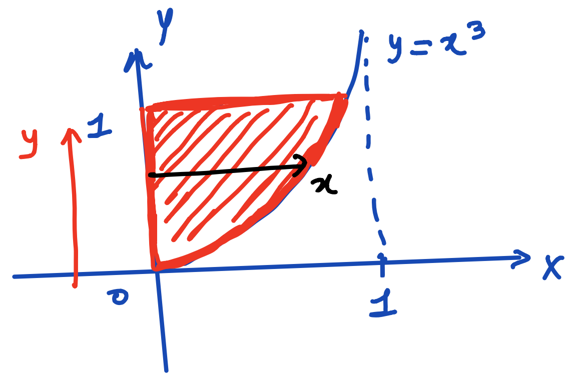
Expectation of a function of X and Y

$$E\big(g(X,Y)\big) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y)\, f(x,y)\, dx\, dy$$

Example 2: Let X and Y have the joint pdf

$$f(x,y) = c \quad \text{for } 0 < x < 1, \quad x^3 < y < 1 \Rightarrow S_{X,Y}$$

a) Find the constant c such that $f(x,y)$ is a valid joint pdf.
b) Find $P(X > Y)$.
c) Find the marginal distributions of X and Y.
d) Find the mean of X.



a) $$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = 1$$

$$= \int_0^1 \int_0^{y^{1/3}} c\,dx\,dy = 1$$

$$= c\int_0^1 \left(x\Big|_0^{y^{1/3}} \right) dy = 1 \qquad \Rightarrow \boxed{c = \frac{4}{3}}$$

$$= c\int_0^1 y^{1/3}\,dy = 1$$

$$= \frac{c\,y^{4/3}}{4/3}\Bigg|_0^1 = 1$$

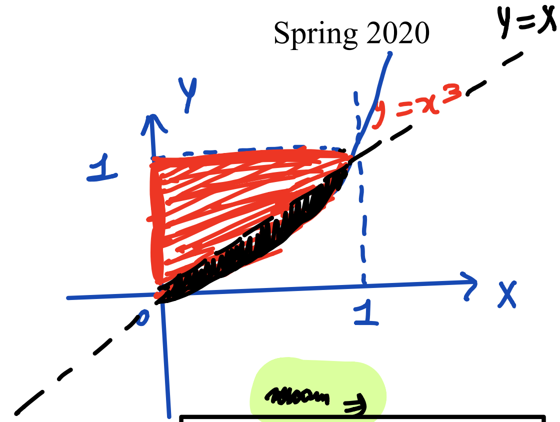
$$= \boxed{\frac{3}{4}c} = 1$$

$$\circledast \quad \int_0^1\int_{x^3}^1 c\,dy\,dx = 1 \quad = \quad c\int_0^1\int_{x^3}^1 dy\,dx = c\int_0^1 (1-x^3)\,dx =$$

$$c\left[x - \frac{x^4}{4}\right]_0^1 = c\left(1 - \frac{1}{4}\right) = \frac{3c}{4} = 1$$

$$\Rightarrow \boxed{c = \frac{4}{3}}$$

b) $$P(X>Y) = \iint_{\{X>Y\}} f(x,y)\,dx\,dy$$



zoom ⇒

$0<x<1,\ x^3<y<1,\ x>y$

$$\boxed{I = \int_0^1\int_y^{y^{1/3}} \frac{4}{3}\,dx\,dy = \int_0^1\int_{x^3}^{x} \frac{4}{3}\,dy\,dx}$$

$$= I = \frac{4}{3}\int_0^1 \left(y^{\frac{1}{3}} - y\right)dy = \frac{4}{3}\left[\frac{y^{\frac{4}{3}}}{\frac{4}{3}} - \frac{y^2}{2}\right]_0^1 =$$

$$= \frac{4}{3}\left[\left(\frac{3}{4} - \frac{1}{2}\right) - 0\right] = \frac{4}{3}\left(\frac{3-2}{4}\right) = \boxed{\frac{1}{3}}$$

Two continuous random variables X and Y are independent if and only if

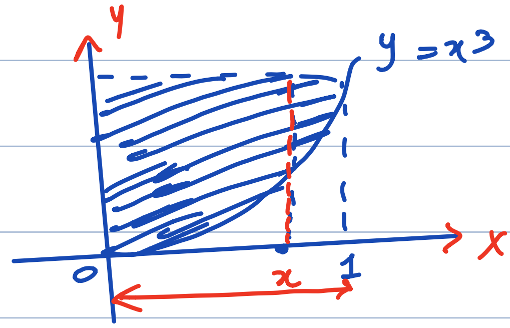
Already seen the conditions before

$$\boxed{f(x,y) = f_X(x)\,f_Y(y) \quad \text{for all } x,y} \in S_{X,Y}$$

c) $f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

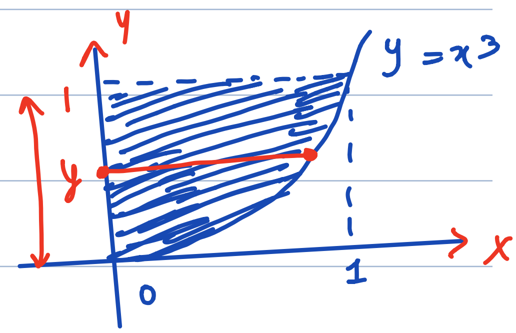
$= \int_{x^3}^{1} \frac{4}{3}\,dy = \frac{4}{3}\,y\Big|_{x^3}^{1}$

$= \frac{4}{3}(1-x^3)$, $0<x<1 = S_X$



⊛ $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,dx = \int_0^{y^{1/3}} \frac{4}{3}\,dx = \frac{4}{3}\,x\Big|_0^{y^{1/3}} = \frac{4y^{1/3}}{3}$ for $0<y<1 = S_Y$

9) Calculation of marginal is only over shaded support



---

d) $E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx = \int_0^1 x \cdot \frac{4}{3}(1-x^3)\,dx =$

$\frac{4}{3}\left(\frac{x^2}{2} - \frac{x^5}{5}\right)\Big|_0^1 = \frac{2}{5} = \boxed{0.4}$

// marginal

10) OR

$\underbrace{E(X)}_{=g(X,Y)} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x \cdot f(x,y)\,dx\,dy$

Calculation of E is over complete support

// joint

$= \int_0^1 \int_0^{y^{1/3}} x \cdot \frac{4}{3}\,dx\,dy = \frac{4}{3}\int_0^1 \left(\frac{x^2}{2}\right)\Big|_0^{y^{1/3}} dy = \boxed{0.4}$

e) $Var(X) = \boxed{0.062}$ $\Rightarrow$ find second moment first and solve

f) $E(Y) = \boxed{0.571}$ $= \int_0^1 y \cdot \frac{4}{3} y^{\frac{1}{3}} dy$

g) $Var(Y) = \boxed{0.074}$ $\int_0^1 y^2 \cdot \frac{4}{3} y^{\frac{1}{3}} dy - \left(\int_0^1 y \cdot \frac{4}{3} y^{\frac{1}{3}} dy \right)^2$

solve for these at home

9)

Example 2 (cont.): Let X and Y have the joint pdf

$$f(x,y) = c \qquad \text{for } 0 < x < 1, \quad x^3 < y < 1$$

Are X and Y independent?

$\iint f(x,y) =$

$\left(\int f_X(x)\,dy\right)$

$\left(\int f_Y(y)\,dx\right)$

$f(x,y) = \frac{y}{3}$

$0 < x < 1$

$x^3 < y < 1$

$\overset{?}{=}$

$\frac{y}{3}(1-x^3) \Rightarrow$ Marginal density of X

$0 < x < 1$

$\frac{y}{3}\, y^{1/3} \Rightarrow$ marginal density of Y

$0 < y < 1$

$$\boxed{\text{So,}\quad \frac{y}{3} \overset{?}{\neq} \frac{y}{3}(1-x^3)\cdot\frac{y}{3}(y^{1/3})}$$

(Also see that the support should match)

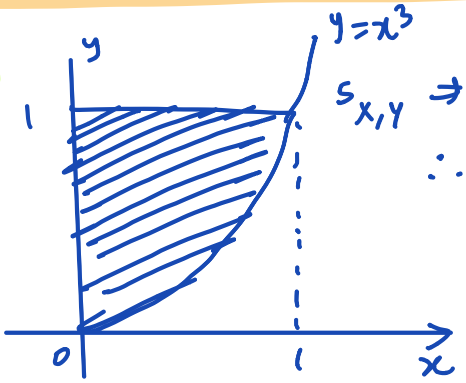
11)

⊛ Note: X, Y – independent $\Rightarrow S_{X,Y}$ has to be rectangular

But if the joint support is rectangular $\Rightarrow$ not necessary indep.

⊛ If the support is not rectangular $\Rightarrow X-Y$ not independent

Example



$S_{X,Y} \Rightarrow$ not rectangular

$\therefore$ X & Y are not indep. ✓

h) find $E\left(\frac{X}{\sqrt{Y}}\right)$ ⇒ part of example 2 only

$\frac{X}{\sqrt{Y}} \Longrightarrow g(X,Y)$

$$E\left(\frac{X}{\sqrt{Y}}\right) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{x}{\sqrt{y}}\, f(x,y)\, dx\, dy$$

⇒ ∵ we can't use the other methods we use for $E(X)$ or $E(Y)$

$$= \int_0^1 \int_0^{y^{1/3}} \frac{x}{\sqrt{y}}\, \frac{y}{3}\, dx\, dy$$

$$= \int_0^1 \frac{y}{3\sqrt{y}} \left(\frac{x^2}{2} \Big|_0^{y^{1/3}} \right) dy = \frac{y}{3} \int_0^1 \frac{1}{\sqrt{y}} \times \frac{y^{2/3}}{2}\, dy$$

$$= \frac{2}{3} \int_0^1 y^{\frac{2}{3} - \frac{1}{2}}\, dy = \frac{2}{3}\, \frac{y^{7/6}}{7/6} \Bigg|_0^1 = \frac{2}{3} \times \frac{6}{7} = \boxed{\frac{4}{7}}$$

The conditional probability density function of X, given $Y = y$ is

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

$\int_{-\infty}^{\infty} f(x,y)\,dx \leftarrow f_Y(y)$

The conditional probability density function of Y, given $X = x$ is

(2)

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)}$$ & provide $S_{X=x|Y}$

$\hookrightarrow$ as usual $\int_{-\infty}^{\infty} f(x,y)\,dy$

- $f_{X|Y}(x|y) \geq 0$ for any x, y
- $\int_{-\infty}^{\infty} f_{X|Y}(x|y)\,dx = 1$ for any y

The conditional probability

$$P(X \in A | Y = y) = \int_A f_{X|Y}(x|y)\,dx$$

The conditional mean of X, given $Y = y$

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

The conditional variance of X, given $Y = y$

$$Var(X|Y = y) = \sigma^2_{X|Y=y} = E(X^2|Y = y) - [E(X|Y = y)]^2$$

where $E(X^2|Y = y) = \int_{-\infty}^{\infty} x^2\, f_{X|Y}(x|y)\,dx$

Proposition: Continuous random variables X and Y are independent if and only if

- $f_{X|Y}(x|y)$ doesn't depend on y
- $f_{X|Y}(x|y) = f_X(x)$
- $f_{Y|X}(y|x)$ doesn't depend on x
- $f_{Y|X}(y|x) = f_Y(y)$

all of these for all x, y

i) Example 2 cont $\Rightarrow$ Find the conditional mean of Y given $X = \frac{1}{2}$

x

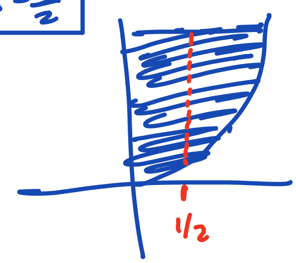
$$E(Y \mid X = \tfrac{1}{2}) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid 1/2)\, dy$$

$$f_{Y|X}(y \mid \tfrac{1}{2}) = \frac{f(1/2, y)}{f_X(\frac{1}{2})} = \frac{4/3}{\frac{4}{3}\left(1 - (\frac{1}{2})^3\right)} = \boxed{\frac{8}{7}}$$

$$\boxed{\text{Support} \Rightarrow \quad \tfrac{1}{8} < y < 1 \quad \text{so,} \quad S_{Y|X=\frac{1}{2}}}$$

so, $E(Y \mid X = \frac{1}{2}) =$

$$\int_{1/8}^{1} y \cdot \frac{8}{7}\, dy = \frac{8}{7} \left.\frac{y^2}{2}\right|_{1/8}^{1} = \boxed{\frac{9}{16}}$$



---

j) Find conditional mean of y given $X = x$

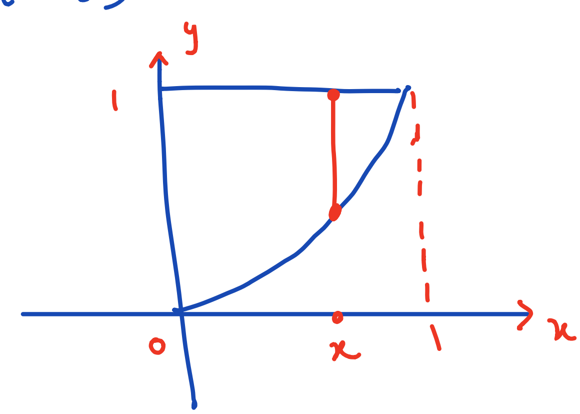
$$E(Y \mid X = x) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y \mid x)\, dy$$

$$f_{Y|X}(y \mid x) = \frac{f(x,y)}{f_X(x)} = \frac{4/3}{4/3\,(1 - x^3)}$$

$$\boxed{x^3 < y < 1, \quad 0 < x < 1}$$

$\Downarrow$

$S_{Y|X=x}$



$$E(Y \mid X = x) = \int_{x^3}^{1} y \cdot \left(\frac{1}{1 - x^3}\right) dy = \frac{1}{1 - x^3} \left.\frac{y^2}{2}\right|_{x^3}^{1} = \frac{1 - x^6}{2(1 - x^3)} = \boxed{\frac{1 + x^3}{2}}$$

$$\boxed{0 < x < 1}$$

* we found $f_{Y|X}(y|x) = \dfrac{1}{1-x^3}$ $\quad \overset{a}{x^3} < y < \overset{b}{1}$, $0 < x < 1$

$= \dfrac{1}{b-a}$ } → uniform distribution form
⇓
won't happen always though

So,

$$\boxed{Y \mid X = x \sim \text{Unif}\,(x^3, 1) \qquad 0 < x < 1}$$

So from concepts of uniform ⇒ conditional exp $= \dfrac{a+b}{2}$

So, $$\boxed{E(Y|X=x) = \frac{1+x^3}{2}}$$

⇒ Plug $x = \frac{1}{2}$ and you'll again get $\frac{9}{16}$ ✓

---

* $f_{Y|X}(y|x) = \dfrac{1}{1-x^3}$ $\quad x^3 < y < 1$, $0 < x < 1$

⇒ X & Y are not independent

Tips

* Marginal things for Y will only have y & no x ↗ ∵ $\int f(x,y)\,dy$
* Marginal things for X will only have x & no y
* Conditionals and joint can have both