DATE
PAGE

Ch1 Introduction to Probability and Statistics

A random variable is represented by Symbol X, Y or H is an outcome of an experiment.

For eg –
Rolling of a fair dice is an experiment, then random variable X can take one of value as {1, 2, 3, 4, 5, 6}

Tossing of an unbiased coin is an experiment, then random variable Y can take one of value as {H, T}

In dice roll experiment $X = \{1, 2, 3, 4, 5, 6\}$

Probability that random variable X can take value 1 is written as

$$P(X=1) = \frac{1}{6} = P(X=2)$$

$$P(X=\text{even}) = \frac{1}{2} \text{ or}$$

$$P(X=2) + P(X=4) + P(X=6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

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

Similarly $P(X=\text{odd}) = \frac{1}{2}$

In Short we can write $P(X=x_1)$ can be written as $P(x_1)$.

Discrete Random Variable

Discrete It is a random variable which can take value from a finite set of variables

For eg - Experiment of rolling of a dice has finite set values as - $X = \{1, 2, 3, 4, 5, 6\}$.

Continuous Random Variable

Continuous It is a random variable which can take any value in a range in an experiment.

For eg - Height of a randomly picked student
Y varies between 120cm - 190cm
Y can take value as 162.45 cm
or Y can take value as 132.62 cm

Outlier observation.

Human error or Observation error in our observations.

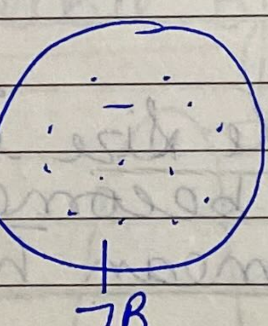
For eg—

Y = Height of a Student

$\{122.2,\ 146.2,\ 132.5, \ldots, 12.26,\ 156.23\}$

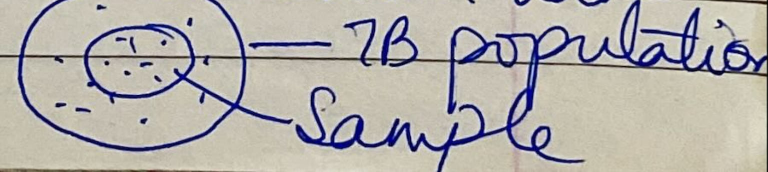
12.26 ↓ outlier

Ch2 Population and Sample

Suppose this  is set of all the people in the world, ie, the population of the world

mean of the population represented as $\mu = \frac{1}{7B}\sum_{i=1}^{7B} h_i$ but impossible to calculate

In order to estimate the mean height of a human we can take a sample.



DATE
PAGE

Sample is the subset of the population.

Size of sample = 1000

Randomly picked sample is called Random sample.

$$\bar{h} = \frac{1}{1000}\sum_{i=1}^{1000} h_i$$

Mean of a sample.

As sample size increases the population mean μ becomes equal to sample mean $\bar{h}$.

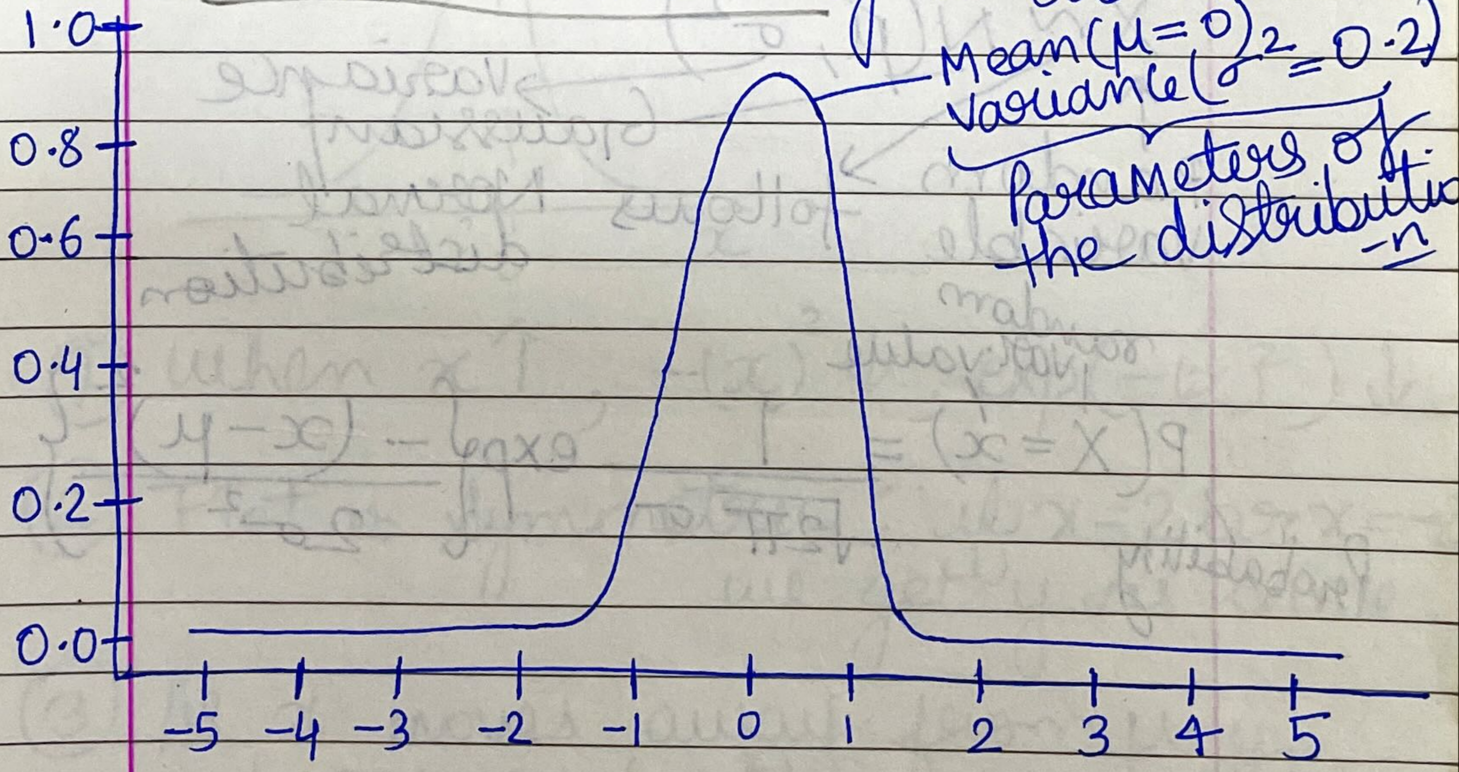
The height in this experiment is termed as the property or statistic.

Ch3 Gaussian/Normal distribution and its Probability Density Function (PDF)

The bell shaped curve is PDF of a Gaussian distribution.

Generally some natural data occur as Gaussian distribution like height, weight.

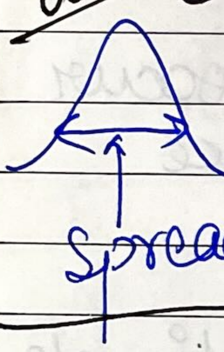
Gaussian distribution is a simple model to learn various properties of a random variable of a distribution



Given the mean and variance of a distribution (Gaussian) we can draw its PDF.

---

done ~~As~~ variance is the spread also called as scale when it is small the PDF is peaked and when it is large the PDF is Fatter and Shorter.



The peak of the PDF of a Gaussian distribution is at μ.

Mean

$$X \sim N(\mu, \sigma^2)$$

variance

Gaussian

Random Variable follows Normal distribution

random var. value

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

Probability

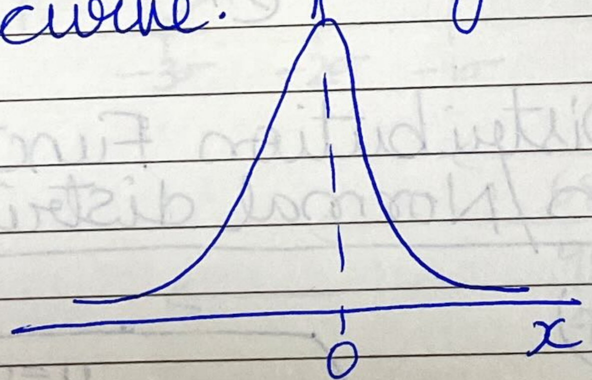
Let $\mu = 0, \sigma^2 = 0, \sigma = 0$

$$P(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\boxed{\frac{1}{2}}\, x^2\right\} = y$$

$\uparrow$ Constant $\qquad$ $\uparrow$ Constant

$$y = \exp\{-x^2\} = \frac{1}{e^{x^2}}$$

when we plot y we get a bell shaped curve.



1. when $x \uparrow$, $-(x)^2 \downarrow$, $\exp(-(x)^2) \downarrow$
2. Plot is symmetric $\because$ if $x = 2$ or $x = -2$ we get y as same
3. As x moves away from μ y reduces $\exp(-x^2)$ i.e. ~~exponentially~~ quadratically

$y = \exp(-x^2)$

$x = 0, \quad y = 1$

$x = 1, \quad y = \exp(-1) = \frac{1}{e^1} = 0.3678$

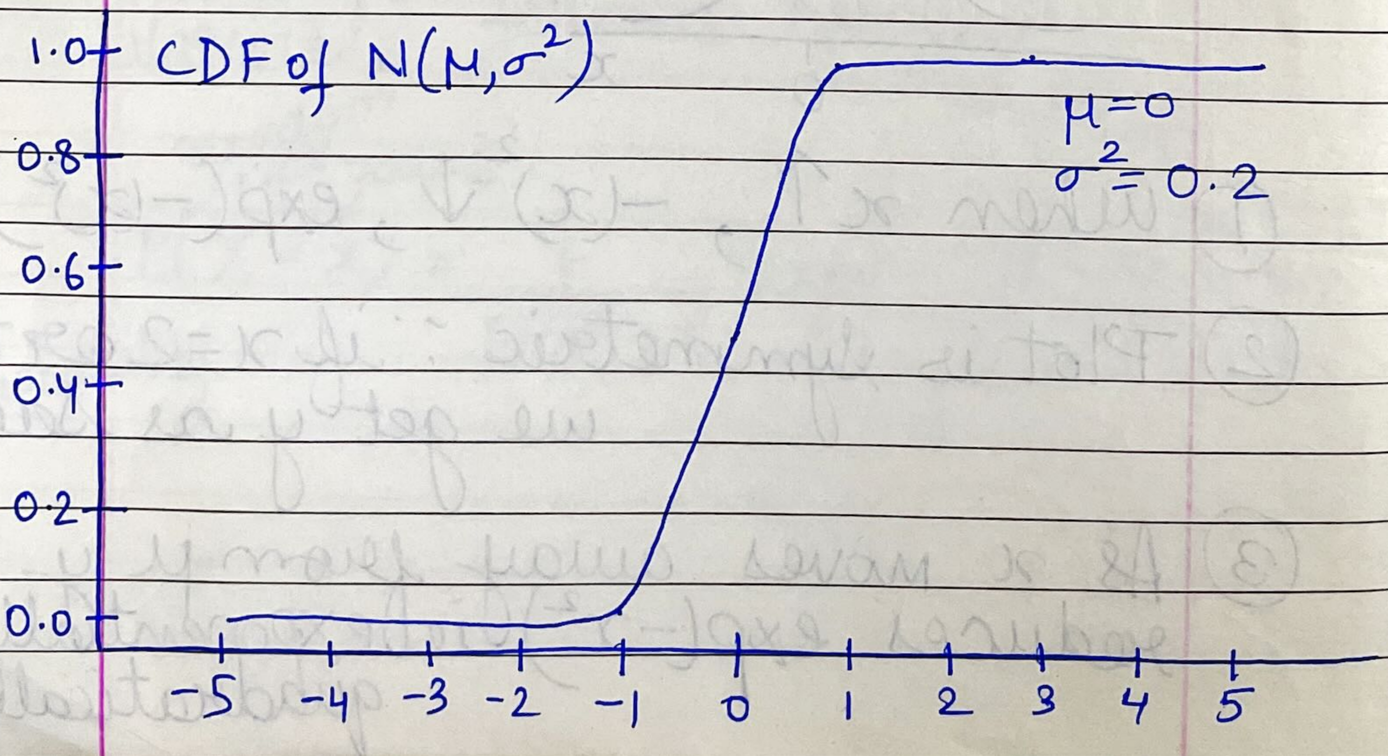
$\downarrow 2x$ $\qquad\qquad\qquad\qquad \downarrow 20x$

$x = 2, \quad y = \exp(-4) = \frac{1}{e^4} = 0.018$

$\downarrow 1.5x$ $\qquad\qquad\qquad\qquad \downarrow 100x$

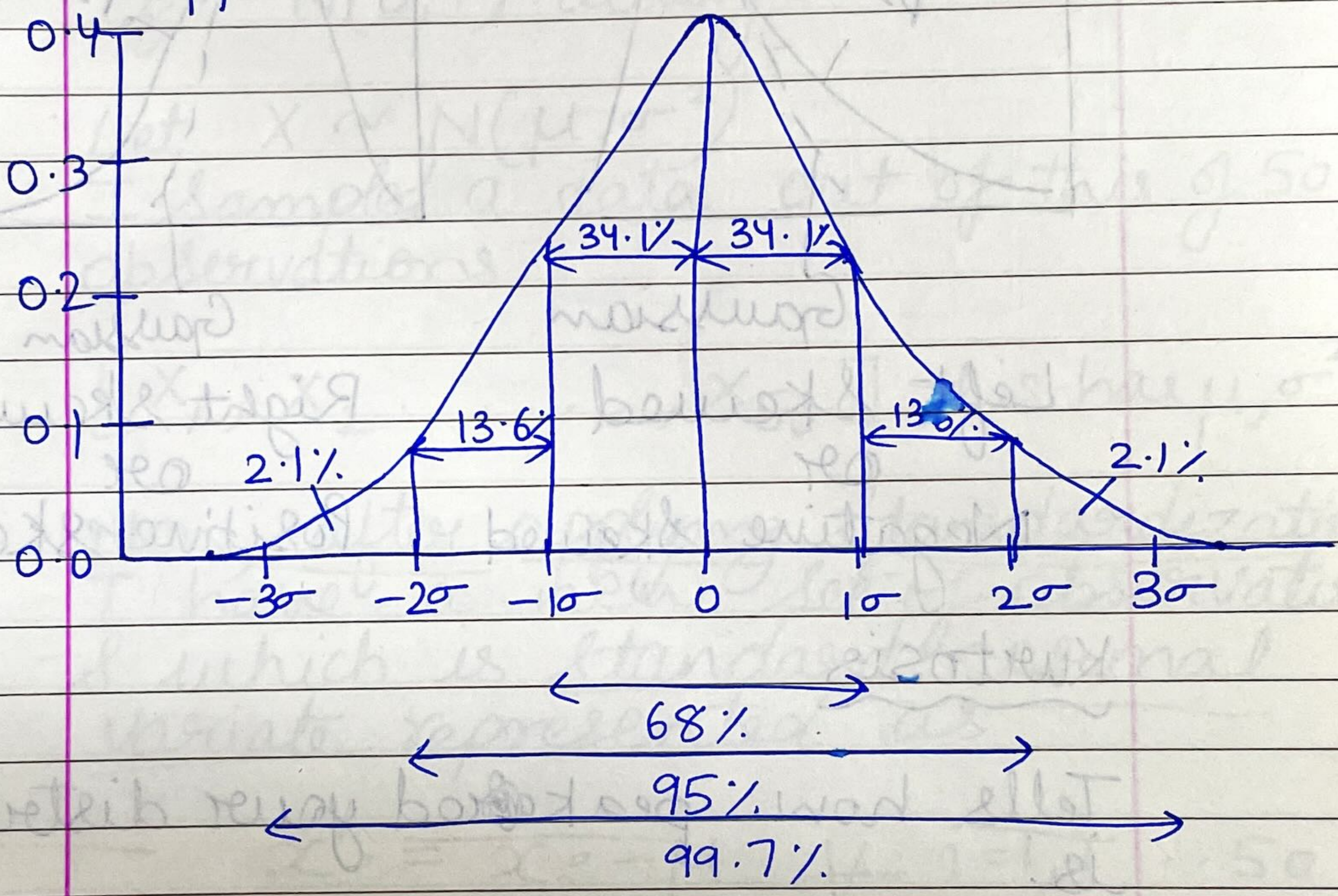
$x = 3, \quad y = \exp(-9) = \frac{1}{e^9} = 0.000123$

Ch4 Cumulative Distribution Function of Gaussian/Normal distribution



68-95-99.7 rule

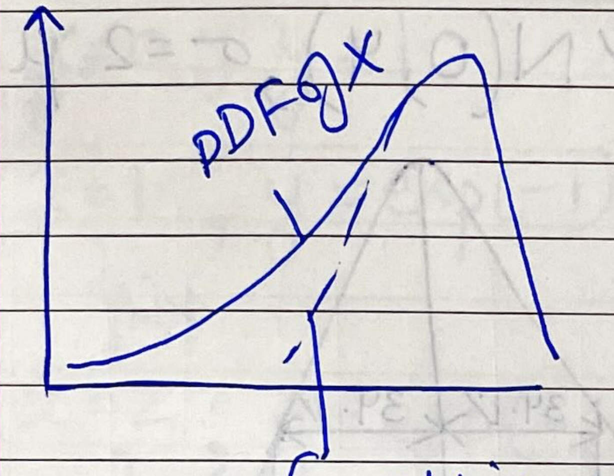
Suppose $X \sim N(0,4)$, $\sigma = 2$. $\mu = 0$.

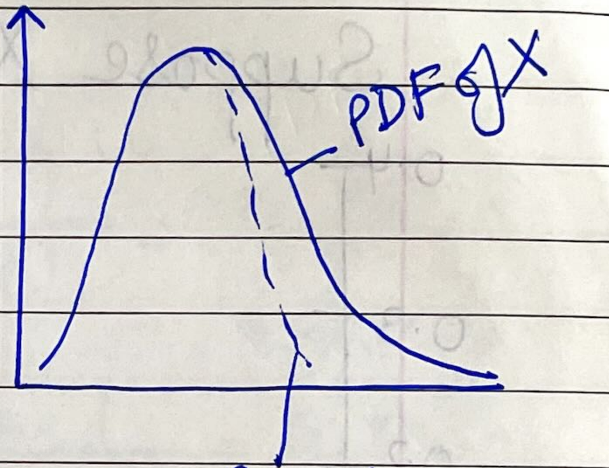


ch5 Symmetric distribution, Skewness and Kurtosis

We know symmetric distribution through Gaussian distribution where 50% of points lie of left side of $\mu = 0$ and 50% on right side.

Skewness



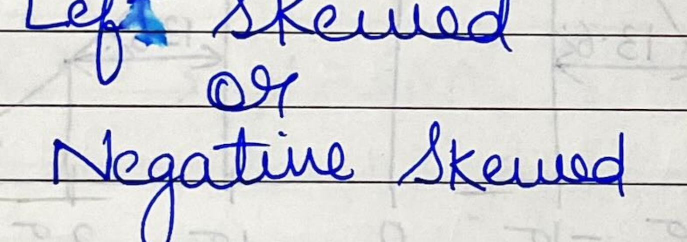


Gaussian	Gaussian
Left Skewed or Negative Skewed	Right skewed or Positive Skewed

Kurtosis

Tells how peaked your distribution is.

Gaussian distribution has kurtosis 3



etc.

Ch6 Standard Normal Variate (z) and Standardization

$z \sim N(0,1)$ always.

Let $X \sim N(\mu, \sigma^2)$
I sampled a data out of this of 50 observations

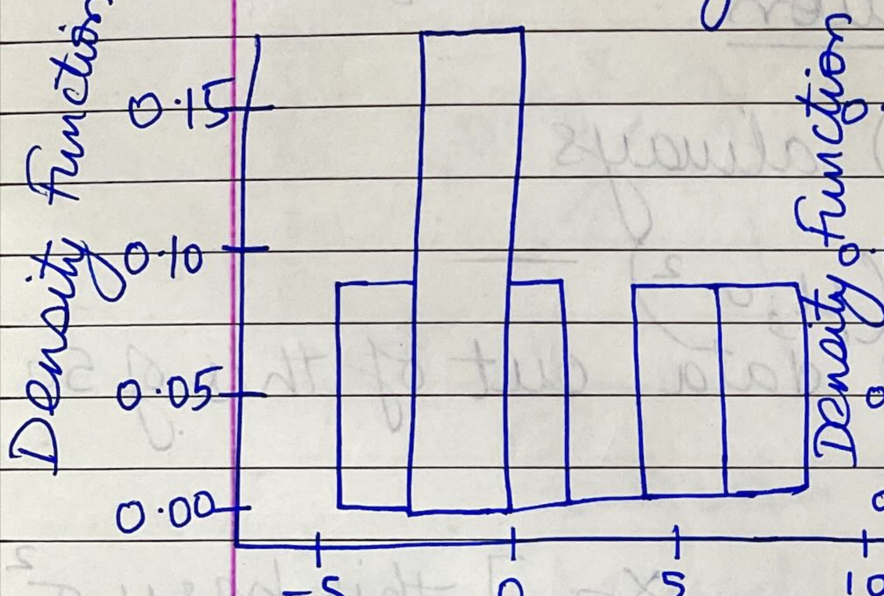
$[x_1, x_2 \ldots\ldots, x_{50}]$ this has μ, σ^2

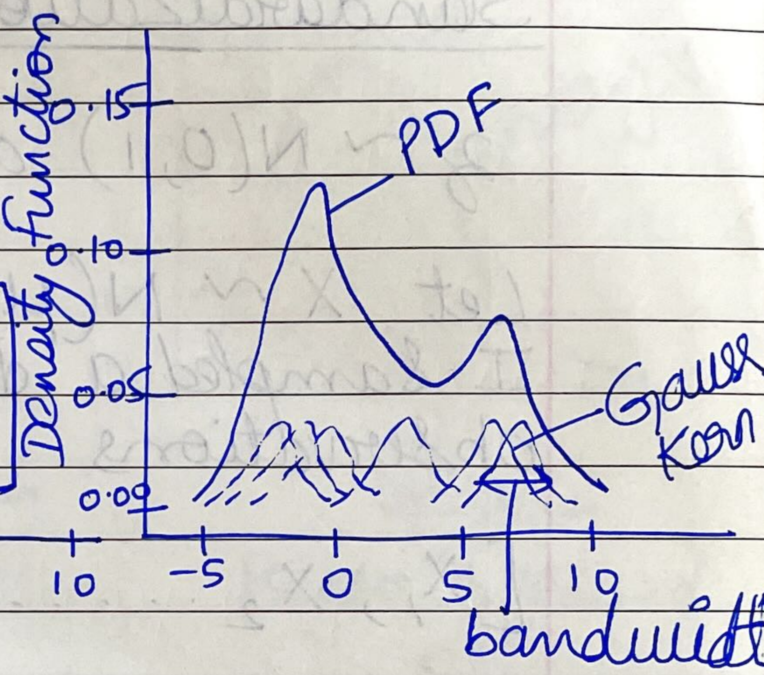
Now, after applying Standardization I have a new set of observation which is standard normal variate represented as

$$x_i' = \frac{x_i - \mu}{\sigma} \quad \forall \; i = 1, 2 \ldots 50$$

Here $x_i' \sim N(0,1)$ forall

Ch7 Kernel density estimation





PDF of a histogram is constructed using Kernel density estimation.

We build Kernels for each point which is also called a Gaussian Kernel for each point.

At any point x we sum up the Kernels to get the height of a point to get the PDF.

Note— As the no. of bins ↑ the PDF becomes jaggered.

If high density of points are there for kernels we get high peak in PDF.

Small — jaggered pdf
Large — fatter pdf

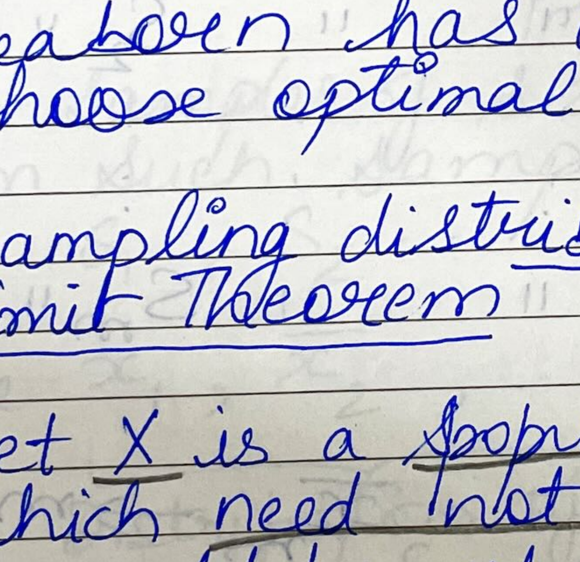
If you choose a small bandwidth you get jaggered PDF or large bandwidths you get a fatter PDF. So after ~~order to~~ choosing a medium bandwidth you get optimal PDF.

Seaborn has a method to choose optimal bandwidths.

Ch8 Sampling distribution & Central Limit Theorem

Let X is a population sample which need not be Gaussian or it could be where X represents distribution of incomes

X:



Now, we pick a random sample of size n say S_1

We again picked a random sample of size n say S_2.

We picked m such samples and the last one being S_m.

Let mean of sample S_1 is $\overline{x}_1$

" " " S_2 " $\overline{x}_2$

⋮

" " " S_m " $\overline{x}_m$

We have $\underbrace{\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_m}$ → m sample means

$\overline{x}_i$s have distribution called Sampling distribution of sample means.

There are some distributions such as pareto distributions have infinite mean or undefined. For them Central Limit Theorem does not apply.

Central Limit Theorem

Let the population distribution X has a finite μ and σ^2.

pop. mean   variance

If samples of size n are taken. m such samples are taken.

$S_1, S_2 \ldots S_m$

Sample mean $\underbrace{\bar{x}_1, \bar{x}_2, \ldots \bar{x}_m}$

Distribution of sample mean $\bar{x}_i$.

Also called Sampling distribution of sample mean.

Then Central Limit Theorem says:

$\bar{x}_i \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$ as $n \rightarrow \infty$

n → size of sample

Gaussian distribution

Sampling distribution of sample mean $\bar{x}$ has mean equal to population mean μ which is Gaussian distribution and variance equal to population variance divided by sample size. $\frac{\sigma^2}{n}$

Ch 9 Q-Q plot: How to test if a random variable is normally distributed or not?

probplot in Scipy

Q-Q stands for Quantile Quantile plot.

X is a random variable and there are 500 observations from X

$X: x_1, x_2, x_3, \ldots x_{500}$

We need to find whether X is Gaussian distribution or not?

For that there is Q-Q plot technique. Apart from Q-Q plot there are other techniques such as KS-Test and Anderson Darling Te

So to do QQ plot following are the steps ÷

1) Sort x_i and compute its percentile

$x_1, x_2 \dots x_{500}$

$\downarrow$ sort (ascending)

$x'_1, x'_2, \dots x'_{500}$

$\downarrow$ Percentile

So x'_5 — 1st percentile $\Rightarrow x^{(1)} \rightarrow$ percentile

x'_{10} — 2nd " $\Rightarrow x^{(2)}$

x'_{15} — 3rd " $\Rightarrow x^{(3)}$

:

:

x_{500} — 100^{th} percentile $\Rightarrow x^{(100)}$

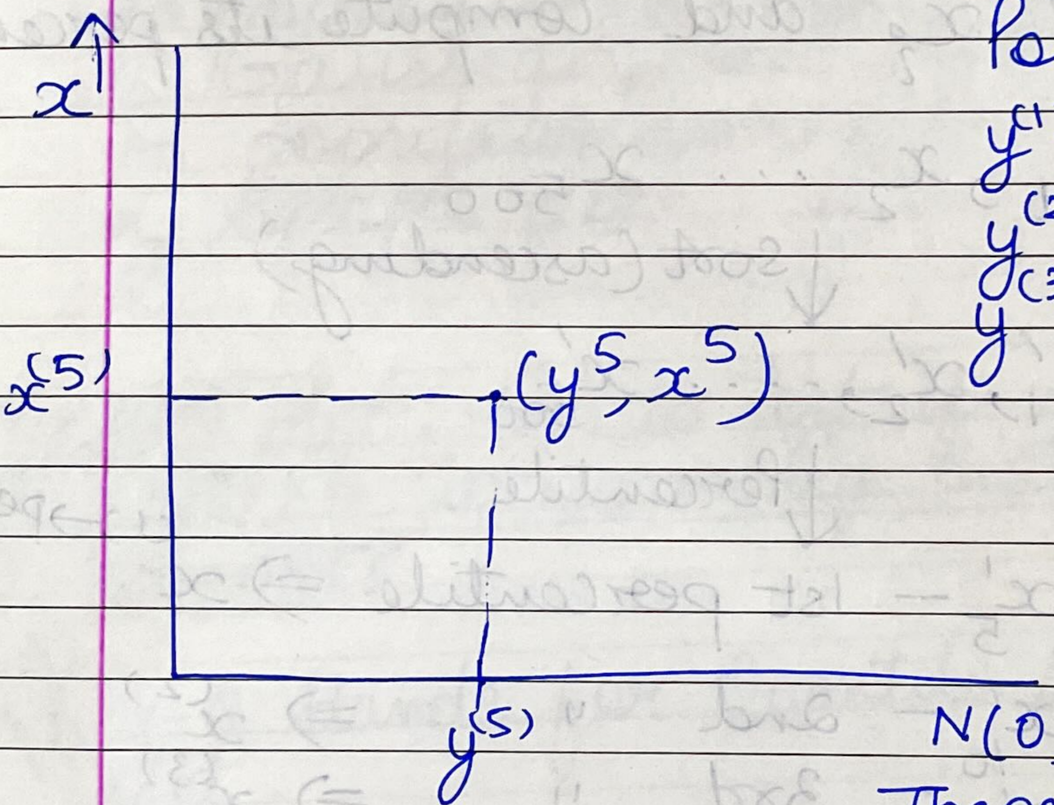
2) Create a random variable Y which belongs to $N(0,1) \rightarrow$ Standard Gaussian distribution

$y_1, y_2, y_3 \dots y_{100}$

$\downarrow$ Sort (asc)

$y'_1, y'_2, y'_3 \dots y'_{100} \xrightarrow{\text{percentile}} y^{(1)}, y^{(2)}, y^{(3)}, \dots y^{(100)}$

3) Plot Q-Q plot using $x^{(1)}, x^{(2)} \ldots x^{(100)}$
$y^{(1)}, y^{(2)} \ldots y^{(100)}$



Points;
$y^{(1)}, x^{(1)}$
$y^{(2)}, x^{(2)}$
$y^{(3)}, x^{(3)}$
100 pair

$N(0,1) : y \rightarrow$
Theoretical Quantiles

If $y^{(i)}, x^{(i)}$ $\forall$ $i \to 1$ to 100 lie on a straight line the random variable X and Y have similar distribution. are Gaussian distribution.

Limitations – If no. of observations are small it is difficult to read a QQ plot.

ch10 How distributions are used?

Q1 Company XYZ
Task:- To order t-shirts for all employees. (100K)
S, M, L, XL

How many XL t-shirts should you order?

Cost of collecting data of 100K employees is too high.

So, you get an idea

height ≥ 180 cm — XL t-shirts
[160, 180cm] — L t-shirts

Collect heights of 500 random employees
Sample mean and standard deviation

If height $\sim N(\mu, \sigma)$

I can easily compute Probability ($h \geq 180$ cm)
I can read this from my CDF

It gives us a theoretical model obsew in many natural phenomenon (Gaus Distribution).

Q2 Salaries →

If $S \sim N(\mu, \sigma)$

How many employees make a salary $\geq 100K\$$

How many employees make a salary between $50K\$ - 70K\$$.

I can answer all these questions from CDF.

You identify a distribution is a Gaussian distribution from dddplot.

ed
sian

Ch 11 Chebyshev's Inequality

We know the 68-95-99.7 rule for Gaussian distribution and that 68% of points lie between $\mu-1\sigma$ and $\mu+1\sigma$. 95% of points lie between $\mu-2\sigma$ and $\mu+2\sigma$. And 99.7% points lie between $\mu-3\sigma$ and $\mu+3\sigma$.

What if I don't know about the distribution, but I know the μ & σ which is finite and nonzero finite respectively.

If I have to answer what % of ~~peo~~ points lie b/w ($\mu-2\sigma$ to $\mu+2\sigma$) then we can use Chebyshev's inequality.

Chebyshev's Inequality States that:-

random variable.

$$P(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$$

or

$$P\begin{pmatrix} X \geq \mu+k\sigma \\ X \leq \mu-k\sigma \end{pmatrix} \leq \frac{1}{k^2}$$

DATE

PAGE

or

$$P\left(\mu - K\sigma < X < \mu + K\sigma\right) > 1 - \frac{1}{K^2}$$

Q If $\mu = 40K\$$ & σ $\$10K$.

Applying above formula to get probability

$$\$20K < X < 60K\$ > 1 - \frac{1}{2^2} \quad \because K = 2.$$

$$1 - \frac{1}{4} = \frac{3}{4} \approx 75\%$$

So, $[20K, 60K] > 75\%$

Ch12 Discrete and Continuous Uniform distributions

PMF ~~~~ mass

If a random variable is discrete and it follows uniform distribution, it is called Discrete Uniform Distribution

PDF ~~~~ den

If a random variable is continuous and it follows uniform distribution, it is called Continuous Uniform Distribution.

PDF and PMF (Probability Mass Function) are one and the same thing except one difference.

We write PDF for Continuous random Variables.

We write PMF for discrete random Variables.

For Discrete Random Variable :-

In a Uniform distribution, all values are equiprobable.

Parameters $\Rightarrow$ a, b.

$n = b - a + 1.$ $\quad PMF = \frac{1}{n}.$

P.T.O

$\left.\begin{array}{l}\text{Mean}\\ \text{Median}\end{array}\right\} \rightarrow \frac{a+b}{2}$

Skewness $\rightarrow 0$

For Continuous Random Variable :-

$PDF = \frac{1}{b-a}$ for $x \in [a, b]$

Parameters = a, b

$-\infty < a < b < \infty$

$\left.\begin{array}{l}\text{Mean}\\ \text{Median}\end{array}\right\} \rightarrow \frac{a+b}{2}$

Ch 13 How to randomly sample data points (Uniform Distribution)

Code generation

Ch 14 Bernoulli and Binomial Distribution

Only 2 outcomes in Bernoulli probability.
They are discrete distributions.

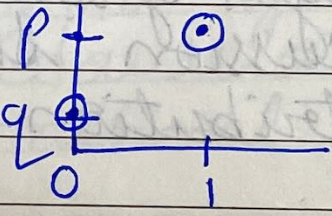
1 Bernoulli Random Variable

Bernoulli parameters $\Rightarrow 0 < P < 1$

$[k=0, \; k=1]$

PMF $\Rightarrow q \rightarrow (1-p), \; p$

$p + q = 1$



For $k=0$

$q = p-1$

For $k=1$

p

2 Binomial Random Variable

In a coin toss expt. we want to check no. of times we get head represented by Y

So, $Y \in \{0, 1, 2, \ldots 10\}$ If a fair coin is tossed 10 times.

So $Y \sim Bin(n, p)$ → Probability of getting head; n → number of trials.

Parameter of Binomial $n, \in N_0$
$p \in [0,1]$

PMF $\binom{n}{k} p^k (1-p)^{n-k}$
given using Combinatorics

ch 15 Log Normal Distribution

Random variable X is considered to be log-normal if natural log of X is normally distributed.

Length of comments posted in Internet discussion Shows log-normal distribution.

User's dwell time on the online articles also follows log normal distribution.

Also human behaviours and Income in economics follows log normal distribution.

How to determine if X follows log normal(μ, σ) distribution?

We take all datapoints:-

$x_1, x_2, x_3, \dots, x_n$

We take natural logarithm of them.

$\ln(x_1), \ln(x_2), \dots, \ln(x_n)$

$\downarrow \quad \downarrow \quad \downarrow$

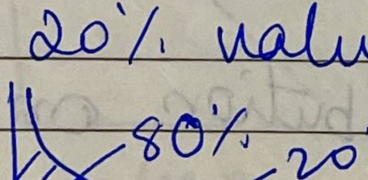
$y_1, y_2, \dots, y_n$

We draw a QQ plot of $y_1 - - y_n$.
If y_i's are Gaussian then X is also log-normal.

Ch16 Power Law distribution

They follow long tail in their plot and something called as the 80-20 rule.

Top 20% value follows 80% of mass



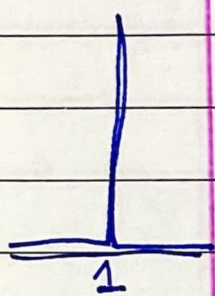
DATE

PAGE

If a distribution follows power law it is called Pareto distribution

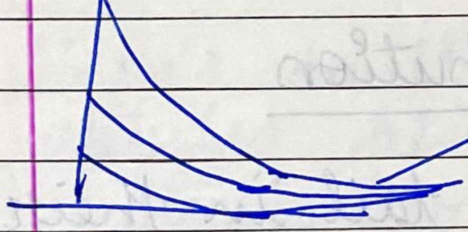
Parameters x_m – Scale
α – shape

As $\alpha \downarrow$ the tail becomes fatter.



when $\alpha = \infty$, the distribution approaches Dirac delta function where 1 value has a peak and everything else has 0.

But pareto distributions fall like this.



very ~~large~~ small value which are too much and few values that are too less.

eg- Size of no. of people in city, town, village.

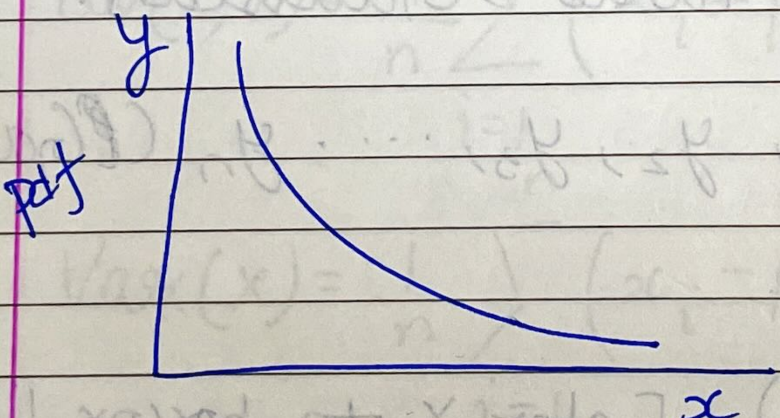
File size distribution on Internet.

Hard disk errors.

Value of oil reserves — few large fields and many many small fields.

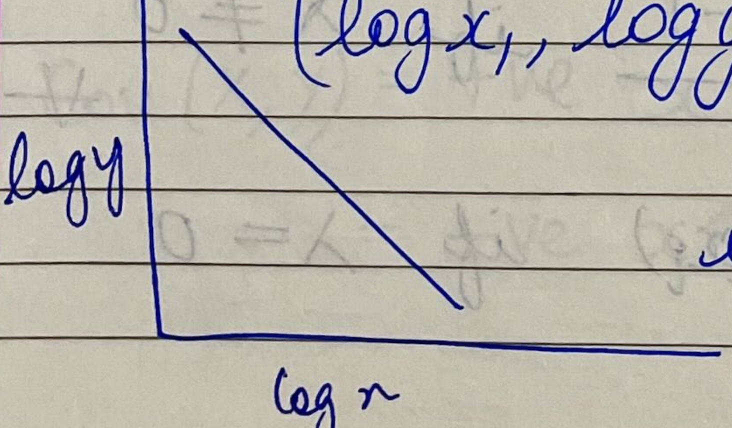
Log-log plot is one which helps figure out whether it follows power law or not.

$(x_1, y_1) \ldots (x_n, y_n)$



plot log of x and y.

$(\log x_1, \log y_1) \ldots (\log x_n, \log y_n)$



if it is a straight line then it follows power law axis of plot is ~~lines~~ log

ch17 Box-cox Transform

If I have to convert a Pareto distribution to Gaussian distribution then Box Cox Transformation is used.

Let, $X = x_1, x_2, \dots x_n$. (Pareto)

To convert Pareto → Gaussian.

$Y = y_1, y_2, y_3, \dots y_n$ (Gaussian)

Method:-

1) BoxCox(X). Enter X to boxcox function λ is generated.

2) $y_i \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \\ \lg(x_i) & \text{if } \lambda = 0 \end{cases} \quad \forall i : 1 \to n$

If $\lambda = 0$, $x \sim$ lognormal.

loggamma distribution is power law distribution.

Ch18 Applications of non Gaussian distributions

Construction of dams, Hydrology, etc.

Ch19 Co-variance

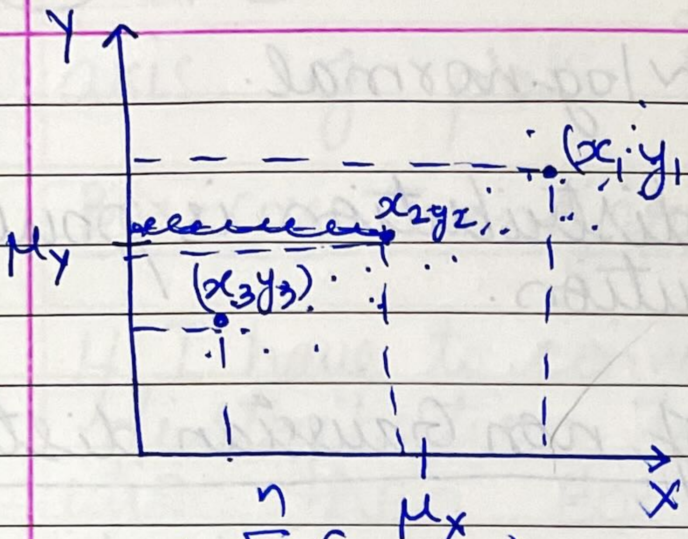
$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x) * (y_i - \mu_y)$$

$$Var(X) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_x) * (x_i - \mu_x)$$

$$Cov(X, X) = Var(X)$$

$Cov(X, Y) = +ve$ $x\uparrow, y\uparrow$

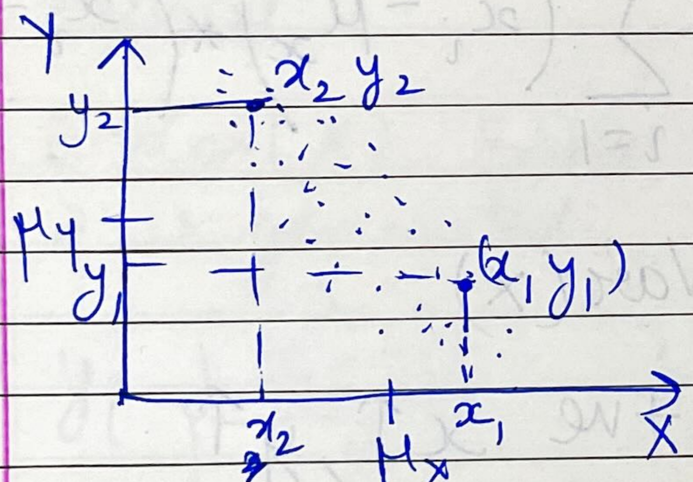
$= -ve$ $x\uparrow$ $y\downarrow$



$$cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu_x)*(y_i-\mu_y)$$

$(x_1\ y_1)$	+	+	+ve
$(x_2\ y_2)$	−	−	+ve
(x_3, y_3)	−	−	+ve

$cov(x,y) \rightarrow$ +ve.



(x_1, y_1)	+	−	−ve
(x_2, y_2)	−	+	−ve

$cov(x,y) \rightarrow$ −ve

Cov(X,Y) → is either too positive or too negative

If metric system changes for observations Covariance changes.

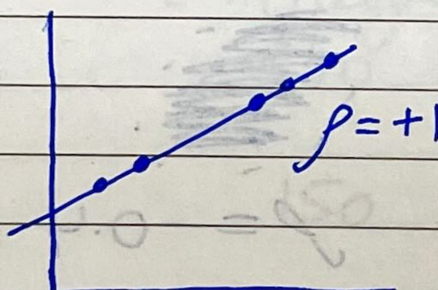
ch20 Pearson Correlation Coefficient (PCC)

Done

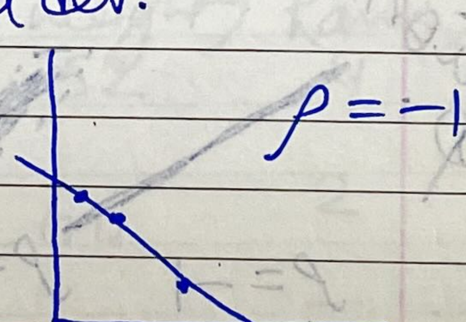
It is good when you have linear relationships

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \; ; \; -1 < \rho < +1$$

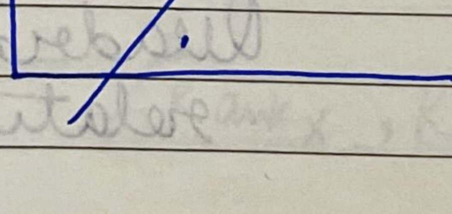
std dev.

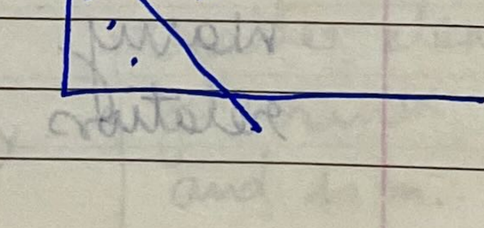


$x\uparrow y\uparrow$



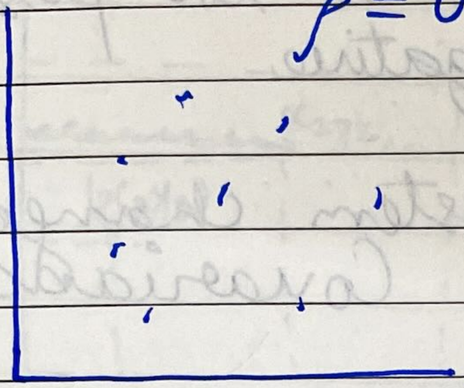
$x\uparrow Y\downarrow$





They say slope of a line doesn't matter.

$\rho = 0$

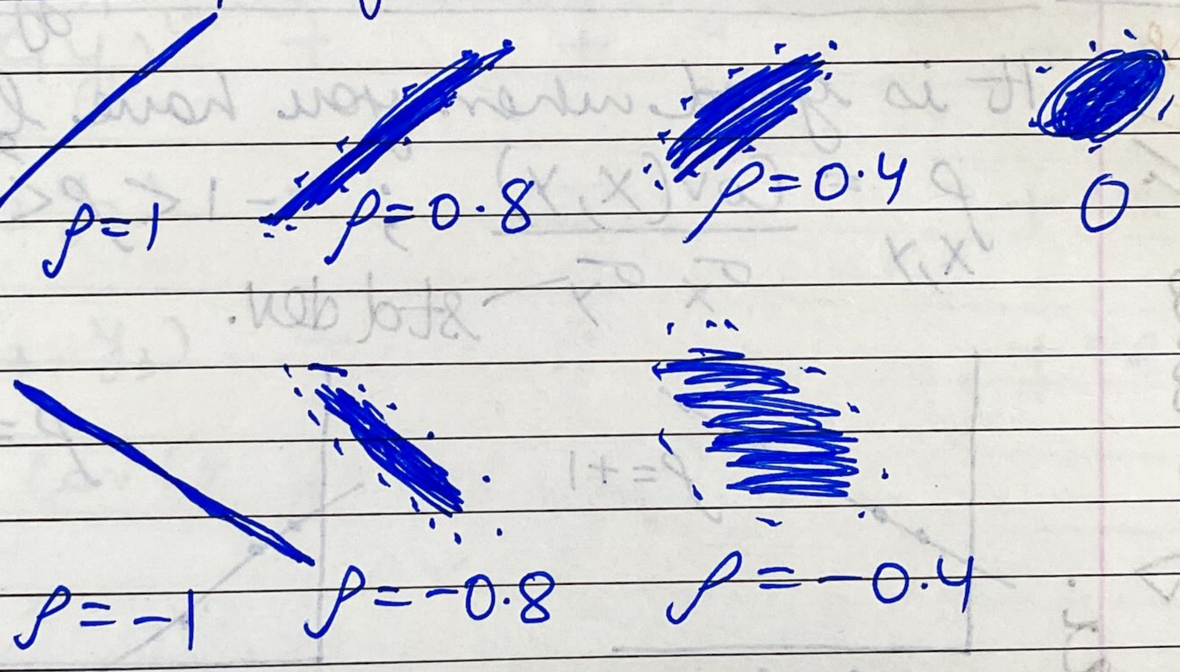


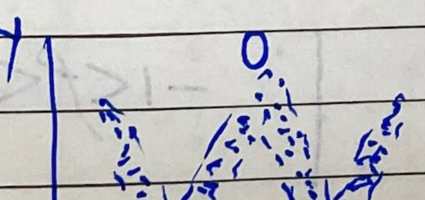
No relation

Slope of a line doesn't matter

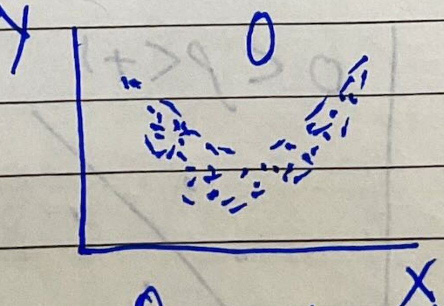
Example

It can be inc/dec relationship.





Wavy relation



Quadratic relation