2-D. Random Variable

Let $\mathcal{E}$ be an expt. Let S be a SS associated with $\mathcal{E}$. Let

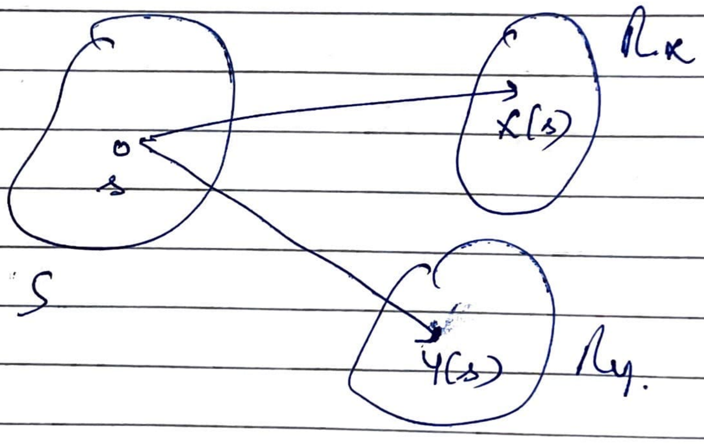
$X = X(s)$ &

$Y = Y(s)$ be two functions

$X : S \to R$ &

$Y : S \to R$.

Then (X, Y) is a 2D R.V. or a random vector



Discrete Case

(i) $P(X = x_i, Y = y_j) = p(x_i; y_j) \geq 0 \; \forall (x_i; y_j)$

(ii) $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} p(x_i; y_j) = 1.$

Cont. Case ~~Joint~~ PDF f

(i) $f(x, y) \geq 0 \quad \forall (x, y)$

(ii) $\iint_{R_Y R_X} f(x, y)\, dx\, dy = 1$ &

(iii) $P(a \le X \le b,\ c \le Y \le d)$

$$= \int_c^d \int_a^b f(x,y)\, dx\, dy$$

$f(x,y)$ is called joint pdf.

Ex: Two production lines manufacture a certain type of item.

Capacity line I → 5 items
line II → 3 items

Actual items produced is a r.v (2D) (X, Y) summarized as prob. dis. as

Y \ X	0	1	2	3	4	5
	0	.01	.03	.05	.07	.09
0	.00	.02				
1	.01					
2	.01					
3	.01					

Table A (6.1)

Y \ X	0	1	2	3	4	5
0	0	.01	.03	.05	.07	.09
1	.01	.02	.04	.05	.07	.08
2	.01	.03	.05	.05	.05	.06
3	.01	.02	.04	.06	.06	.05

~~Ex~~ $p(2,3) = .04$ etc

let $B = \{$ line I produces more items than line II $\}$

find $P(B)$

Sol^n $\cancel{P(B) = p(0,1) + p(0,2}$

$$\boxed{\begin{aligned} P(B) = {} & p(1,0) \\ & + p(2,1) + p(2,0) \\ & + p(3,2) + p(3,1) + p(3,0) \\ & + p(4,3) + p(4,2) + p(4,1) + p(4,0) \\ & + p(5,4) + p(5,3) + p(5,2) + p(5,1) + p(5,0) \\ = {} & .01 + .03 + .04 + .05 + .05 + .05 \\ & + .07 + .06 + .05 + .06 \\ & + \cancel{.09 + .08 + .06 + .05} \end{aligned}}$$

$$\begin{aligned} P(B) = {} & p(1,0) \\ & + p(2,0) + p(2,1) \\ & + p(3,0) + p(3,1) + p(3,2) \\ & + p(4,0) + p(4,1) + p(4,2) + p(4,3) \\ & + p(5,0) + p(5,1) + p(5,2) + p(5,3) + \end{aligned}$$

$$\begin{aligned} = {} & .01 \\ & + .03 + .04 \\ & + .05 + .05 + .05 \\ & + .07 + .06 + .05 + .06 \\ & + .09 + .08 + .06 + .05 \end{aligned}$$

$$= 0.75$$

Ex) Suppose

$$f(x, y) = \begin{cases} x^2 + \dfrac{xy}{3}, & 0 \le x \le 1, \\ & 0 \le y \le 2 \\ = 0 & \text{else where} \end{cases}$$

let $B = \{X + Y \geq 1\}$. Find $P(B)$.

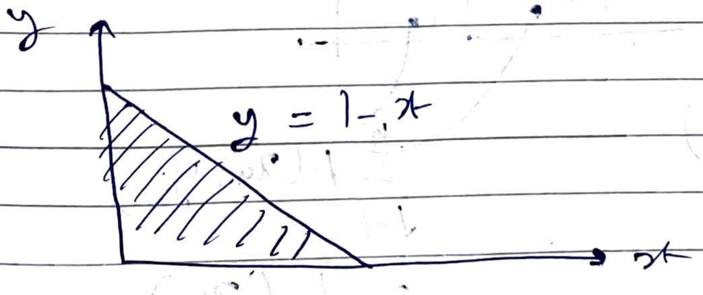
~~y = 1 - x~~

Soln Find $P(\bar{B}) = P(X + Y \le 1)$

or $\cancel{P(y \le x+1)}$ $P(Y < X - 1)$

We know: $P(a \le X \le b,\ c \le Y \le d)$

$$= \int_c^d \int_a^b f(x, y)\, dx\, dy$$



$\bar{B} = \{X + Y < 1$ or $Y < X - 1\}$.

$$P(\bar{B}) = \int_0^1 \int_0^{1-x} \left(x^2 + \frac{xy}{3}\right) dy\, dx$$

$$= \cancel{65/72}\ \ 7/72$$

$\therefore P(B) = 1 - P(\bar{B}) = 65/72$

Marginal and Conditional Probability Distn

Consider table 'A'

Find $P(Y=1)$
i.e. Second line produces 1 item.

Let (X, Y) be a 2D r.v. Then

$$\boxed{P(X = x_i) = \sum_{j=1}^{\infty} p(x_i, y_j)}$$

is the ~~marginal~~ probability $P(X = x_i)$.

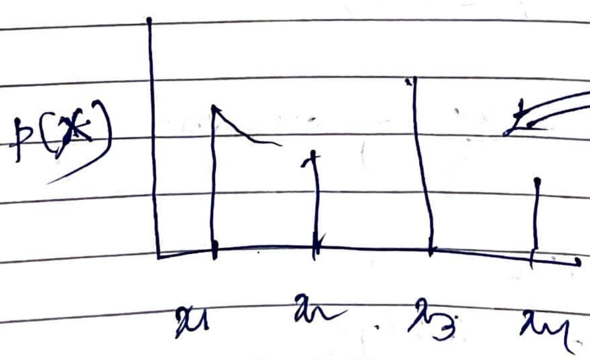
The f^n $p(x_i) = P(X = x_i)$ is the marginal probability dist of X.

Y \ X	x_1	x_2	x_3	x_4	Sum
y_1	-	-	-	-	$p(y_1)$
y_2	-				$p(y_2)$
y_3	-				$p(y_3)$
y_4	-				$p(y_4)$
Sum	$p(x_1)$	$p(x_2)$	$p(x_3)$	$p(x_4)$	1

$\sum_{i=1}^{4} p(x_i, y_1) = p(y_1)$

$\sum_{j=1}^{4} p(x_1, y_j) = p(x_1)$

$\sum_{j=1}^{4} p(x_4, y_j) = p(x_4)$



Marginal Prob. dist of X

Continuous Case

Let (X, Y) be a 2D cont r.v. Then define pdf $g(x)$ & $h(y)$ as follows:

$$\boxed{g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy}$$ Marginal pdf of x.

$$\boxed{h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx}$$ Marginal pdf of y

Also, $P(a \leq X \leq b)$

$= P(a \leq X \leq b, -\infty < Y < \infty)$

$$= \int_{a}^{b} \underbrace{\left[\int_{-\infty}^{\infty} f(x, y)\, dy\right]}_{g(x)} dx$$

$$\boxed{P(a \leq X \leq b) = \int_{a}^{b} g(x)\, dx}$$

Notes

→ Marginal pdfs are legitimate pdfs

∴ $\sum p(x_i) = 1$ & $\sum p(y_j) = 1$

& $\int_{R_x} g(x)\, dx = 1$ & $\int_{R_y} h(y)\, dy = 1$

Eg let $f(x,y) = 2(x+y-2xy)$,

$0 \le x \le 1$,

$0 \le y \le 1$.

Find $g(x)$ & $h(y)$

Sol

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

$$= \int_{0}^{1} 2(x+y-2xy)\, dy$$

$$= \left[2\left(xy + \frac{y^2}{2} - xy^2\right)\right]_0^1$$

$$= 2\left(x + \frac{1}{2} - x\right)$$

$$= 1$$ [$\therefore$ x is uniformly distr over $[0,1]$]

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

$$= \int_{0}^{1} 2(x+y-2xy)\, dx$$

$$= 2\left(x^2/2 + xy - x^2y\right)_0^1$$

$$= 2\left(\frac{1}{2} + y - y\right)$$

$$= 1$$

Independent Random Variables:

X & Y are independent r.v. if the out come of one does not affect the out come of the other.

Discrete

(X, Y) are ind. R.V. iff

~~$p(x_i, y_j) = p(x_i) \cdot q(y_j)$~~

$$\boxed{p(x_i, y_j) = p(x_i)\, q(y_j)} \quad \forall\, i, j$$

Cont.

$$\boxed{f(x, y) = g(x)\, h(y)} \quad \forall\, (x, y)$$

i.e. Joint pdf can be factored into marginal pdfs.

Also

$$p(x_i \mid y_j) = p(x_i) \quad \forall\, i, j$$

$$\& \quad q(y_j \mid x_i) = q(y_j) \quad \forall\, i, j$$

&

$$g(x \mid y) = g(x)$$

$$\& \quad h(y \mid x) = h(y) \quad \forall\, (x, y)$$

Table

Y \ X	0	1	2	q(y_j)
0	.1	.2	.2	.5
1	.04	.08	.08	.2
2	.06	.12	.12	.3
p(x_i)	.2	.4	.4	1.0

$p(0) = 0.2$ $\quad$ $q(0) = 0.5$

$p(1) = 0.4$ $\quad$ $q(1) = 0.2$

$p(2) = 0.4$ $\quad$ $q(2) = 0.3$

$p(1,2) = \cancel{0.4 * 0.3 = 0.12}$

$= 0.12$

$p(1) \cdot q(2) = 0.4 * 0.3$

$= 0.12$

$\therefore \; p(1,2) = p(1)\, q(2)$

Similarly for all i, j $\quad p(x_i, y_j)$

$= p(x_i)\, q(y_j)$

Th. Let (X, Y) be a 2D r.v.
Let A & B be events whose occurrence depends only on X and Y respectively
i.e. $A \subseteq R_X$ & $B \subseteq R_Y$
Then if X & Y are ind. then

$$P(A \cap B) = P(A) \cdot P(B)$$

Proof (Cont. Case)

$$P(A \cap B) = \iint_{A \cap B} f(x, y)\, dx\, dy$$

$$= \iint_{A \cap B} g(x)\, h(y)\, dx\, dy \quad [\because X \text{ \& } Y \text{ are ind}]$$

$$= \int_A g(x)\, dx \int_B h(y)\, dy$$

$$= P(A)\, P(B)$$

$$\therefore \boxed{P(A \cap B) = P(A)\, P(B)}$$

Conditional Distributions

Y \ X	0	1	2	3	4	5	
0	0	0.01	0.03	0.05	0.07	0.09	
1	0.01	0.02	0.04	0.05	0.06	0.08	
2	0.01	0.03	0.05	0.05	0.05	0.06	$= 0.25 = P(Y=2) = q(y_2) = q(2)$
3	0.01	0.02	0.04	0.06	0.06	0.05	

Find $P(X=2 \mid Y=2)$

Soln We know $P(A/B) = \dfrac{P(A\cap B)}{P(B)}$

$$\therefore P(X=2/Y=2) = \frac{P(X=2, Y=2)}{P(Y=2)}$$

$$= \frac{0.05}{\sum_{i=0}^{5} P(X=x_i, Y=2)} = \frac{0.05}{0.25} = 0.20$$

In General

$$\boxed{p(x_i/y_j) = \frac{p(x_i, y_j)}{q(y_j)}}$$

$$\boxed{q(y_j/x_i) = \frac{p(x_i, y_j)}{p(x_i)}}$$

Continuous Case

Conditional pdf of X for given $Y (=y)$

$$\boxed{g(x/y) = \frac{f(x,y)}{h(y)}, \quad h(y) > 0}$$

& $$\boxed{h(y/x) = \frac{f(x,y)}{g(x)}, \quad g(x) > 0}$$

Note :

$$\sum_{R_X} p(x_i/y_j) = \sum_{R_X} \frac{p(x_i, y_j)}{q(y_j)} = \frac{\sum_i p(x_i, y_j)}{q(y_j)}$$

$$= \frac{q(y_j)}{q(y_j)} = 1$$

Similarly $\sum_{R_Y} q(y_j/x_i) = 1$

Also

$$\int_{R_X} g(x/y)\, dx = \int_{R_X} \frac{f(x,y)}{h(y)}\, dx$$

$$= \frac{1}{h(y)} \int_{R_X} f(x,y)\, dx$$

$$= \frac{h(y)}{h(y)} = 1$$

Similarly $\int_{R_Y} h(y/x)\, dy = 1$

eg let $f(x,y) = x^2 + \frac{xy}{3}$; $0 \le x \le 1$
$0 \le y \le 2$

$$g(x) = \int_0^2 f(x,y)\,dy$$

$$= \int_0^2 \left(x^2 + \frac{xy}{3}\right) dy$$

$$= \left[x^2 y + \frac{x}{3}\cdot\frac{y^2}{2}\right]_0^2$$

$$\boxed{g(x) = 2x^2 + \frac{2}{3}x}$$

$$h(y) = \int_0^1 \left(x^2 + \frac{xy}{3}\right) \cancel{dy}\,dx = \left[\frac{x^3}{3} + \frac{x^2}{2}\cdot\frac{y}{3}\right]_0^1$$

$$= \cancel{\frac{y}{3}+} \frac{1}{3} + \frac{y}{6}$$

$$\Rightarrow \boxed{h(y) = \frac{y}{6} + \frac{1}{3}}$$

$$g(x|y) = \frac{f(x,y)}{h(y)} = \frac{x^2 + \frac{xy}{3}}{\frac{1}{3} + \frac{y}{6}}$$

$$\boxed{g(x|y) = \frac{6x^2 + 2xy}{2+y}}$$

$$\& \quad \boxed{h(y|x) = \frac{3x+y}{6x+2}}$$

Check that $g(x|y)$ is a pdf

$$\int_0^1 g(x|y)\,dx$$

$$= \int_0^1 \frac{6x^2 + 2xy}{2+y}\,dx$$

$$= \frac{1}{2+y}\left[6 \cdot \frac{x^3}{3} + 2y \cdot \frac{x^2}{2}\right]_0^1$$

$$= \frac{1}{2+y}\left[2x^3 + yx^2\right]_0^1$$

$$= \frac{1}{2+y}\left[2+y\right]$$

$$= 1$$

Similarly $\int_0^2 h(y|x)\,dy = 1$

Variance of a R.V.

Let X be a R.V. Variance of X, $V(X)$, or σ_X^2 is defined as:

$$V(X) = E[X - E(X)]^2$$

σ_X is called S.D. of X.

Note: (i) $V(X)$ is in square units of X

(ii) $V(X)$ is called the 2^{nd} moment of X

The k^{th} moment is

$$E[X - E(X)]^k.$$

Th $V(X) = E(X^2) - [E(X)]^2$

Proof:
$$V(X) = E[X - E(X)]^2$$
$$= E\{X^2 - 2X\,E(X) + [E(X)]^2\}$$
$$= E(X^2) - 2\,E(X)\,E(X) + [E(X)]^2 \quad (\because E(X) \text{ is a number})$$
$$= E(X^2) - 2[E(X)]^2 + [E(X)]^2$$

$$\boxed{V(X) = E(X^2) - [E(X)]^2}$$

~~Ex~~ **Ex:** Suppose the weather department classifies clouds as 0 ≡ perfectly clear sky

⋮

10 ≡ completely invisible sky.

Let X be a r.v s.t. $R_X = \{ \cancel{p_0}\, 0, \cancel{p_1}\, 1, \dots, \cancel{p_{10}}\, 10 \}$.

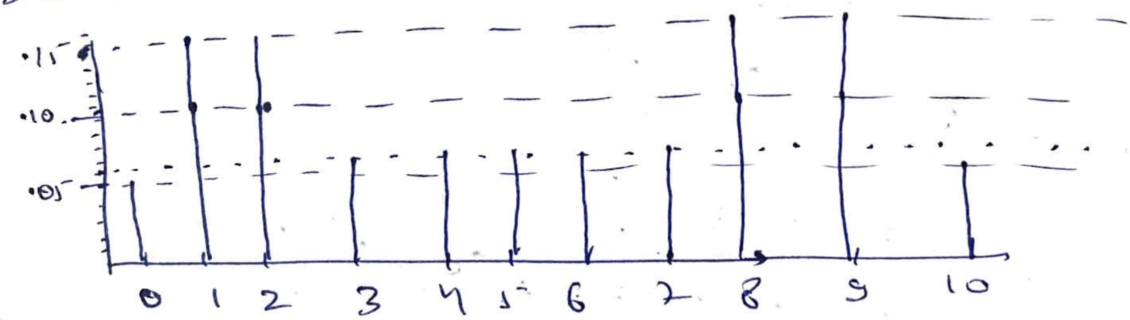
Let the probability dist of X be

~~$P(X = p_0) = p(p_0)$~~

$P(X=0) = p(0) = p_0 = p_{10} = 0.05$

$p_1 = p_2 = p_8 = p_9 = 0.15$

$p_3 = p_4 = p_5 = p_6 = p_7 = 0.06$



$.05 * 2 + .15 * 4 + 0.6 * 5$

$= .10 + .60 + .30$

$= 1$

Find $E(X)$ & $V(X)$

$E(X) = \sum x_i\, p(x_i) = 0 * .05 + 10 * .05 + 2 * .15 + 8 * .15 + 9 * .15$

$+ 1 * .15 + 2 * .15 + 8 * .15 + 9 * .15$

$+ 3 * .06 + 4 * .06 + 5 * .06 + 6 * .06 + 7 * .06$

$= 0.5 + 3.00 + 1.5$

$\Rightarrow \boxed{E(X) = 5}$

$V(X) = E(X^2) - [E(X)]^2$

Calculation of $E(X^2)$

$E(X^2) = \sum x_i^2 \cdot p(x_i)$ $\qquad [E[f(X)] = \sum f(x_j) \cdot p(x_j)]$

$= 0^2(.05)^2 + 100(.05)$
$+ 1(.15) + 4(.15) + 64(.15) + 81(.15)$
$+ 9(.06) + 16(.06) + 25(.06) + 36(.06) + 49(.06)$

$= 35.6$

$\therefore V(X) = E(X^2) - [E(X)]^2$

$= 35.6 - (5)^2$

$= 35.6 - 25$

$\boxed{V(X) = 10.6}$

$\therefore \sigma_X = \sqrt{10.6}$

$\boxed{\sigma_X = 3.25}$

The Correlation Coefficient

Let (X, Y) be a 2D RV. Then the C.C between X & Y is :

$$\boxed{\rho_{XY} = \frac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X)V(Y)}}}$$

Th: If X & Y are independent then $\rho = 0$

Proof: $\rho = \dfrac{E\{[X - E(X)][Y - E(Y)]\}}{\sqrt{V(X) V(Y)}}$

$$E\{[X - E(X)][Y - E(Y)]\} = E\{XY - X E(Y) - Y E(X) + E(X) E(Y)\}$$

$$= E(XY) - E(X) \cancel{E(Y)} - E(Y) \cancel{E(X)} + \cancel{E(X) E(Y)}$$

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

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

$\Rightarrow \rho = 0$

Chebyshev's Inequality:

If pdf of X is known then $E(X)$ & $V(X)$ can be computed. However if $E(X)$ & $V(X)$ are known, we cannot reconstruct the pdf and hence we can't get quantities like $P[X - E(X)]$.

However from the knowledge of $E(X)$ & $V(X)$ we can get upper & lower bounds of quantities such as $P[|X-c|]$ etc.

<u>Chebyshev's inequality</u>

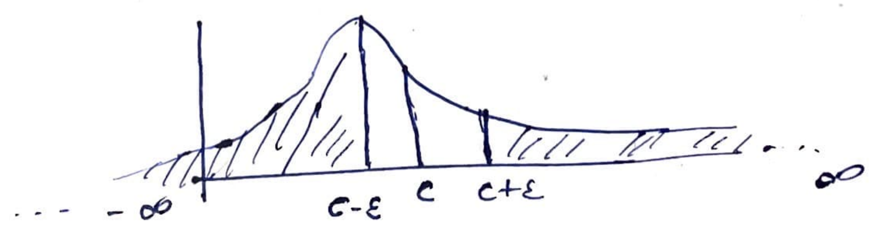
Let X be a rv & $E(X) = \mu$.
Let $c \in \mathbb{R}$. Then if $E[(X-c)^2]$ is finite
and $\varepsilon \in \mathbb{R}^+$, then

$$P[|X-c| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E[(X-c)]^2$$

<u>Proof</u> ~~P[[|X-c|~~

$$P([|X-c| \geq \varepsilon]) = \int_{x:|x-c|\geq\varepsilon} f(x)\,dx$$

<u>Note</u> limit of integration is <u>$-\infty$ to $c-\varepsilon$</u> & <u>$c+\varepsilon$ to ∞</u>



$$|x-c| \geq \varepsilon \Rightarrow \frac{(x-c)^2}{\varepsilon^2} \geq 1.$$

$c+\varepsilon$ to ∞

$$\therefore \int_{-\infty \text{ to } c-\varepsilon} f(x)\,dx \leq \int_R \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx$$

$$R = \{x : |x-c| \geq \varepsilon\}$$

~~This integral~~

$$\int_{x:\,|x-c|\geq\varepsilon} f(x)\,dx \leq \int_{R} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx$$

$$R = \{x : |x-c| \geq \varepsilon\}$$

$$\leq \int_{-\infty}^{\infty} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx \quad \text{—} (1)$$

But $$\int_{-\infty}^{\infty} \frac{(x-c)^2}{\varepsilon^2} f(x)\,dx = \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty} (x-c)^2 f(x)\,dx$$

$$= \frac{1}{\varepsilon^2} E[(X-c)^2]$$

$$\therefore \int_{x:\,|x-c|\geq\varepsilon} f(x)\,dx \leq \frac{1}{\varepsilon^2} E[(X-c)^2]$$

or $$\boxed{P[|X-c| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E[(X-c)^2]}$$

Consequences

1. Choosing $c = \mu$

$$P[|X-\mu| \geq \varepsilon] \leq \frac{1}{\varepsilon^2} E[(X-\mu)^2]$$

$$\Rightarrow \boxed{P[|X-\mu| \geq \varepsilon] \leq \frac{Var(X)}{\varepsilon^2}} \quad \text{--- } (1)$$

2. Choosing $c = \mu$ & $\varepsilon = k\sigma$; $\sigma^2 = Var(X)$

$$P[|X-\mu| \geq k\sigma] \leq \frac{\sigma^2}{(k\sigma)^2}$$

$$\Rightarrow \boxed{P[|X-\mu| \geq k\sigma] \leq \frac{1}{k^2}} \quad \text{--- } (2)$$

$$\Rightarrow P[|X-\mu| \geq \sigma] \leq 1$$

3. If $V(X) = 0$ then $P[X = \mu] = 1$.

Proof (1).

$P[|X-\mu| \geq \varepsilon] = 0$ [$\because Var(X) = 0$]

$\therefore P[|X-\mu| < \varepsilon] = 1$ for any $\varepsilon > 0$

$\because \varepsilon$ is arbitrary $\therefore$ proved.