03.09.24

Math-243 (Probability and Statistics)

Ref: Probability and Statistics for Engineers and Scientists by Ronald E Walpole, Raymond H. Myers, Sharon L Myers and Keying Ye.

- Random Variable.
- Discrete and Continuous probability distributions
- Joint probability distributions
- Marginal distributions and independence
- Expectations, variance and covariance of random variables and their properties
- Chebyshev's theorem
- Discrete probability distributions : Binomial, multinomial, Poisson distribution and their properties
- Continuous " " : Uniform, normal, chi-square distribution and their properties
  - exponential distribution
- Confidence intervals
- Hypothesis testing

No. of Heads occuring in tossing a coin 10 times

0 1 2 3 . . . . 10 → কতবার head আসতে পারে (0 - 10)

no. of Heads আসাটা → variable (random variable)

random variable – যেই variable এর সাথে probability র association আছে।
কখন কোনটা আসবে সেটার সাথে একটা certain probability আছে।

random variable
↳ discrete – coin toss এ # Heads আসা। 1 এবং 2 এর মাঝে অন্য কিছু আসতে পারবে না।
↳ continuous – temperature, height

Probability Distributions: random variable এর বিভিন্ন value-র সাথে probability র association বোঝায় যে function.

২টো Ludo dice roll এর sample space

all possible outcomes of rolling a dice (successively) 2 times →

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

sum of upper faces of 2 dices : minimum 2 → (1, 1)

maximum 12 → (6, 6)

→ একটা variable

→ discrete value এর জন্য

Probability distribution / Probability function → P(x)

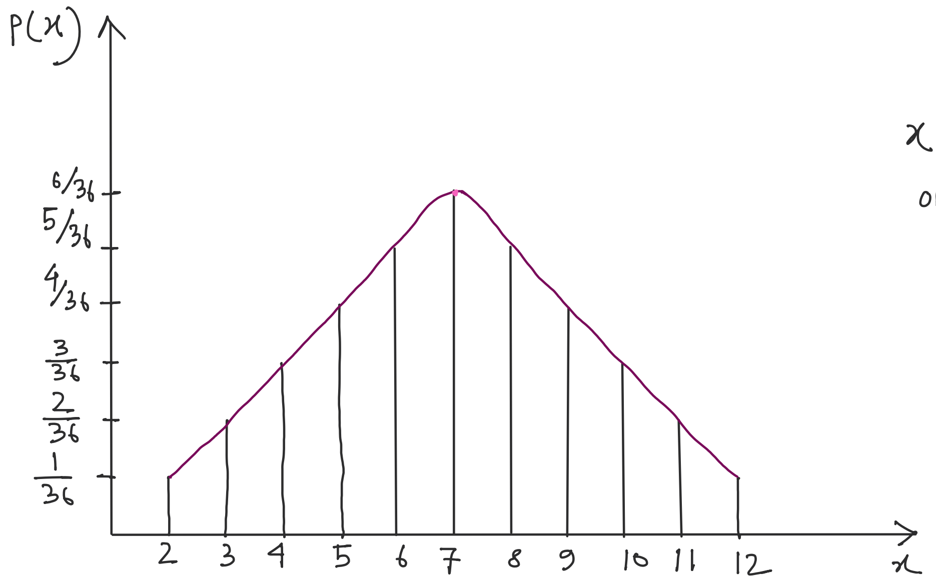
X	2	3	4	5	6	7	8	9	10	11	12
P(x)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

most likely occur

least likely occur

x axis — random variable (discrete value)

y axis — probability



$x \to P(x)$

one to one mapping

distribution graph

symmetric / uniform distribution হয়েছে

বিভিন্ন shape এর distribution হতে পারে। e.g. $p(x) = e^{-x}$

X, Y, Z → random variables

$P(X) = f(x)$

$P(X, Y)$ — joint distribution

(multi variable এর জন্য)

joint distribution থেকে single distribution বের করা

$P(X,Y)$ থেকে $P(X)$

— marginal distribution

random variable X , random variable Y এর উপর depend করলে

— dependency → Independence এ পড়বো

imp variance → measure of dispersion

— distribution analysis এ help করে।

10.09.24

Probability mass function → discrete probability র জন্য function

Probability density → continuous probability র জন্য যে function

$$P(X=x) = \underline{f(x)}$$

→ probability distribution or probability mass function (P.M.F)

if (i) $f(x) \geq 0$

(ii) $\sum_{x} f(x) = 1$

random variable যে value গুলো নিতে পারে তাদের probability -র sum.

e.g. $f(x) = x^2$

$x = 0 \rightarrow f(0) = 0$

$f(1) = 1$

$f(2) = 4$

but 2nd condition hold করে না। so P.M.F বলতে পারবো না।

3.5 Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$;

(b) $f(x) = c\binom{2}{x}\binom{3}{3-x}$, for $x = 0, 1, 2$.

3.5(a) value of c = ?

for $f(x)$ to serve as a probability distribution of random variable X.

$$f(x) = c(x^2 + 4), \quad x = 0, 1, 2, 3$$

$\hookrightarrow$ discrete random variable

$f(0) = 4c$ $\qquad$ $f(1) = 8c$

$f(1) = 5c$ $\qquad$ $f(3) = 13c$

$$\rightarrow (4 + 5 + 8 + 13)\, c = 1$$

$$c = \frac{1}{30}$$

x	0	1	2	3
$f(x)$	$\frac{4}{30}$	$\frac{5}{30}$	$\frac{8}{30}$	$\frac{13}{30}$

$f(0) = \frac{4}{30} > 0$

$f(1) = \frac{5}{30} > 0$

Probability density function : if (i) $f(x) \geq 0$

(P.D.F) $\qquad$ (ii) $\int f(x)\, dx = 1$

$X \rightarrow$ continuous random variable

$$P(a < x < b) = \int_a^b f(x)\, dx$$

discrete case এ → particular point এ probability calculate

continuos ,, এ → একটা interval এ ,, ,,

(point এ নিলে probability = 0)

discrete point এ height থাকে but width থাকে না, so area zero হবে। integration করলে area পাবো, তাই particular point এ P.D.F zero হবে।

$\boxed{c \quad < a < x < b \quad < d}$

random variable x এর value c থেকে d range এ হতে পারে।
এই range এ Σprobability নিলে 1।
but $a<x<b$ range এ probability will be<1.

$\boxed{\text{Walpole}}$ → exercise + example

exercise

3.6 The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20{,}000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

$P \geq 200$

$\int_{200}^{\infty} \frac{20000}{(x+100)^3} dx$

Find the probability that a bottle of this medicine will have a shell life of
(a) at least 200 days;
(b) anywhere from 80 to 120 days.

$$(b)\ P(80 < x < 120) = \int_{80}^{120} \frac{20000}{(x+100)^3} dx$$

$$= 0.102$$

$$P(0<x<200) = \int_0^{200} \frac{20000}{(x+100)^3} \, dx$$

$$= 0.889$$

at least 200

$P(x \geq 200) = 1 - P(0<x<200)$

$= 0.11$

(a) $$P(X \geq 200) = \int_{200}^{\infty} \frac{20000}{(x+100)^3}$$

$$= \frac{1}{9}$$

প্রতি ৯ টার মধ্যে 1 টার shell life ≥ 200

100 11

Example 3.8: A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

real phenomena to probability max function convert করা।

$T = 20$

$D = 3$

$ND = 17$

অনেক combination আসতে পারে।

D-D, ND-D, ND-ND,

D-ND

X: no. of defective

↳ 0, 1, 2 হতে পারে।

as computer দুইটাই নিবো তাই x এর value > 2 হবে না।

$P(x=0) = f(0)$

$$= \frac{\boxed{{}^{17}C_2 \times {}^{3}C_0}}{{}^{20}C_2}$$ → event এর sample size.

$$= \frac{68}{95}$$

17টি ND থেকে 2টি

3টি D থেকে একটিও না

Total ${}^{20}C_2$ টি computer → sample space এর size

$P(x=1) = f(1)$

$$= \frac{{}^{3}C_1 \times {}^{17}C_1}{{}^{20}C_2} = \frac{51}{190}$$

$P(x=2) = f(2)$

$$= \frac{{}^{3}C_2 \times {}^{17}C_0}{{}^{20}C_2} = \frac{3}{190}$$

$\therefore P(x=0,1,2) = P(x=1) + P(x=2) + P(x=3)$

$= 1$

17.09.24

Joint Probability Distribution :

X, Y are discrete random variables.

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

ii) $\sum_x \sum_y f(x,y) = 1$ over the domain probability sum $=1$

iii) $P(X=x, Y=y) = f(x,y)$ xy plane

Probability distribution for more than 1 random variable

dependency থাকলে
joint probability distribution

Let, A, B, C, D, Y random variables
Y কে পাওয়ার জন্য A, B, C, D লাগে

$X=x$ straight line parallel to y axis
$Y=y$ straight line parallel to x axis

xy plane এর সব জায়গা domain এ নেই।
(discrete)

continuous $\int_x \int_y f(x,y)\, dy\, dx = 1$ (partial integration দিয়ে করি)

Example 3.14: Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) the joint probability function $f(x, y)$,

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) | x + y \leq 1\}$.

B − 3
R − 2
G − 3

2 pens need to be selected

X : no. of blue pens
Y : no. of red pens

X random variable as defined as no. of blue pens.

Probability distribution of $f(x,y)$:

$y \backslash x$	0	1	2
0	$3/28$	$9/28$	$3/28$
1	$3/14$	$3/14$	0
2	$1/28$	0	0

≥ 3 হবে না cz pen ২টোই select করবো।

3টা G থেকেই ২টো নিচ্ছি

$$f(0,0) = f(X=0, Y=0) = \frac{{}^3C_2 \times {}^2C_0 \times {}^3C_0}{{}^8C_2}$$

$$= \frac{3}{28}$$

$$f(2,2) = f(1,2) = f(2,1) = 0$$

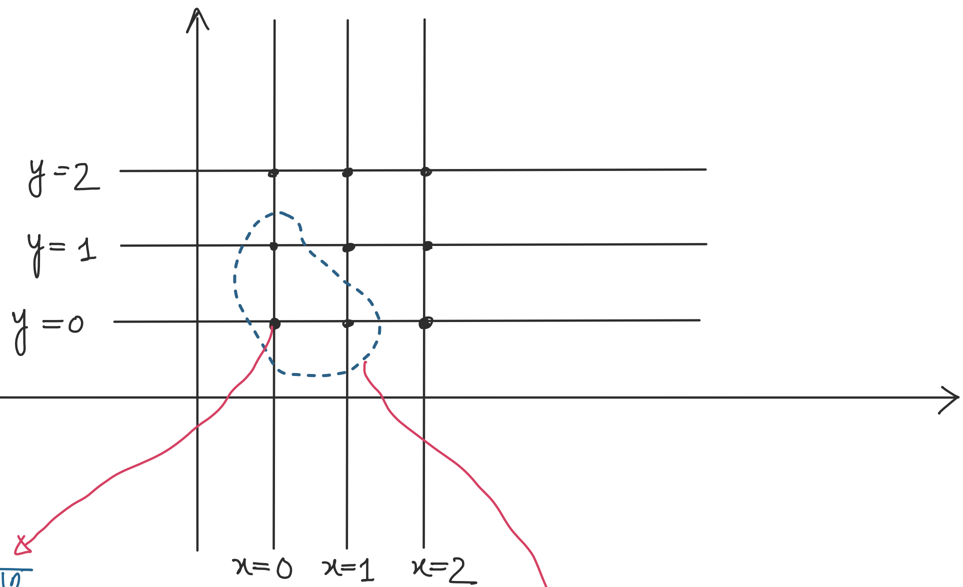
২এর বেশি pen select করবোনা।

$$f(1,0) = \frac{{}^3C_1 \times {}^3C_1 \times {}^2C_0}{{}^8C_2} = \frac{9}{28}$$

$$f(x,y) = \frac{{}^{3}C_{x} \cdot {}^{2}C_{y} \cdot {}^{3}C_{2-(x+y)}}{{}^{8}C_{2}}$$

domain : $x+y \leq 2$

$$f(x,y) = \begin{cases} \dfrac{{}^{3}C_{x} \cdot {}^{2}C_{y} \cdot {}^{3}C_{2-(x+y)}}{{}^{8}C_{2}} & , \; x+y \leq 2 \\ 0 & , \; x+y > 2 \end{cases}$$



3D তে bar রাখা হলে height হবে $3/28$. (এই point এ)

region এ probability calculate — প্রতিটা probability-র sum আসবে।

Example 3.15: A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X,Y) \in A]$, where $A = \{(x,y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

(b) $$\int_0^{1/2}\int_{1/4}^{1/2} 2/5\,(2x+3y)\,dy\,dx$$

$$= \frac{2}{5}\int_0^{1/2} \left[2xy + \frac{3}{2}y^2\right]_{1/4}^{1/2} dx$$

$$= \frac{2}{5}\int_0^{1/2}\left(\frac{x}{2} + \frac{9}{32}\right) dx$$

$$= \frac{13}{160}$$

(a) $$\int_0^1\int_0^1 2/5\,(2x+3y)\,dy\,dx$$

$0 \to 1$ x
$0 \to 1$ y] unit square (1st quadrant এ)

এই unit square এর বাহিরের সব point এ value $=0$

$\frac{1}{4} < y < \frac{1}{2}$, $0 < x < \frac{1}{2}$ → domain এ আছে তাই probability থাকবে

condition 1 holds : x, y positive

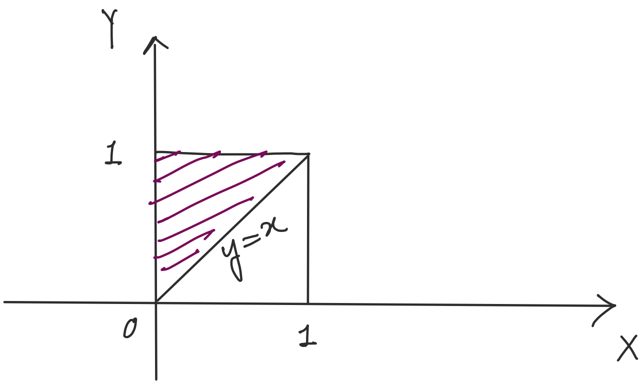
scalar multiplication, addition
সব positive এই থাকবে।

exercise:

3.43 Let X denote the reaction time, in seconds, to a certain stimulus and Y denote the temperature (°F) at which a certain reaction starts to take place. Suppose that two random variables X and Y have the joint density

$$f(x, y) = \begin{cases} 4xy, & 0 < x < 1,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find
(a) $P(0 \le X \le \frac{1}{2}$ and $\frac{1}{4} \le Y \le \frac{1}{2})$;
(b) $P(X < Y)$.



নীচের half

$$\int_0^1 \int_0^1 4xy\,dy\,dx - \boxed{\int_0^1 \int_0^x 4xy\,dy\,dx}$$

$$= 1 - \int_0^1 \left[2xy^2\right]_0^x dx = 1 - \int_0^1 2x^3\,dx$$

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

or

$$\int_0^1 \int_x^1 4xy \, dy \, dx = \frac{1}{2}$$

$x \le y \le 1$

$0 \le x \le 1$

or,

$0 \le y \le 1$

$0 \le x \le y$

$$\int_0^1 \int_0^y 4xy \, dx \, dy = \frac{1}{2}$$

22 . 10 . 24

Definition 3.8: The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \ge 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum\sum_A f(x, y)$.

domain এর subset

domain discrete — summation

full xy plane এ continuous — integration

$$\int_0^5 x^2 y \, dx$$ → x এর জন্য integration করছি। y এর function পাবো।

$$= y \int_0^5 x^2 \, dx$$

$$= \frac{125}{3} y$$

2 variable এর function কে কোনো একটার জন্য integrate করলে অন্যটার equation পাবো। — marginal distribution

y এর জন্য integration / sum — marginal distribution of x

x — y

joint distribution থেকে marginal distribution এভাবে বের করবো।

Definition 3.10: The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x,y) \quad \text{and} \quad h(y) = \sum_x f(x,y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx$$

for the continuous case.

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y) & 0 \le x \le 1 \\ & 0 \le y \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

continuous domain

↳ integrate

$$g(x) = \int f(x,y)\,dy$$

$$= \frac{2}{5}\int_{y=0}^{1}(2x+3y)\,dy$$

$$= \frac{2}{5}\left[2xy + \frac{3}{2}y^2\right]_0^1$$

$$= \frac{2}{5}\left(2x + \frac{3}{2}\right)$$

$$= \frac{4}{5}x + \frac{3}{5}$$

এখানে, marginal of x এর domain $\underline{0 \text{ to } 1}$

as joint distribution এর মাঝে আছে।

যদিও straight line এর domain $\mathbb{R}$

$$h(y) = \int f(x,y)\,dx$$

$$= \frac{2}{5}\int_{x=0}^{1}(2x+3y)\,dx$$

$$=\frac{2}{5}\left[x^2 + 3xy\right]_0^1$$

$$=\frac{2}{5}(1+3y)$$ $0 \le y \le 1 \to$ domain

$$g(x) = \frac{1}{5}(4x+3)$$

$$g(x)\cdot h(y) = \frac{2}{25}(1+3y)(4x+3)$$

$$\neq f(x,y)$$

↳ ∴ x, y are dependent.

marginal এর multiplication যদি joint এর সমান হয় — independent

Definition 3.12: Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Definition 3.11: Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x, y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x, y)}{h(y)}, \text{ provided } h(y) > 0.$$

$f(x|y) = f(x)$ given y

conditional probability র function.

$$f(x|y) = \frac{f(x,y)}{h(y)} \qquad h(y) > 0 \qquad \frac{\text{joint distribution}}{\text{marginal}}$$

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

$g(x) > 0$

probability বোঝায় so (+ve)

$$P(a < X < b \mid \underline{Y = y}) = \int_a^b f(x|y)\, dx.$$

conditional probability কে x এর domain এ integrate করলেই পাবো। (Y given)

random variable Y যখন y value নিবে তখন $a < X < b$ হওয়ার probability

Example 3.19: The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x,y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$. $f(x|y)$

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

continuous random variable

$P(Y > 0.5 \mid x = 0.25)$

$$= \int_{0.5}^{1} f(y|x)\, dy$$

Chapter - 4

Mathematical Expectation

$\mu \rightarrow$ expectation of random variable

$\mu_X = E[X]$ = mean of random variable

$= \int x f(x)\, dx$ when x is continuous

$= \sum_x x f(x)$ when x is continuous

$E[g(X)] = ?$

random variable এর যেকোনো function itself একটা random variable

(marginal না)

$Y = 3X + 5$

$g(X)$

Find $E[3x+5] = ?$

Theorem 4.1: Let X be a random variable with probability distribution $f(x)$. The expected value of the random variable $g(X)$ is

$$\mu_{g(X)} = E[g(X)] = \sum_x g(x) f(x)$$

$f(g(x))$ না

if X is discrete, and

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

if X is continuous.

Example 4.3: Let X be the random variable that denotes the life in hours of a certain electronic device. The probability density function is

$$f(x) = \begin{cases} \frac{20{,}000}{x^3}, & x > 100, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected life of this type of device.

mean বের করা লাগবে।

$$\mu = E(X) = \int_{100}^{\infty} x \cdot \frac{20000}{x^3} \, dx$$

$$= 200$$

29.10.24

3.53 Given the joint density function

$$f(x,y) = \begin{cases} \frac{6-x-y}{8}, & 0 < x < 2,\ 2 < y < 4, \\ 0, & \text{elsewhere}, \end{cases}$$

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

x constant so (y|x) হবে।

$$P(1<y<3 \mid x=1) = \int_1^3 f(y|x)\,dy$$

conditional function

$$= \frac{3}{২} \int_2^3 \frac{f(x,y)}{g(x)}\,dy$$

joint function

$2<y<4$ ছিল

$$g(x) = \int_2^4 f(x,y)\,dy$$

যে domainএ function defined

$$= \int_2^4 \frac{6-x-y}{8}\,dy$$

$$= \left[\frac{6-x}{8}\cdot y - \frac{y^2}{16}\right]_2^4$$

$$= \frac{6-x}{8} \times 2 - \frac{16}{16} + \frac{4}{16}$$

$$= \frac{6-x}{4} - 1 + \frac{1}{4}$$

$$= \frac{6-x-3}{4}$$

$$= \frac{3-x}{4} \implies \text{domain } 0 \text{ to } 2$$

$$g(x) = \begin{cases} \frac{3-x}{4}, & 0 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

$$P(1<y<3 \mid x=1) = \int_2^3 \frac{6-x-y}{8} \times \frac{4}{3-x}\, dy$$

$$= \int_2^3 \frac{6-x-y}{6-2x}\, dy$$

$$= \left[\frac{6-x}{6-2x} \cdot y - \frac{y^2}{12-4x} \right]_2^3$$

$$= \frac{6-x}{6-2x} - \frac{9}{12-4x} + \frac{4}{12-4x}$$

$$= \frac{6-x}{6-2x} - \frac{5}{12-4x}$$

$$x=1 \Rightarrow \quad \frac{5}{4} - \frac{5}{8} = \frac{5}{8}$$

$$f(y|x) = \frac{f(x,y)}{g(x)} = \frac{6-x-y}{8} \times \frac{4}{3-x}$$

$$= \begin{cases} \dfrac{6-x-y}{2(-x+3)} & 0<x<2, \quad 2<y<4 \\ 0 & \text{elsewhere} \end{cases}$$

$$P(1 < y < 3 \mid x = 1) = \int_1^3 f(y|x)\,dy$$

$$= \underbrace{\int_1^2 f(y|x)\,dy}_{=0} + \int_2^3 f(y|x)\,dy$$

4.43 The length of time, in minutes, for an airplane to obtain clearance for takeoff at a certain airport is a random variable $Y = 3X - 2$, where X has the density function

random variable

$$f(x) = \begin{cases} \frac{1}{4}e^{-x/4}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

exponential decay function

Find the mean and variance of the random variable Y.

$$E[Y] = \int_0^\infty g(x)\, f(x)\, dx$$

$$= \int_0^\infty (3x-2)\, \tfrac{1}{4}\, e^{-x/4}\, dx$$

$$= (3x-2)\,\frac{1}{4}\,\frac{e^{-x/4}}{-1/4} - \int 3 \cdot \frac{1}{4} \cdot \frac{e^{-x/4}}{-1/4}\, dx$$

$$= \Big[-(3x-2)\,e^{-x/4} + 3 \times -4\, e^{-x/4}\Big]_0^\infty$$

$$= \Big[-(3x-2)\,e^{-x/4} - 12\, e^{-x/4}\Big]_0^\infty$$

$= +(-2) + 12$

$= 10$

variance,

$$\sigma_X^2 = E\left[(X-\mu)^2\right]$$

$$\sigma_{g(x)}^2 = E\left[\left\{g(x) - \mu_{g(x)}\right\}^2\right]$$

Here $g(x) = 3x - 2$

$$E\left[(3x-2-10)^2\right] = \frac{1}{4}\int_0^\infty (3x-12)^2 \frac{1}{4} e^{-x/4}\,dx$$

$\boxed{g(x) - A}$

$$= (3x-12)^2 \cdot \frac{e^{-x/4}}{-1/4} - \int 6(3x-12)\cdot\frac{e^{-x/4}}{-1/4}\,dx$$

$$= (3x-12)^2 \cdot (-4)\, e^{-x/4} + 24\int (3x-12)\, e^{-x/4}\,dx$$

$$= \Big[(3x-12)^2 (-4)\, e^{-x/4} + 24\,(3x-12) \times (-4)\, e^{-x/4}$$

$$\boxed{-\,24\int 3 \times (-4)\, e^{-x/4}\,dx}$$

$$-\,24 \times -12 \times -4\; e^{-x/4}\Big]_0^\infty$$

$$= \left[12^2 \times -4 + 24 \times (-12) \times -4 - 24 \times -12 \times -4\right]$$

$$= 576/4 = 144$$

Ans: 144

Theorem 4.5: If a and b are constants, then

$$E(aX + b) = aE(X) + b.$$

random variable এর linear combination থাকলে তার expectation হবে individual expectation এর linear combination এর সমান।

Theorem 4.6: The expected value of the sum or difference of two or more functions of a random variable X is the sum or difference of the expected values of the functions. That is,

$$E[g(X) \pm h(X)] = E[g(X)] \pm E[h(X)].$$

Theorem 4.7: The expected value of the sum or difference of two or more functions of the random variables X and Y is the sum or difference of the expected values of the functions. That is,

$$E[g(X, Y) \pm h(X, Y)] = E[g(X, Y)] \pm E[h(X, Y)].$$

Theorem 4.8: Let X and Y be two independent random variables. Then

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

independent না হলে def^n থেকেই করা লাগবে।

Theorem 4.9: If X and Y are random variables with joint probability distribution $f(x, y)$ and a, b, and c are constants, then

$$\sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

covariance

Corollary 4.8: Setting $b = 0$ and $c = 0$, we see that

$$\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2.$$

$$E[aX] = \int_{-\infty}^{\infty} a x f(x)\, dx$$

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

$$\sigma^2_{ax} = a^2 \sigma^2_x \qquad \mu_{ax} = E[ax] = aE[x]$$

$$E[(aX - \mu_{ax})^2] = E[(ax - a\mu_x)^2]$$

$$= a^2 \boxed{E[(x-\mu_x)^2]}$$

variance

05.11.24

Correlation:

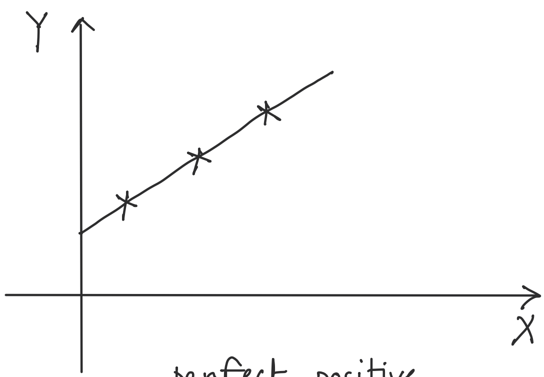
It measures the linear relationship between two or more random variables. It also measures the direction and strength of relationship between variables

direction → +ve/−ve

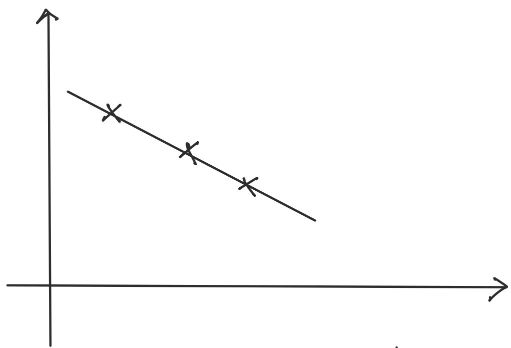
Types of correlation:

1. positive or negative
2. simple or multiple — more than 2
   simple → 2 variable
3. linear or non-linear

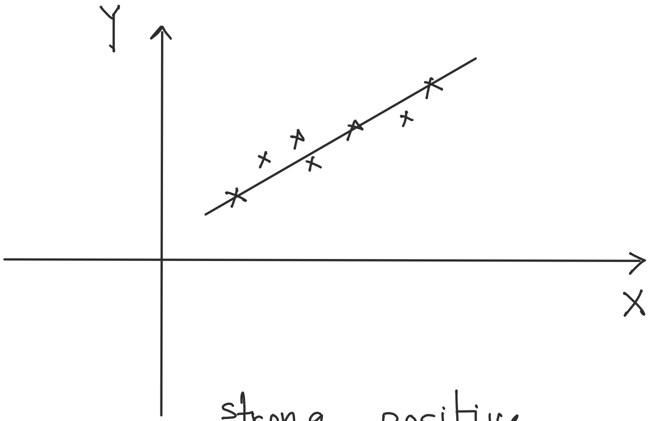
Graphical method to measure correlation: (Scatter diagram)



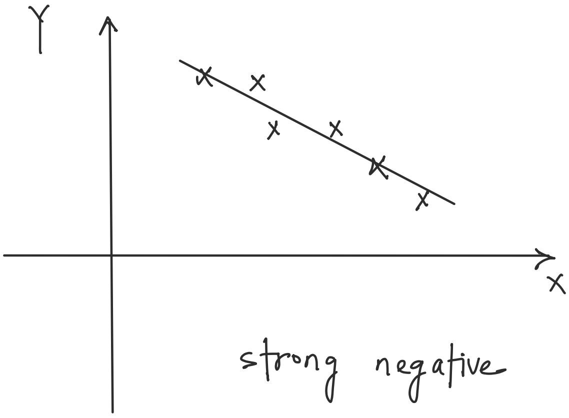
perfect positive



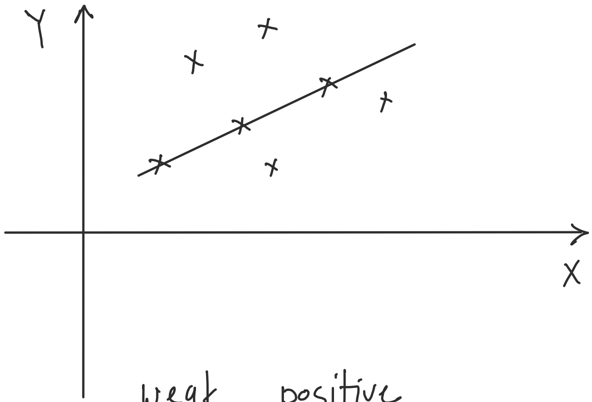
perfect negative



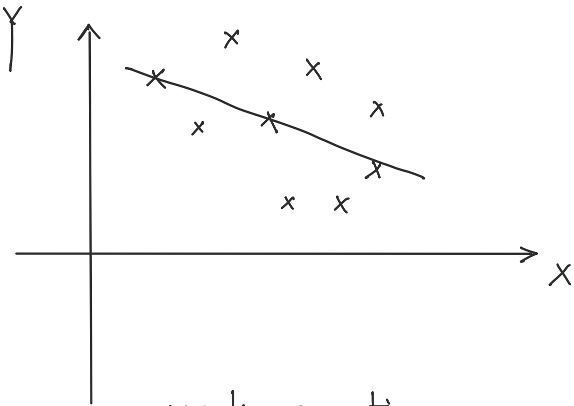
strong positive



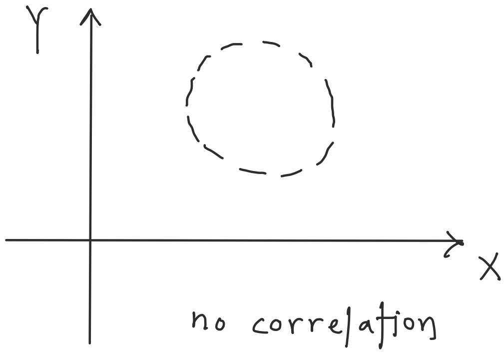
strong negative



weak positive



weak negative



no correlation

→ no linear correlation.

strong — maxm data straight line এর আশেপাশে

weak → straight line থেকে দূরে।

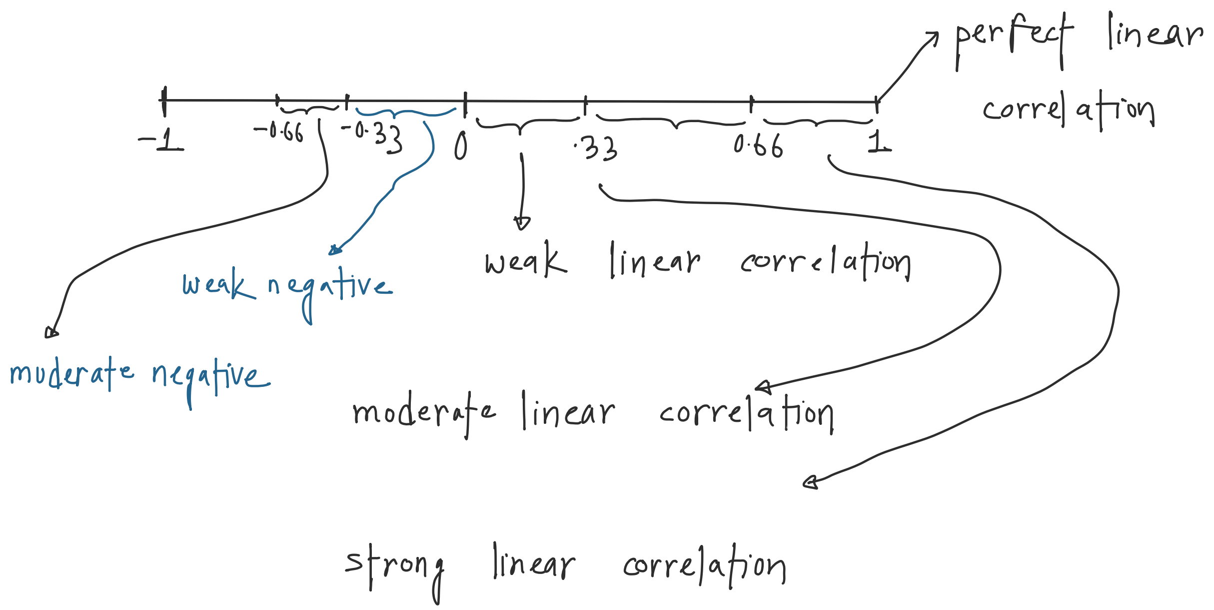
Mathematical method to measure correlation:

by karl pearson correlation coefficient

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2}\sqrt{\sum (y-\bar{y})^2}} = \frac{\text{covariance}}{\text{standard deviation}}$$

random variable এর

It can be proved that $-1 \leq r \leq 1$



Q. Find Karl-Pearson correlation coefficient between the sales and expenses from the table :

Sales (lakhs), X	50	50	55	60	65	65	65	60	60	50
Expenses (lakhs), Y	11	13	14	16	16	15	15	14	13	13

$X \quad Y \quad (X-\bar{X}) \quad (Y-\bar{Y}) \quad (X-\bar{X})(Y-\bar{Y}) \quad (X-\bar{X})^2 \quad (Y-\bar{Y})^2$

random variable catagorical / qualitative type এর হলে —
Karl pearson

let $r = 0.92$
↳ strong linear

0.6
↳ moderate linear

Spearman's Rank Correlation coefficient:

→ quantitative / qualitative ২ ধরনের data র জন্যই কাজ করে।

$$r_s = 1 - \frac{6\sum D^2}{N^3 - N}$$

where N = no. of cases

D = difference in ranks

Tie in Rank:

$$r_s = 1 - \frac{6}{N^3 - N}\left\{\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \dots\right\}$$

Q.

Marks in Bangla	R_1	Marks in English	R_2	$D^2 = (R_1 - R_2)^2$
15	2	40	6	16
20	$\frac{3+4}{2} = 3.5$ ~~3~~ tie	30	$\frac{3+4+5}{3}$ 4	⋮
28	5 → 4 দিতে পারবো না।	50	7	
12	minimum এর rank 1 দিলাম 1	30	4	
40	6	20	2	
60	7	10	1	
20	3.5 ~~3~~	30	4	
80	8	60	8	

R_2 → R_1 এ minimum কে 1 ধরা হয়েছে। এখানেও তাই করা লাগবে। maximum কে 1 ধরেও করা যেত। but R_1, R_2 তে same ভাবে করা লাগবে।

Bangla, English marks এর মাঝে correlation আছে কিনা।

20 2 বার আছে $m_1 = 2$

30 3 বার " $m_2 = 3$

আরো repeat হলে $m_3, m_4 \dots$

Ans. 0

rank correlation coefficient = 0

→ no correlation

independent.

12.11.24

Regression Analysis :

(i) To estimate the relationship that exists, on the average, between the dependent variable and independent variable.

(ii) To determine the effect of each independent variable on the dependent variable controlling the effects of the others independent variables.

(iii) To predict the value of dependent variable for a given value of the explanatory variables.

Simple linear regression model :

$$\hat{Y} = bX + a$$

where b = slope of the line

= regression coefficient of Y on X

a = intercept.

the straight line $\hat{Y} = bX + a$ is called the best fitted line of Y for the given data.

The least Square Method:

$$\sum e_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (y_i - bx_i - a)^2$$

$\sum e_i^2$ → sum of square of errors (SSE)

partial differentiat-ion

$$\frac{\partial (SSE)}{\partial a} = -2 \sum (Y_i - bX_i - a) = 0$$ slope $= 0$

$$\frac{\partial (SSE)}{\partial b} = -2 \sum (Y_i - bX_i - a) X_i = 0$$

$\sum Y_i = \sum b x_i + \sum a$

$$\sum Y_i = na + b \sum X_i$$

$\sum x_i Y_i = \sum b x_i^2 + \sum a x_i$

$$\sum x_i y_i = a \sum X_i + b \sum X_i^2$$

X_i
Y_i

$\sum X_i, \sum Y_i, \sum X_i^2, \sum x_i y_i$ বের করতে পারবো।

let

$2 = 10a + 5b$

$6 = 5a + 25b$

Solve করে a, b পাবো। যেটা দিয়ে best fitted line টা পাবো।

$$b = \frac{n \sum x_i y_i - \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

$$a = \bar{Y} - b\bar{X}$$

Regression line of Y on X: $Y - \bar{Y} = b_{yx} (X - \bar{X})$

where $b_{yx} = r \dfrac{\sigma_y}{\sigma_x} = \dfrac{\sum xy}{\sqrt{\sum x^2} \sqrt{\sum y^2}} \cdot \dfrac{\sqrt{\sum y^2}}{\sqrt{\sum x^2}}$

r

regression coefficient

correlation coefficient

$$= \frac{\sum xy}{\sum x^2}$$

$\dfrac{\text{Covariance}}{\text{variance of } x}$

$$\hat{Y} = b_1 X_1 + b_2 X_2 + a \qquad b_1, b_2, a \rightarrow 3\text{টা unknown}$$

২এর বেশি variable. multiple হবে। (simple না)

↳ multiple linear regression

$$\sum Y_i = \sum b_1 x_{1i} + \sum b_2 x_{2i} + na$$

$$\sum X_{1i} Y_i = \sum b_1 X_{1i}^2 + \sum b_2 X_{1i} X_{2i} + \sum a X_{1i}$$

$$\sum X_{2i} Y_i = \sum b_1 X_{1i} X_{2i} + \sum b_2 X_{2i}^2 + \sum a X_{2i}$$

$\underbrace{\hat{Y}}_{\text{salary}} = b_1 \overbrace{X_1}^{\text{age}} + b_2 \overbrace{X_2}^{\text{working hour}} + a$

$$Y = 0.6 X_1 + 0.9 X_2 + 10$$

interpret কর।

$X_1 = 0 \rightarrow$ কাজ আজকেই শুরু করেছে

$X_2 = 0$

$a = 10 \Rightarrow$ minimum salary

Age এর effect জানতে চাই। 1 বছর। (working hour fixed)

↳ আগে 10hr কাজ করলে এখনো 10hr কাজ করছে।

$X_1 = 1$

$X_2 = 0$

$b = 10$

$$Y = 10.6$$

working hour এর effect জানতে $X_1 = 0$

$$Y = 10.9$$

$\therefore$ working hour এর effect বেশি।

Regression line of X on Y :

$$X - \bar{X} = b_{xy} (Y - \bar{Y})$$

where $b_{xy} = r \dfrac{\sigma_x}{\sigma_y} = \dfrac{\Sigma xy}{\Sigma y^2}$ $\dfrac{\text{covariance}}{\text{variance of } y}$

Age, X	25	30	35	27	45	50
Salary, Y	10000	20000	50000	40000	70000	80000

$X - \bar{X} = x$

$Y - \bar{Y} = y$

$$r = \frac{(X-\bar{X})(Y-\bar{Y})}{\sqrt{\Sigma(X-\bar{X})^2}\sqrt{\Sigma(Y-\bar{Y})^2}} = \frac{xy}{\sqrt{x^2}\sqrt{y^2}}$$

pearson correlation coefficient.

let $r = 0.9$ ⟹ random variable X and Y positively related

↳ strong positive linear correlation between X and Y

$r = 0.7$ ⟹ moderate

যার age 26 তার salary কত? — table থেকে পাচ্ছি না।

↳ regression use

(value predict/estimate করতে)

Age — independent variable

Salary — dependent "

প্রতিটা independent variable এর কি effect আছে dependent variable এর উপর

simple — ২টা variable এর মধ্যে

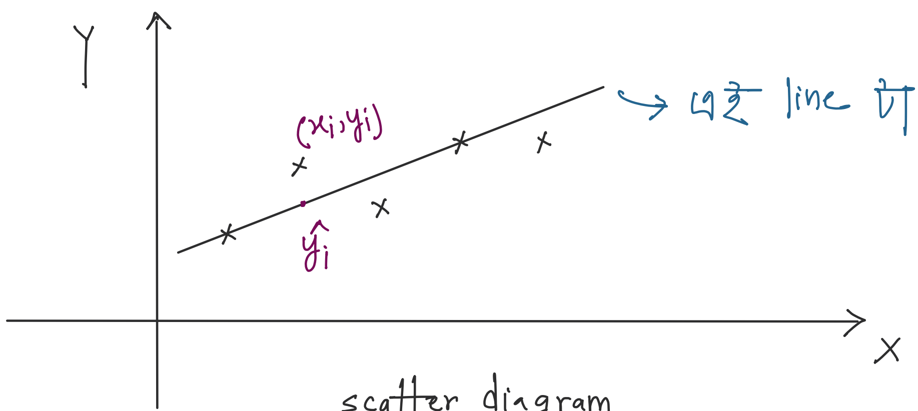
x	1	2	3	4
Y	5	8	4	9

$Y = mx + c$

↳ function /relation বের করা লাগবে, যেখানে independent variable এর value বসালে dependent টা জানতে পারবো।

regression — predict করতে পারছি।

— independent 1 unit change dependent কতটুকু change হচ্ছে সেটাও বের করতে পারি।



→ এই line টা value গুলোকে best represent করে।

scatter diagram

x_i, y_i → observed

$\hat{y}_i$ — estimated

$$y_i - \hat{y}_i \longrightarrow \text{error}$$

$$\Sigma (x - \bar{x}) = 0$$

mean

তাই difference কে square করে sum করা লাগবে।

13. 11. 24

Graphical representation is a way of analysing numerical data. It exhibits the relation between data, ideas, information and concepts in a diagram. There are different types of graphical representation. Some of them are as follows:

Line graphs: Line graph is used to display the continuous data and it is useful for predicting future events over time.

Bar Graphs: Bar graph is used to display the catagory of data and it compares the data using solid bars to represent the quantities.

Histograms: The graph that uses bars to represent the frequency of numerical data that are organised into intervals. Since all the intervals are equal and continuous, all the bars have the same width.

Circle graph: Also known as the piechart that shows the relationships as the parts of the whole. The circle is considered with 100% and the catagories occupied is represented with the specific percentage like

15% , 56% etc.

Stem and leaf plot: In the stem and leaf plot, the data are organized from least value to the greatest value. The digits of the least place values from the leaves and the next place value digit forms the stems.
A stem and leaf plot also called a stem and leaf diagram is a way of organising data into a form that makes it easy to observe the frequency of different types of values. Each data value is broken into a stem and a leaf.

This "|" symbol is used to show stem values and leaf values and it is called as stem and leaf plot key.
For example 25 is represented as 2 on the stem and 5 on the leaf and shown using stem and leaf plot key like this 2|5.

In the image given below the stem values are listed one below the other in ascending order and the list values are listed left to right from the stem values in descending order.

As the stem and leaf plot definition states:

6|7 → 6 on the stem and 7 on the leaf read as 67

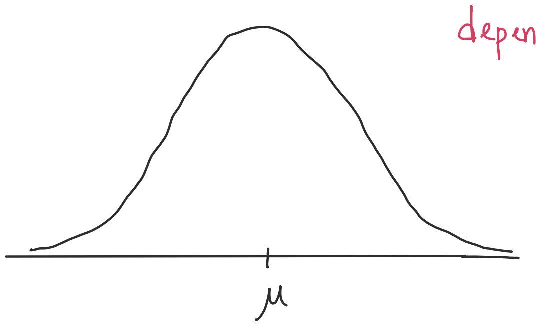
6|8 → 6 on " " " 8 " " " " " 68

9|0 → 9 on " " " 0 " " " " " 90

Normal distribution :

→ mean, standard deviation এর উপর depend করে

19.11.24



$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where $-\infty < x < \infty$

$X \sim n(x, \mu, \sigma) \rightarrow X$ normal distribution follow করে

(i) $E(X) = \mu$ (ii) $Var(X) = \sigma^2$

Proof (i):

$$E(X-\mu) = \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{a}\right)^2} dx$$

$E(X) - E(\mu) = 0$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \sigma z \, e^{-\frac{1}{2}z^2} \sigma dz$$

let,

$\frac{x-\mu}{\sigma} = z$

$\rightarrow dx = \sigma dz$

$\therefore E(X) = \mu$

$$= \frac{\sigma^2}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} z e^{-\frac{1}{2}z^2} dz$$

$$= 0 \quad \left[\because f(z) = z e^{-\frac{1}{2}z} \text{ is odd function}\right]$$

$\therefore E(x) = \mu$

Proof (ii)

Now $Var[x] = E\left[(X-\mu)^2\right]$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \sigma^2 z^2 \, e^{-\frac{1}{2}z^2} \sigma \, dz$$

$$= \frac{2\sigma^3}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} z^2 \, e^{-\frac{1}{2}z^2} \, dz \qquad \left[\because f(z) = z^2 e^{-\frac{1}{2}z^2} \text{ is even function}\right]$$

$$= \frac{2\sigma^3}{\sqrt{2\pi\sigma^2}} \int_{0}^{\infty} 2u \, e^{-u} \frac{du}{\sqrt{2u}}$$

let, $\frac{1}{2}z^2 = u$

$\rightarrow \frac{1}{2} \cdot 2 \cdot z \, dz = du$

$\rightarrow z \, dz = du$

$\rightarrow dz = \frac{du}{z}$

$= \frac{du}{\sqrt{2u}}$

$$= \frac{2\sigma^3}{\sqrt{2\pi\sigma^2}} \cdot \frac{1}{\sqrt{2}} \int_{0}^{\infty} e^{-u} \cdot u^{\left(\frac{1}{2}+1\right)-1} du$$

$u^{1-\frac{1}{2}}$

$e^{-u} \, u^{n-1} = \Gamma n$

$\Gamma \frac{3}{2} = \frac{1}{2}\Gamma \frac{1}{2} = \frac{1}{2}\sqrt{\pi}$

$$= \frac{4\sigma^3}{\sqrt{2\pi\sigma^2}} \frac{1}{\sqrt{2}} \cdot \frac{1}{2} \cdot \sqrt{\pi} = \sigma^2$$

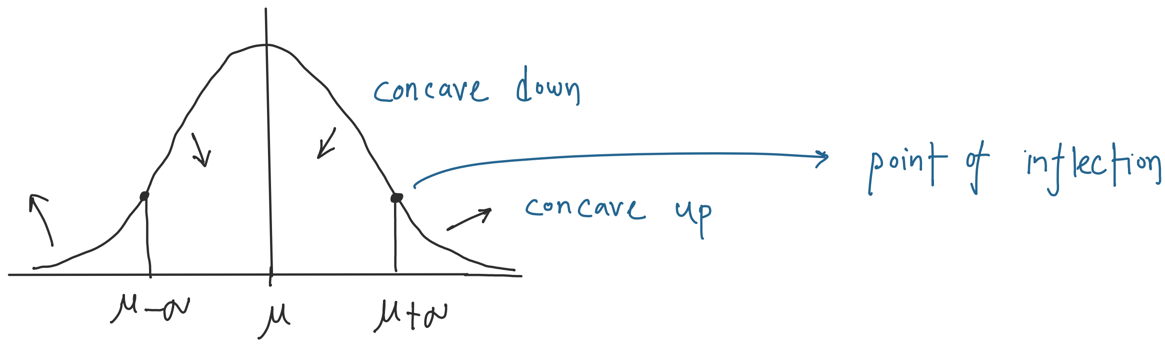
$\text{Var}(x) = \sigma^2$

$*\quad \int_{0}^{\infty} e^{-x} x^{n-1} \, dx = \Gamma(n)$

$*\quad \Gamma(n+1) = n \Gamma n$

$*\quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$

$$*\quad \int_{-\infty}^{\infty} f(x) \, dx = \begin{cases} 2\int_{0}^{\infty} f(x)\,dx & \text{when } f \text{ is even} \\ 0 & \text{when } f \text{ is odd} \end{cases}$$



$$\int_{-\infty}^{\infty} n(\quad) = 1$$ (over the domain integration)

$$Z = \frac{X-\mu}{\sigma}$$ here $X \sim n(x, \mu, \sigma)$

and $z \sim n(z, \mu = 0, \sigma = 1)$

$$E[z] = E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}\{E[X] - E[\mu]\}$$

$$= \frac{1}{\sigma}\{\mu - \mu\} = 0$$

$$E[z^2] = E\left[\left(\frac{x-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} E\left[(x-\mu)^2\right] - \frac{1}{\sigma^2} \cdot \sigma^2 = 1$$

$$\therefore Var[z] = E[z^2] - \{E[z]\}^2 = 1 - 0 = 1$$

$$\therefore P(X_1 < X < X_2) = p\left(\frac{x_1-\mu}{\sigma} < \frac{x-\mu}{\sigma} < \frac{x_2-\mu}{\sigma}\right)$$

let $\frac{x_1-\mu}{\sigma} = z_1$

and $\frac{x_2-\mu}{\delta} = z^2$

$$= P(z_1 < z < z_2)$$

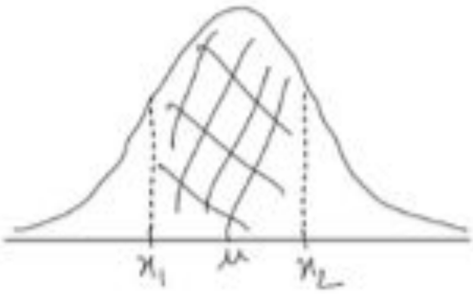
$$= P(z < z_2) - P(z < z_1)$$

integration over the domain 1 পাবো। (probability 1) inequality : $x_1 < X < x_2$

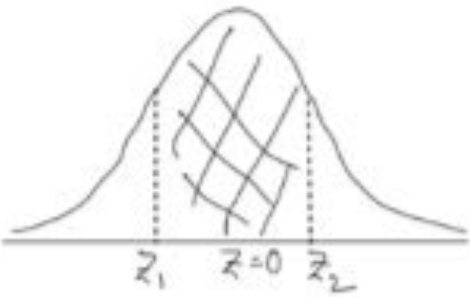
$$\Rightarrow x_1 - \mu < X - \mu < x_2 - \mu$$

$$\Rightarrow \frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_2 - \mu}{\sigma} \qquad [\sigma \text{ positive}]$$

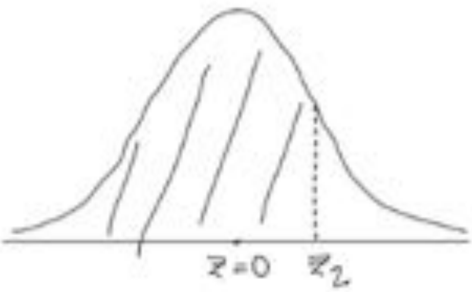
$$\Rightarrow z_1 < z < z_2$$



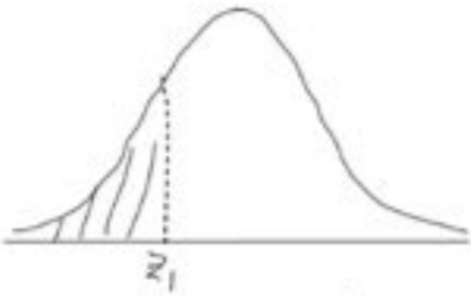
$P(x_1 < X < x_2)$



$P(z_1 < z < z_2)$



$P(z < z_2)$



$P(z < z_1)$

$$P(z < z_2) - P(z < z_1) = P(z_1 < z < z_2)$$

$$\int_{-5}^{3} f(x)\,dx = \int_{-\infty}^{3} f(x)\,dx - \int_{-\infty}^{-5} f(x)\,dx$$

$$\int_{-5}^{3} f(x)\,dx = \int_{-\infty}^{3} f(x)\,dx - \int_{-\infty}^{-5} f(x)\,dx$$

Example 6.7: A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

***Solution*:** First construct a diagram such as Figure 6.14, showing the given distribution of battery lives and the desired area. To find $P(X < 2.3)$, we need to evaluate the area under the normal curve to the left of 2.3. This is accomplished by finding the area to the left of the corresponding z value. Hence, we find that

$$z = \frac{2.3 - 3}{0.5} = -1.4,$$

$P(X<2.3)$

$P\left(\frac{X-\mu}{\sigma} < \frac{2.3-3}{0.5}\right)$

$P(Z<-1.4)$

$= \int_{-\infty}^{-1.4} \frac{1}{\sqrt{\ }} e^{-\frac{1}{2^2}} dz$

and then, using Table A.3, we have

$$P(X < 2.3) = P(Z < -1.4) = 0.0808.$$

complex integration table থেকে value পাবো

−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681

$P(Z < -1.4) = 0.0808$

যদি $Z < -1.45$ হতো →

$\therefore P(Z < -1.45) = 0.0735$

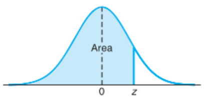
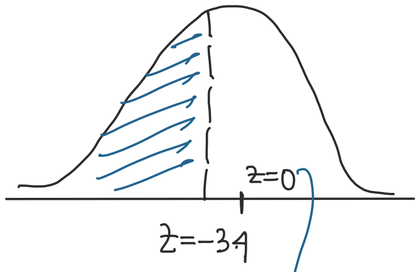


Table A.3 Areas under the Normal Curve

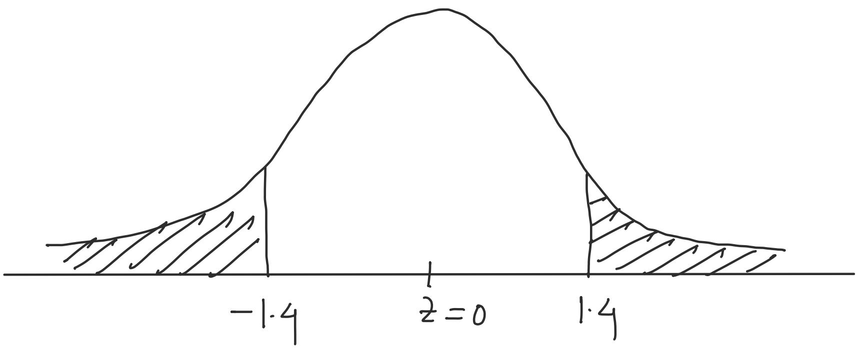
z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
−3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
−3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
−3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
−3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
−0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

$P(z < -3.4) = 0.0003$



$1 - 0.0003$

-1.40
0.05 (deviation)
-1.45



$$P(z < -1.4) = P(z > 1.4)$$
$$= 1 - P(z < 1.4)$$
$$= 1 - 0.9192$$
$$= 0.0808$$

Q: probability দেওয়া থাকবে z, x বের করতে হবে।