Recap

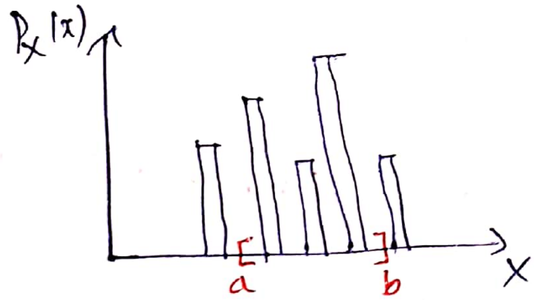
Random Variable $[X : S \to \mathbb{R}$
$s \mapsto X(s) = x \in \mathbb{R}$

Discrete [R_X is finite or Countable]

Single Variable

Probability Mass function (PMF)

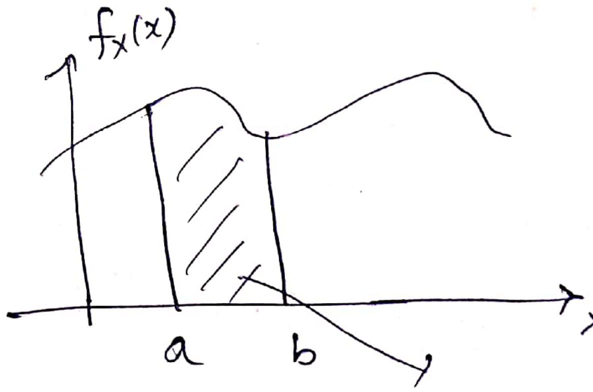
- $P_X(x) = P(X = x)$, $x \in R_X$ ↳ Range of X.
- $P_X(x) \geq 0$
- $\sum_{x \in R_X} P_X(x) = 1$
- $P(a \leq X \leq b) = \sum_{x: a \leq x \leq b} P_X(x)$.



Continuous [R_X is interval (OR) Union of intervals]

Probability density function (PDF)

- $P(X \in B) = \int_B f_X(x)\,dx$, $B \subseteq \mathbb{R}$.
- $f_X(x) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$



$P(a \leq X \leq b)$.

Cumulative Distribution function.

- $F_X(x) = P(X \leq x) = \sum_{y \leq x} P_X(y)$

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$
$$= F_X(b) - F_X(a).$$

- $P(X = a) = P(a \leq X \leq a) = \int_a^a f_X(x)\,dx = 0$
- $F_X(x) = P(X \leq x) = \int_{-\infty}^{x} f_X(y)\,dy$
- $\boxed{f_X(x) = \frac{d}{dx} F_X(x)}$

Mean & Variance of a r.v X

$$\mu_X = E(X) = \begin{cases} \sum_{x \in R_X} x\, P_X(x) & \text{if } X \text{ is discrete r.v} \\ \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx & \text{if } X \text{ is continuous r.v} \end{cases}$$

It describes the center of the distribution.

$$\sigma_X^2 = Var(X) = E((X-\mu_X)^2) = \begin{cases} \sum_{x \in R_X} (x-\mu)^2 P_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)\, dx & \text{if } X \text{ is continuous} \end{cases}$$

Measure of dispersion of X.

$$Var(x) = E(X-\mu_X)^2 = E(X^2) - \mu_X^2 = E(x^2) - (E(X))^2.$$

$$\sigma_X = S.D(X) = \sqrt{Var(X)}.$$

$$E(g(X)) = \begin{cases} \sum_{x \in R_X} g(x)\, P_X(x) & \text{if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx & \text{if } X \text{ is continuous.} \end{cases}$$

Joint Probability Distributions [Pair of r.vs X & Y]

Joint Probability Mass function

- $\boxed{P_{X,Y}(x,y) = P(X=x, Y=y)}$ "and" = "$\cap$"
- $P_{X,Y}(x,y) \geq 0$
- $\sum_{x \in R_X} \sum_{y \in R_Y} P_{X,Y}(x,y) = 1$

Joint Probability Density function.

- $f_{X,Y}(x,y) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx\, dy = 1$
- For any region D of the two dimensional space $\mathbb{R}^2$.

$$\boxed{P((X,Y) \in D) = \iint_D f_{X,Y}(x,y)\, dx\, dy}$$

Hypergeometric Distribution:

In binomial distribution, independence among trials is required the Sampling done with replacement

On the Other hand, the hypergeometric distribution does not require independence and is based on Sampling done without replacement

A set of N objects contains

K Objects Classified as Successes

N-K Objects Classified as failures

A sample of size 'n' Objects is selected randomly (without replacement) from the N objects where $K \le N$ & $n \le N$.

The p.m.f of the hypergeometric r.v X, the number of successes in a random Sample of Size 'n' Selected from N items of which 'K' are labelled success and N-K labelled failure, is

$$\boxed{P_X(x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}},} \quad \max\{0,\ n-(N-K)\} \le x \le \min\{n, K\}$$

Note that x and $n-x$ are no more than K & N-K respectively and both of them cannot be less than 0.

Usually, when both K (Number of successes) and N-K (the number of failures) are larger than the sample size 'n', the range of a hypergeometric r.v will be $x = 0, 1, 2, \dots n$.

~~Basic Assumption~~ [$n \le K$, $n-x \le N-K$

Note: Sampling without replacement is frequently used for inspection.

(P1). A batch of parts contains 100 from a local supplier of tubing and 200 from a supplier of tubing in the next state. If four parts are selected randomly and without replacement,

a) What is the probability they are all from the local supplier?

Soln:

Let X equal the number of parts in the sample from the local supplier. Then X has a hypergeometric distribution

The required prob. is $P(X=4) = \dfrac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}} = 0.0119$

(b). What is the probability that two or more parts in the sample are from the local supplier?

$$P(X \geq 2) = \frac{\binom{100}{2}\binom{200}{2}}{\binom{300}{4}} + \frac{\binom{100}{3}\binom{200}{1}}{\binom{300}{4}} + \frac{\binom{100}{4}\binom{200}{0}}{\binom{300}{4}}$$

$$= 0.298 + 0.098 + 0.0119 = 0.408$$

(c) What is the probability that at least one part in the sample is from the local supplier?

$$P(X \geq 1) = 1 - P(X=0) = 1 - \frac{\binom{100}{0}\binom{200}{4}}{\binom{300}{4}} = 0.804.$$

Marginal Probability Distributions:

If (X, Y) has Joint PMF $P_{X,Y}(x,y)$, then the Marginal PMF's of X & Y are

$$P_X(x) = \sum_{y \in R_Y} P_{X,Y}(x,y)$$

$$P_Y(y) = \sum_{x \in R_X} P_{X,Y}(x,y).$$

Note that earlier notation for Range of X is S_X. Here I used it as R_X

If (X, Y) are continuous r.vs with joint PDF $f_{X,Y}(x,y)$, then Marginal PDF's of X & Y are

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy \quad \text{—} ①$$

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx \quad \text{—} ②$$

(P_1). The joint density function of X & Y is

$$f_{X,Y}(x,y) = \begin{cases} \frac{5y}{4}, & -1 \le x \le 1, \quad x^2 \le y \le 1 \\ 0, & \text{Otherwise} \end{cases}$$

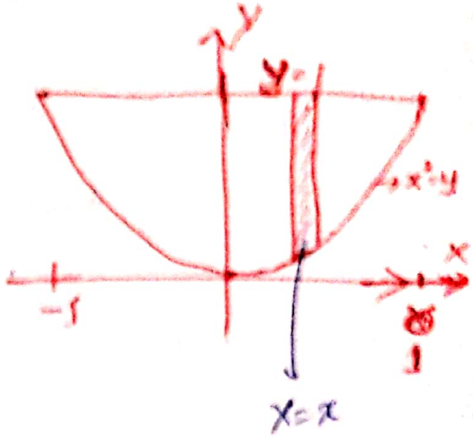
Find the marginal PDFs $f_X(x)$ & $f_Y(y)$.

To find $f_X(x)$:

When $x < -1$ or $x > 1$, $f_{X,Y}(x,y) = 0$

$\therefore$ ① $\Rightarrow$ $f_X(x) = 0$.

For $-1 \le x \le 1$,



$$f_X(x) = \int_{y=x^2}^{y=1} f_{X,Y}(x,y)\,dy$$

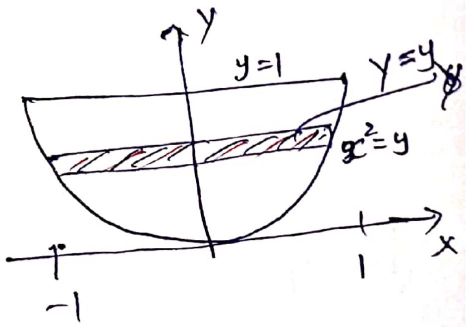
$$= \int_{y=x^2}^{y=1} \frac{5y}{4}\,dy$$

$$= \frac{5}{4}\left[\frac{y^2}{2}\right]_{y=x^2}^{y=1} = \frac{5}{8}\left[1-x^4\right]$$

$$\therefore f_X(x) = \begin{cases} \frac{5}{8}(1-x^4), & -1 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

To find $f_Y(y)$:

We note that for $y<0$. or $y>1$, $f_{X,Y}(x,y) = 0$



② $\Rightarrow$ $f_Y(y) = 0$

For $0 \le y \le 1$, we integrate over the horizontal bar, marked $Y=y$.

The boundaries of the bar are $x=-\sqrt{y}$ to $x=\sqrt{y}$

$$\therefore f_Y(y) = \int_{x=-\sqrt{y}}^{x=\sqrt{y}} f_{X,Y}(x,y)\,dx = \int_{x=-\sqrt{y}}^{x=\sqrt{y}} \frac{5y}{4}\,dx = \frac{5y}{4}\left[x\right]_{x=-\sqrt{y}}^{x=\sqrt{y}}$$

$$= \frac{5y^{3/2}}{2}$$

$$f_Y(y) = \begin{cases} \frac{5}{2}y^{3/2}, & 0 \le y \le 1 \\ 0, & \text{Otherwise.} \end{cases}$$

Conditional Probability Distributions:

Conditioning by a r.v

Let us consider the example of Mobile Response Time.

Let X denote the number of bars of service [Signal Strength]

Let Y denote the response time (to the nearest second)

Here the Response time is the speed of page downloads

Joint PMF of X & Y

y = Response time (nearest second)	x = Number of Bars of Signal Strength: x=1	x = 2	x=3	$P_Y(y)$
y = 4	0.15	0.1	0.05	0.3
y = 3	0.02	0.1	0.05	0.17
y = 2	0.02	0.03	0.2	0.25
y = 1	0.01	0.02	0.25	0.28
$P_X(x)$	0.2	0.25	0.55	1

→ Marginal P.M.F of Y

One expects the Probability Y=1 to be greater at X=3 bars than at X=1 bar.

⇒ The knowledge of one r.v can change the Probabilities that associate with the values of the other.

We write the above conditional Probabilities as

$P(Y=1 \mid X=3)$ and $P(Y=1 \mid X=3)$.

⇒ Conditional P.M.F

Defn: Conditional Probability Mass function:

Given discrete r.vs X & Y with joint PMF $P_{X,Y}(x,y)$. For any event $Y=y$ s.t $P_Y(y) > 0$, the Conditional PMF of X given $Y=y$ is

Recall

$P(A|B) = \frac{P(A \cap B)}{P(B)}$.

$$P_{X|Y}(x|y) = P(X=x \mid Y=y)$$

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

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

$= P(B|A) \cdot P(A)$.

$$\boxed{P_{X|Y}(x|y) = \frac{P_{X,Y}(x,y)}{P_Y(y)}}$$

Result: For discrete r.vs X, Y with joint PMF $P_{X,Y}(x,y)$ and x, y s.t $P_X(x) > 0$ & $P_Y(y) > 0$

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

Example: Conditional Probabilities for Mobile Response time

$$P(Y=1 \mid X=3) = \frac{P(Y=1, X=3)}{P(X=3)} = \frac{P_{X,Y}(3,1)}{P_X(3)} = \frac{0.25}{0.55}$$

$$= 0.454.$$

$$P(Y=2 \mid X=3) = \frac{P(Y=2, X=3)}{P(X=3)} = \frac{P_{X,Y}(3,2)}{P_X(3)} = \frac{0.2}{0.55}$$

$$= 0.364.$$

Defn. Conditional PDF

Given continuous r.vs X & Y with joint PDF $f_{X,Y}(x,y)$

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

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

Similarly,

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

Result:

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

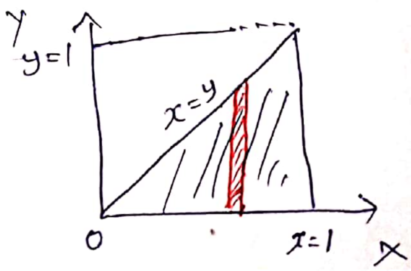
(P1). R.vs X & Y have joint PDF. $f_{X,Y}(x,y) = \begin{cases} 2, & 0 \le y \le x \le 1 \\ 0, & \text{Otherwise.} \end{cases}$

For $0 \le x \le 1$, find the conditional PDF $f_{Y|X}(y|x)$ &

For $0 \le y \le 1$, find the conditional PDF $f_{X|Y}(x|y)$.

Soln:

For $0 \le x \le 1$, $f_X(x) = \displaystyle\int_{y=0}^{y=x} f_{X,Y}(x,y)\, dy$



$$= \int_{y=0}^{y=x} 2\, dy = 2[y]_{y=0}^{y=x} = 2x.$$

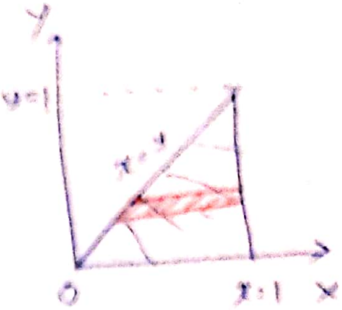
$$\Rightarrow f_{Y|X}(x,y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \begin{cases} \frac{2}{2x}, & 0 \le y \le x \\ 0, & \text{otherwise.} \end{cases}$$

Scanned with CamScanner

For $0 \le y \le 1$, The Marginal PDF of Y.

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx = \int_{x=y}^{x=1} 2\,dx = 2[x]_{x=y}^{x=1}$$

$$= 2[1-y]$$



For y choose Horizontal strip. Limits for x vary from Left boundary to right boundary Here $x=y$ to $x=1$

The Conditional PDF

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

$$= \begin{cases} \dfrac{2}{2(1-y)}, & y \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Conditional Expectation (Mean) and Variance:

The Conditional mean of Y given $X=x$, denoted as $E(Y|X=x)$ or $\mu_{Y|X}$, is

$$E(Y|X=x) = \begin{cases} \displaystyle\sum_{y \in R_Y} y \cdot P_{Y|X}(x|y) & \text{if } X,Y \text{ are discrete r.vs} \\ & (R_Y - \text{Range of } Y) \\ \displaystyle\int_{-\infty}^{\infty} y \cdot f_{Y|X}(x,y)\,dy & \text{If } X,Y \text{ are Continuous r.vs.} \end{cases}$$

The conditional variance of Y given $X=x$, $Var(Y|X=x)$ or $\sigma^2_{Y|X}$ is

$$Var(Y|X=x) = \begin{cases} \displaystyle\sum_{y \in R_Y} (y-\mu_{Y|X})^2 P_{Y|X}(y|x) & [\text{Discrete Case}] \\ \displaystyle\int_{-\infty}^{\infty} (y-\mu_{Y|X})^2 f_{Y|X}(y|x)\,dy & [\text{Continuous}]. \end{cases}$$