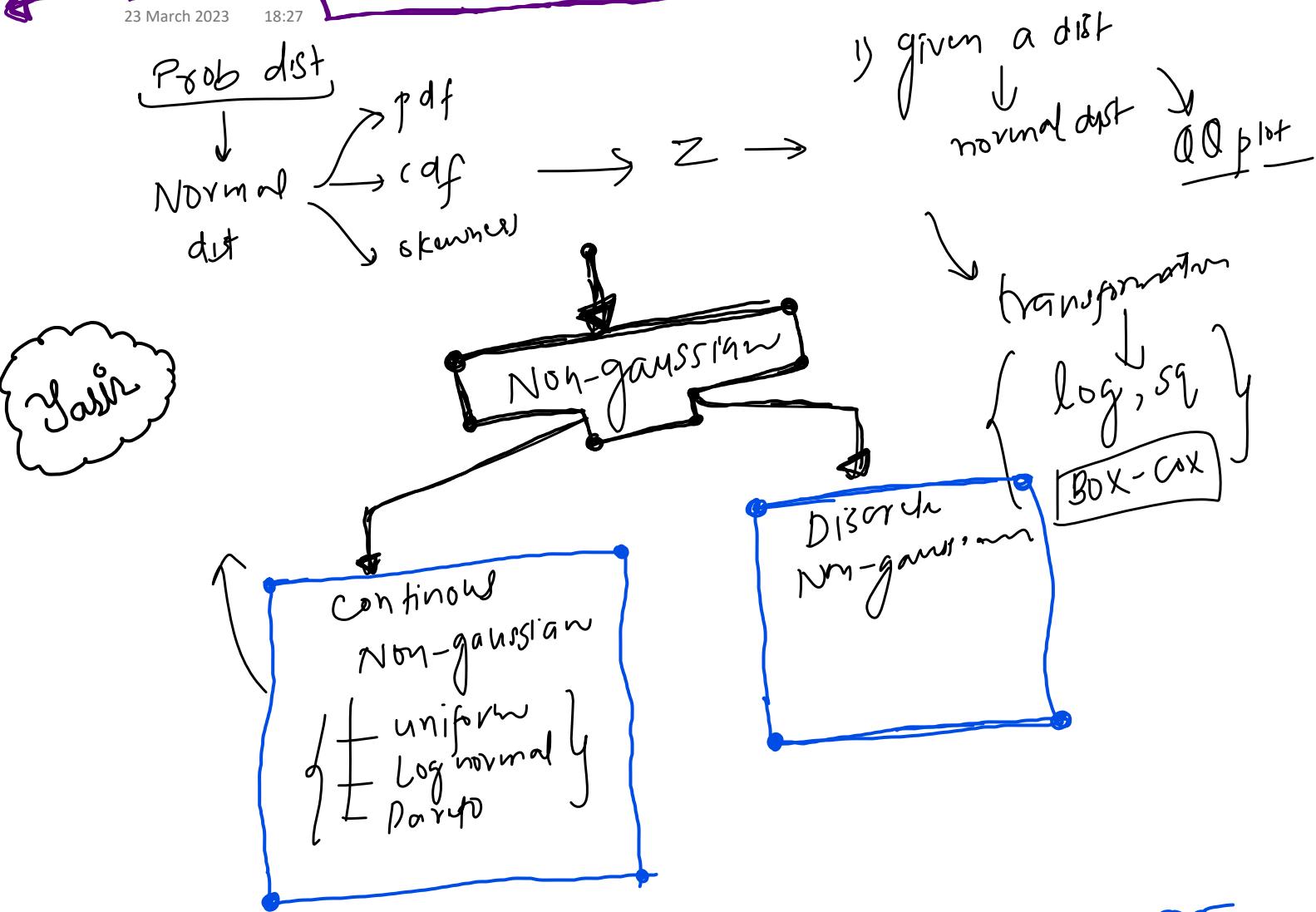


23 March 2023 18:27



Non-gaussian distribution continuous ଏବଂ ନେଟ୍ ବାବାର
 discrete ଏବଂ ମାତ୍ର ।

- Check, given a distribution is normal or not
 (with QQ-plot, quantile-quantile plot)
- Transformation, infere, non-gaussian distribution କେ
 gaussion distribution (୧ convert -logit ।
 (with, log, sq, box-cox transformation)

There are 4 moments in stat.

They are,

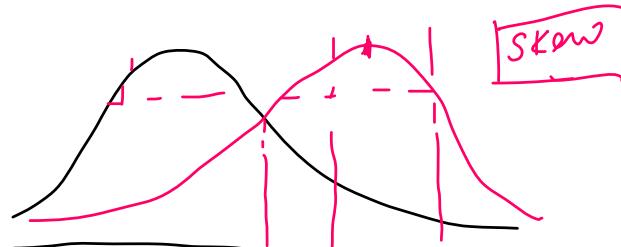
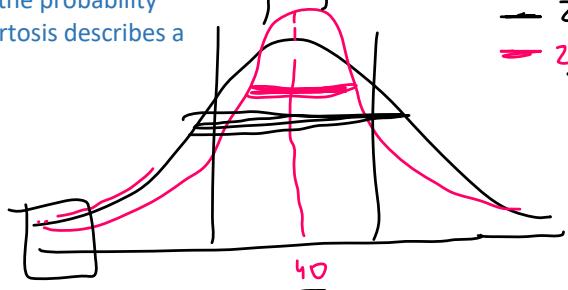
- 1st → mean
- 2nd → Std
- 3rd → skewness
- 4th → kurtosis

— 2017
— 2008

• What is Kurtosis?

Kurtosis is the 4th statistical moment. In probability theory and statistics, kurtosis (meaning "curved, arching") is a measure of the "tailedness" of the probability distribution of a real-valued random variable. Like skewness, kurtosis describes a particular aspect of a probability distribution.

Batsman → Sachin
2017 → 100 matches → 40 avg
2008 → 100 matches → 40 avg



Kurtosis परिष्ठी आमतौर पर
fatter को जा check करें। निचे
1 रो लाई tail fatter है मात्र
outliers को compare to 2 रो
graph

False notation about Kurtosis

<https://en.wikipedia.org/wiki/Kurtosis>

• Formula sample_kurtosis

$$\left\{ \frac{n * (n+1)}{(n-1) * (n-2) * (n-3)} * \sum_i^n \left(\frac{x_i - \bar{x}}{s} \right)^4 \right\} - \frac{3 * (n-1)^2}{(n-2) * (n-3)}$$

• Practical Use-case

(Finance के Kurtosis के बाबत)

In finance, kurtosis risk refers to the risk associated with the possibility of extreme outcomes or "fat tails" in the distribution of returns of a particular asset or portfolio.

If a distribution has high kurtosis, it means that there is a higher likelihood of extreme events occurring, either positive or negative, compared to a normal distribution.

In finance, kurtosis risk is important to consider because it indicates that there is a greater probability of large losses or gains occurring, which can have significant implications for investors. As a result, investors may want to adjust their investment strategies to account for kurtosis risk.

• Excess Kurtosis & Types

Excess kurtosis is a measure of how much more peaked or flat a distribution is

mutual fund के तौर
invest रखाएं मात्र
Kurtosis risk factor
देख लाएं इसे 1

If
inted
tailed

Excess Kurtosis & Types

SK-3

Excess kurtosis is a measure of how much more peaked or flat a distribution is compared to a normal distribution, which is considered to have a kurtosis of 0. It is calculated by subtracting 3 from the sample kurtosis coefficient.

Types of Kurtosis

$$sample \ k - 3 > 0$$

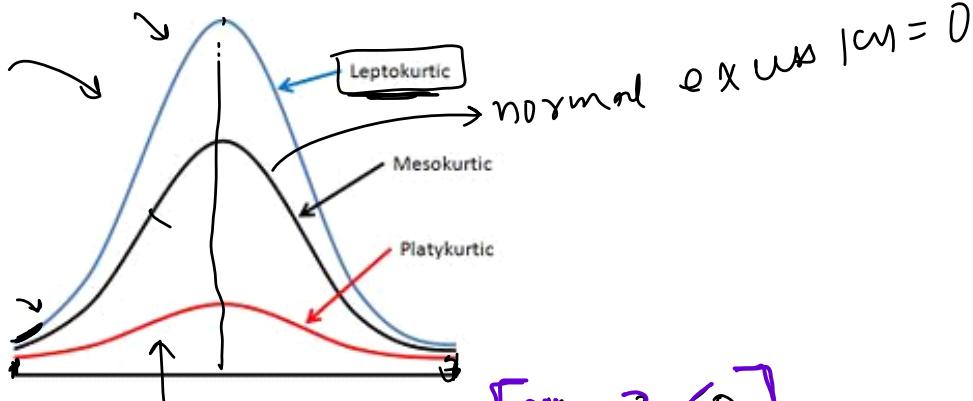
Leptokurtic

A distribution with positive excess kurtosis is called leptokurtic. "Lepto-" means "slender". In terms of shape, a leptokurtic distribution has fatter tails. This indicates that there are more extreme values or outliers in the distribution.

Example - Assets with positive excess kurtosis are riskier and more volatile than those with a normal distribution, and they may experience sudden price movements that can result in significant gains or losses.

SK-3 > 0

Finance example
not gd for investment



Platykurtic

Platykurtic

A distribution with negative excess kurtosis is called platykurtic. "Platy-" means "broad". In terms of shape, a platykurtic distribution has thinner tails. This indicates that there are fewer extreme values or outliers in the distribution.

Assets with negative excess kurtosis are less risky and less volatile than those with a normal distribution, and they may experience more gradual price movements that are less likely to result in large gains or losses.

Finance example
less risk for investment

Mesokurtic

SK-3 = 0

Distributions with zero excess kurtosis are called mesokurtic. The most prominent example of a mesokurtic distribution is the normal distribution family, regardless of the values of its parameters.

(μ, σ)

Mesokurtic is a term used to describe a distribution with a excess kurtosis of 0, indicating that it has the same degree of "peakedness" or "flatness" as a normal distribution.

Example -

In finance, a mesokurtic distribution is considered to be the ideal distribution for assets or portfolios, as it represents a balance between risk and return.

Finance example
stable, ideal
maybe no risk
for investment

For checking a distribution is normal or not



- How to find if a given distribution is normal or not?

Visual inspection One of the easiest ways to check for normality is to visually inspect a histogram or a density plot of the data. A normal distribution has a bell-shaped curve, which means that the majority of the data falls in the middle, and the tails taper off symmetrically. If the distribution looks approximately bell-shaped, it is likely to be normal.

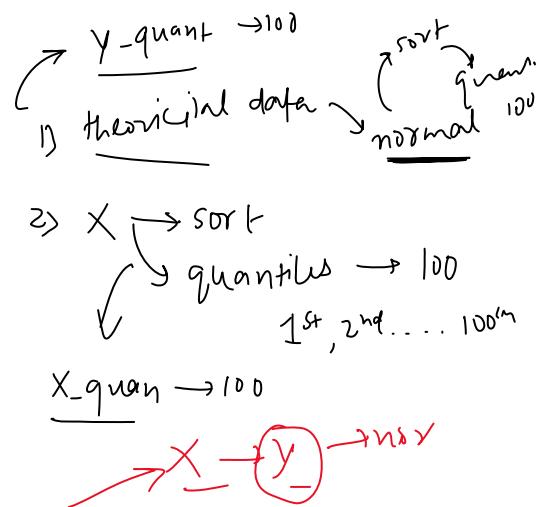
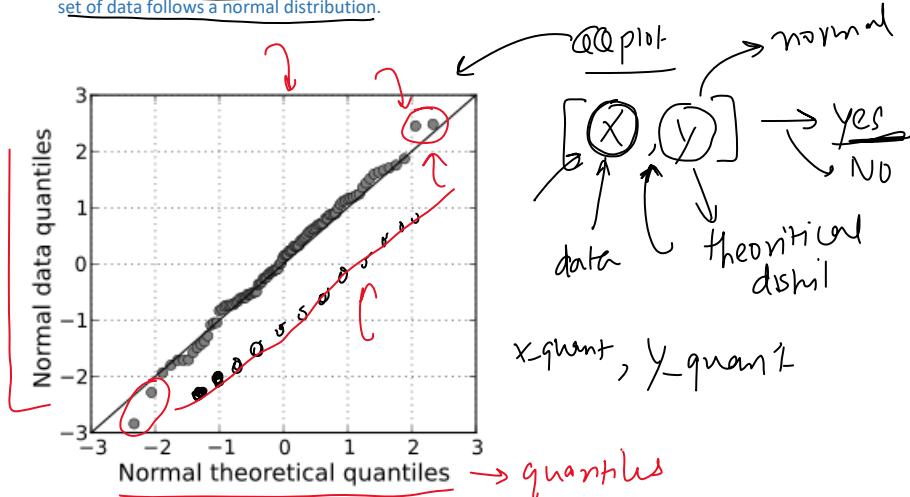


QQ Plot Another way to check for normality is to create a normal probability plot (also known as a Q-Q plot) of the data. A normal probability plot plots the observed data against the expected values of a normal distribution. If the data points fall along a straight line, the distribution is likely to be normal.

Statistical tests There are several statistical tests that can be used to test for normality, such as the Shapiro-Wilk test, the Anderson-Darling test, and the Kolmogorov-Smirnov test. These tests compare the observed data to the expected values of a normal distribution and provide a p-value that indicates whether the data is likely to be normal or not. A p-value less than the significance level (usually 0.05) suggests that the data is not normal.

- What is a QQ Plot and how is it plotted?

A QQ plot (quantile-quantile plot) is a graphical tool used to assess the similarity of the distribution of two sets of data. It is particularly useful for determining whether a set of data follows a normal distribution.

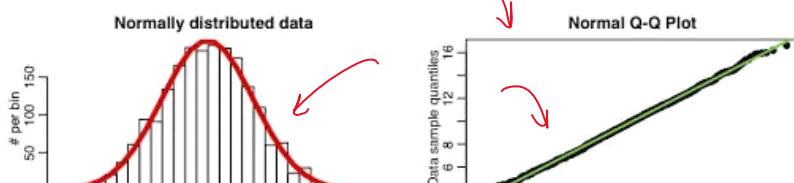


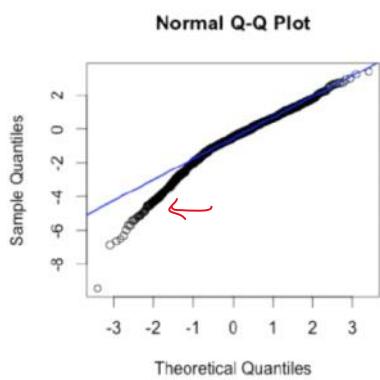
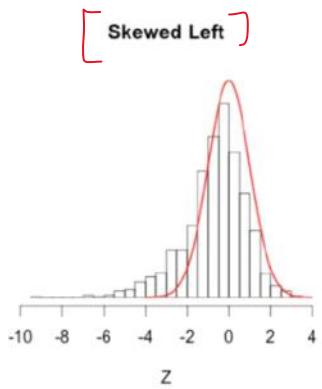
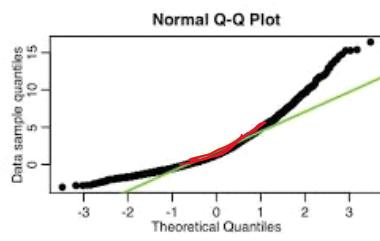
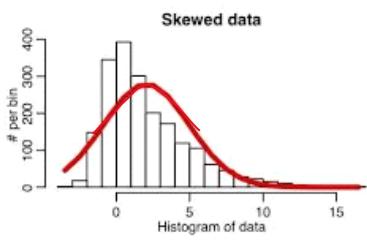
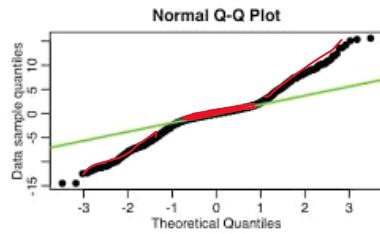
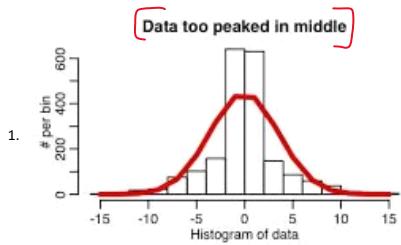
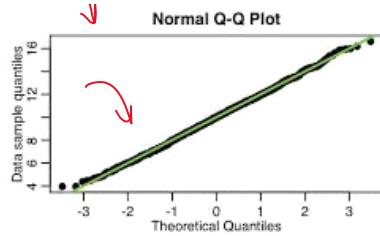
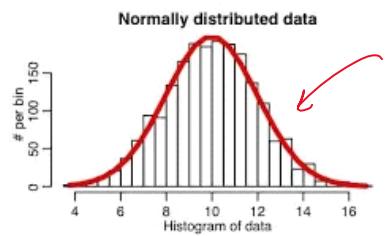
How to draw a QQ-Pbt

In a QQ plot, the quantiles of the two sets of data are plotted against each other. The quantiles of one set of data are plotted on the x-axis, while the quantiles of the other set of data are plotted on the y-axis. If the two sets of data have the same distribution, the points on the QQ plot will fall on a straight line. If the two sets of data do not have the same distribution, the points will deviate from the straight line.

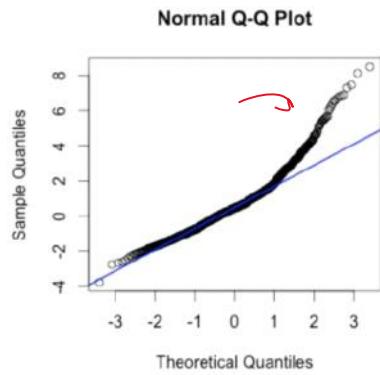
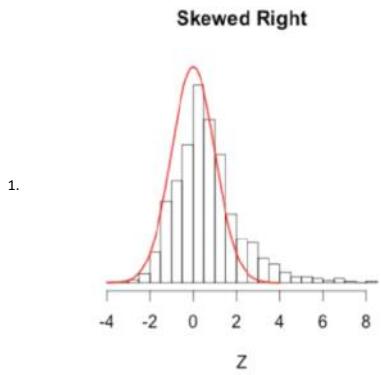
- Python example ✓
- <https://www.statsmodels.org/dev/generated/statsmodels.graphics.gofplots.qqplot.html>

- How to interpret QQ plots

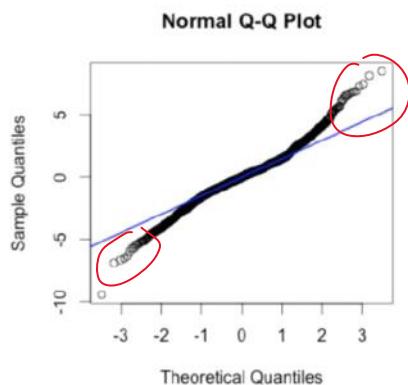
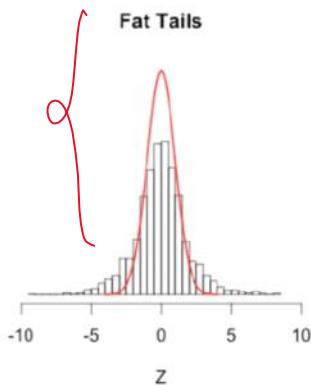




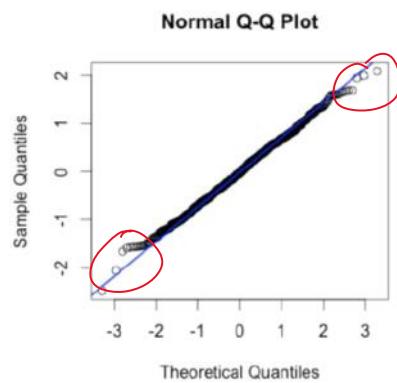
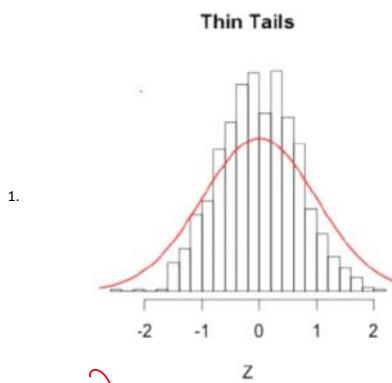
Left skewed, left side ↙ fate I



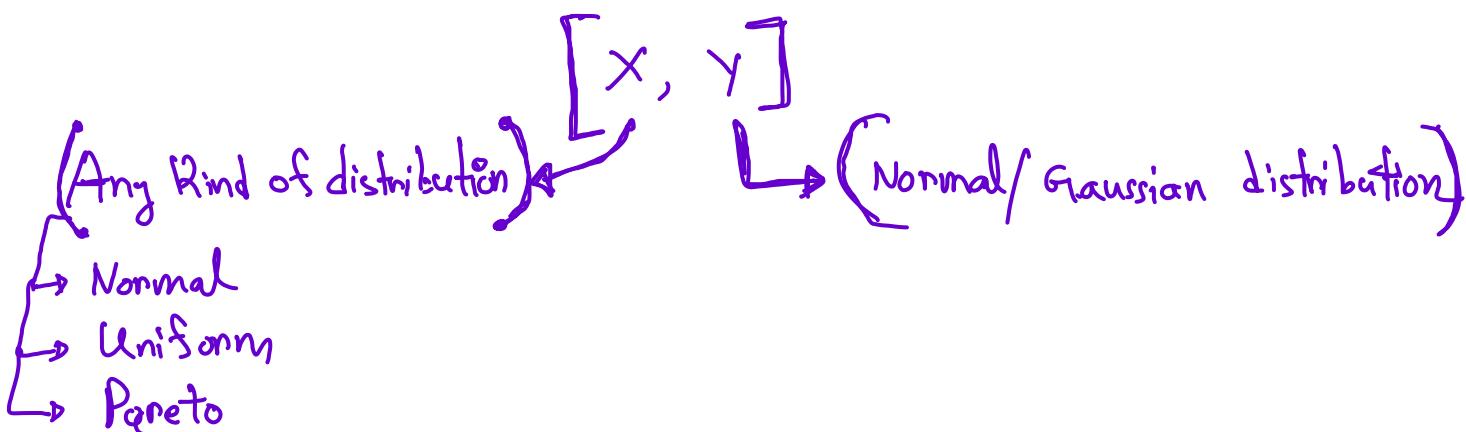
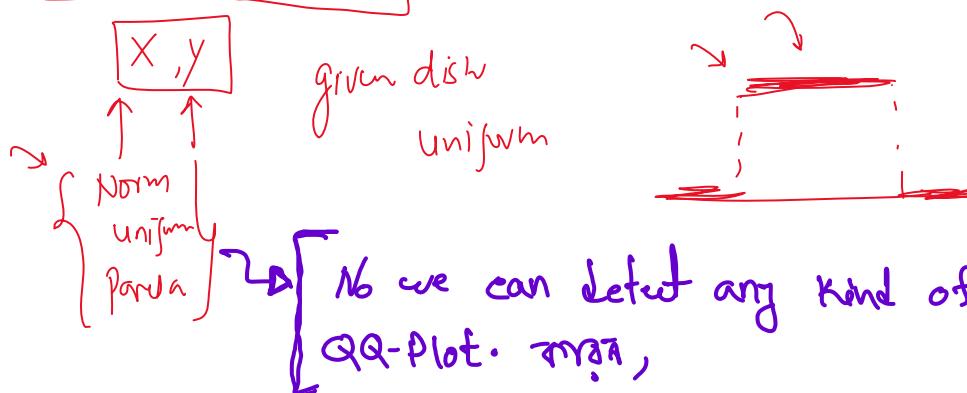
Right skewed,
right side ↗ fatig



combination of
(Left + Right) Skewed



- Does QQ plot only detect normal distribution?



Uniform Distribution

23 March 2023 13:19

P[Always output a value belonging to range]

• What is Uniform Distribution and its types

In probability theory and statistics, a uniform distribution is a probability distribution where all outcomes are equally likely within a given range. This means that if you were to select a random value from this range, any value would be as likely as any other value.

Types

Uniform
Discrete Continuous

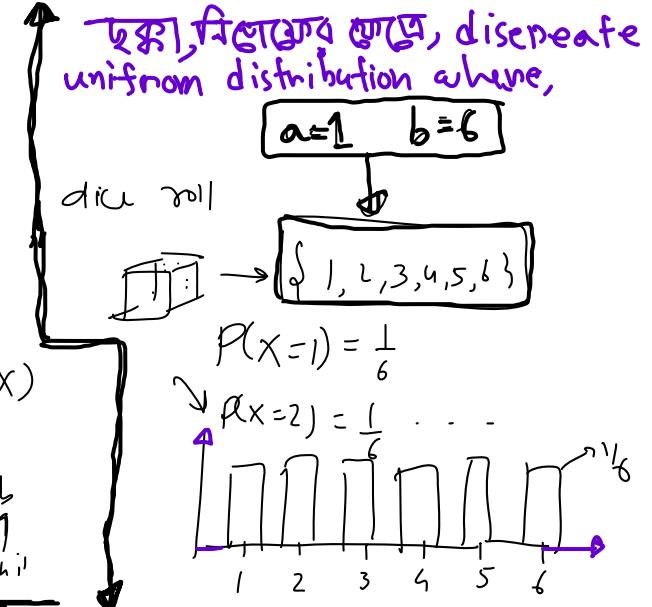
✓ Denoted as

$$X \sim U(a, b) \rightarrow \text{parameters} \begin{matrix} a \leftarrow 1 \\ b \leftarrow 6 \end{matrix}$$

lowest val highest val

• Examples

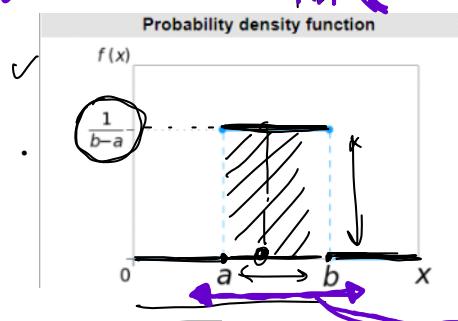
- a. The height of a person randomly selected from a group of individuals whose heights range from 5'6" to 6'0" would follow a continuous uniform distribution.
- b. The time it takes for a machine to produce a product, where the production time ranges from 5 to 10 minutes, would follow a continuous uniform distribution.
- c. The distance that a randomly selected car travels on a tank of gas, where the distance ranges from 300 to 400 miles, would follow a continuous uniform distribution.
- d. The weight of a randomly selected apple from a basket of apples that weighs between 100 and 200 grams, would follow a continuous uniform distribution.



basically इसके बारे में यह कहते हैं कि वर्षा का अवधि (a, b) है।

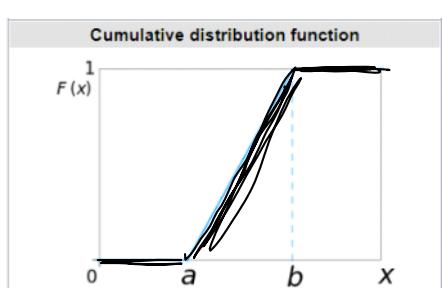
✓ PDF CDF and Graphs

PDF (Lecture #03)



$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

difference $(b-a)$, uniform वितरण, Area under the graph 1, $(1/(b-a))$



CDF (Lecture #04)

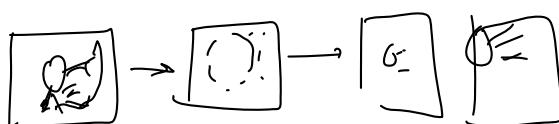
https://en.wikipedia.org/wiki/Continuous_uniform_distribution

• Skewness $\rightarrow 0$ \rightarrow symmetric \rightarrow Normal

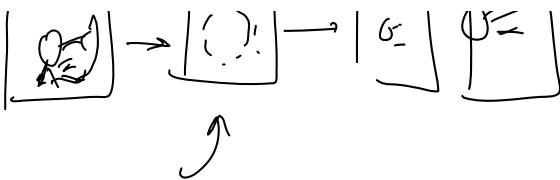
↳ (Uniform distribution का तरीका skewness = 0)

✓ Application in Machine learning and Data Science

✓ **Random initialization** In many machine learning algorithms, such as neural networks and k-means clustering, the initial values of the parameters can have a significant impact on the final result. Uniform distribution is often used to randomly initialize the parameters, as it ensures that all values in the range have an equal probability of being selected.



In many machine learning algorithms, even neural networks and k-means clustering, the initial values of the parameters can have a significant impact on the final result. Uniform distribution is often used to randomly initialize the parameters, as it ensures that all values in the range have an equal probability of being selected.



- b. **Sampling:** Uniform distribution can also be used for sampling. For example, if you have a dataset with an equal number of samples from each class, you can use uniform distribution to randomly select a subset of the data that is representative of all the classes.
- c. **Data augmentation:** In some cases, you may want to artificially increase the size of your dataset by generating new examples that are similar to the original data. Uniform distribution can be used to generate new data points that are within a specified range of the original data.
- d. **Hyperparameter tuning:** Uniform distribution can also be used in hyperparameter tuning, where you need to search for the best combination of hyperparameters for a machine learning model. By defining a uniform prior distribution for each hyperparameter, you can sample from the distribution to explore the hyperparameter space.

→ Deep learning
in NN

NN: একটি ক্লাউডে, Random initialization এবং সমৃদ্ধ Uniform এবং ঘোর
Gaussian Normal initialization, ($\mu=0, \sigma=0.2$) ব্যবহার করা হয়। Gradient
Converge একটি রূপ।

যাতে deep learning,
Xavier/Glorot initialization → Uniform distribution
He initialization → Gaussian Normal initialization
ব্যবহার করা।

Symmetric activation fn: tanh, sign

non-Symmetric activation fn: ReLU, Leaky ReLU, PReLU, ELU,
sigmoid, softplus

Non-Symmetric activation fn এবং সমৃদ্ধ He initialization
ব্যবহার করা।

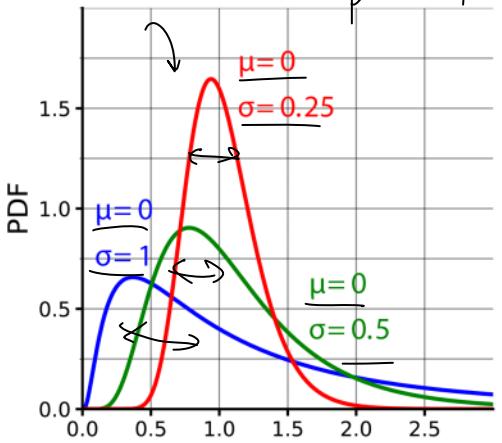
Log Normal Distribution

23 March 2023 13:19

(Right Skewed)

In probability theory and statistics, a lognormal distribution is a heavy tailed continuous probability distribution of a random variable whose logarithm is normally distributed.

parameters $\rightarrow (\mu, \sigma)$



प्रायत्निक R8 SKewed distribution for Log Normal अग्रणी
Age column (R8 ii) का लोगरिदम वल्यू हो (ln(x)) = new value
अग्रणी, आवृत्ति वर्गीकरण अग्रणी नोर्मल अग्रणी अग्रणी लोग
नोर्मल वर्गीकरण अग्रणी अग्रणी

$$\log(x) \sim N(\mu, \sigma)$$

$$X \sim \text{log Normal}$$

$$\log(x) \sim N(\mu, \sigma)$$

Age

$$\ln(27), \ln(28), \ln(29), \dots$$



Examples

- The length of comments posted in Internet discussion forums follows a log-normal distribution.
- Users' dwell time on online articles (jokes, news etc.) follows a log-normal distribution.
- The length of chess games tends to follow a log-normal distribution.
- In economics, there is evidence that the income of 97%–99% of the population is distributed log-normally.

insta, reddit, failnly, words

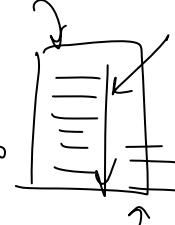
words

video

$$C_1 \rightarrow 23$$

$$C_2 \rightarrow 41$$

$$C_3 \rightarrow 180$$



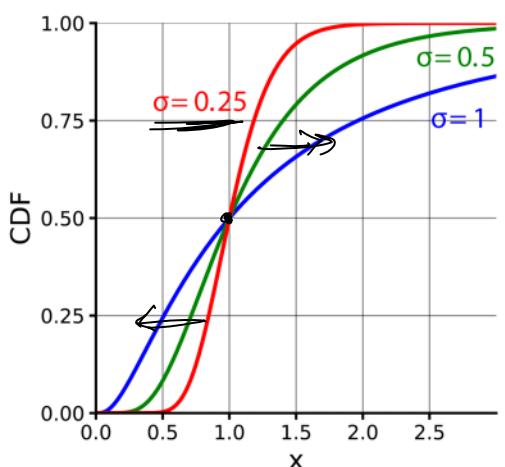
Denoted as

$$X \sim \text{lognormal}(\mu, \sigma) \rightarrow \ln(x) \sim N(\mu, \sigma^2), \text{ denoted by, lognormal } (\mu, \sigma)$$

✓ PDF Equation

$$\frac{1}{x\sqrt{2\pi}} e^{-\left(\frac{\ln x - \mu}{\sigma^2}\right)^2} \quad \begin{bmatrix} \text{similar} \\ \text{Normal dist.} \end{bmatrix}$$

CDF



Skewness

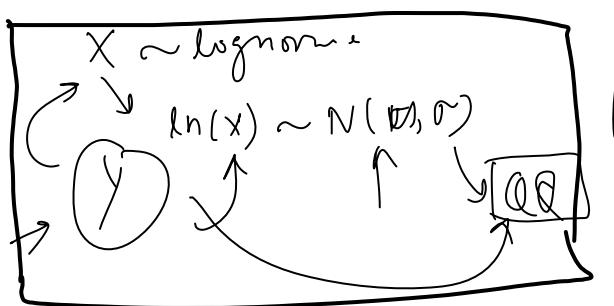
Skewed

How to check if a random variable is log normally distributed?

skewness
skewness



Using QQ plot but we need to follow the definition. Should be plot $\ln(x)$ distribution percentile in QQ-plot.



Pareto Distribution

23 March 2023 13:19

Pareto Distribution

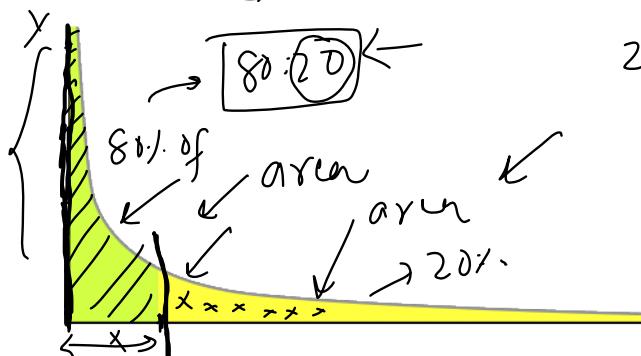
The Pareto distribution is a type of probability distribution that is commonly used to model the distribution of wealth, income, and other quantities that exhibit a similar power-law behaviour

What is Power Law

In mathematics, a power law is a functional relationship between two variables, where one variable is proportional to a power of the other. Specifically, if y and x are two variables related by a power law, then the relationship can be written as:

$$y = k * x^\alpha$$

$$y = K x^\alpha$$



$\frac{20\% \text{ of } x}{80\% \text{ of entire}}$

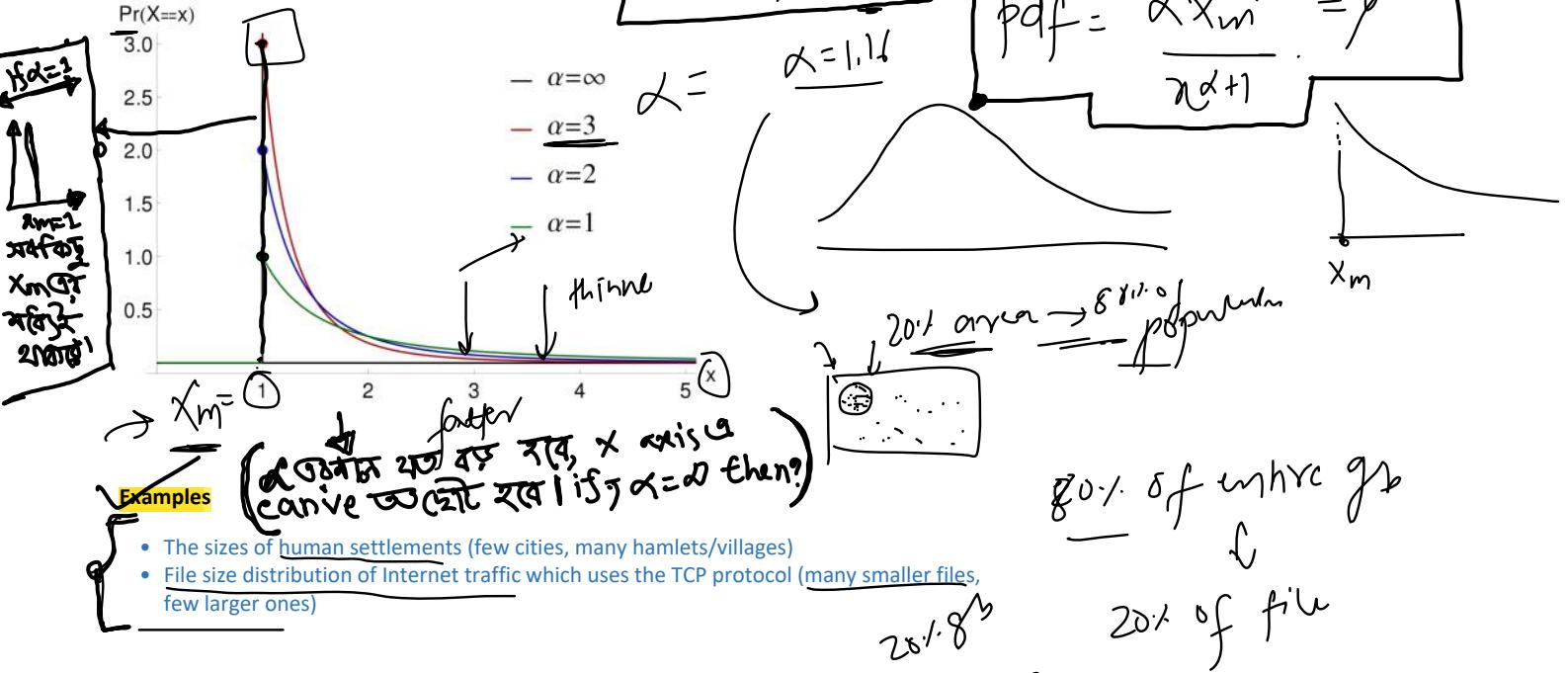
Power-law, 80:20 rule, ८०:२०, ८०% लोक्याते पुरिबीय ८०%. अमेरिका भारत जैसे।

Graph, यदि $x=20$ (20%) तो $y=80%$. अमेरिका दुर्घटना का लिए।

Vilfredo Pareto originally used this distribution to describe the allocation of wealth among individuals since it seemed to show rather well the way that a larger portion of the wealth of any society is owned by a smaller percentage of the people in that society. He also used it to describe distribution of income. This idea is sometimes expressed more simply as the Pareto principle or the "80-20 rule" which says that 20% of the population controls 80% of the wealth

Pareto(wealth, income)

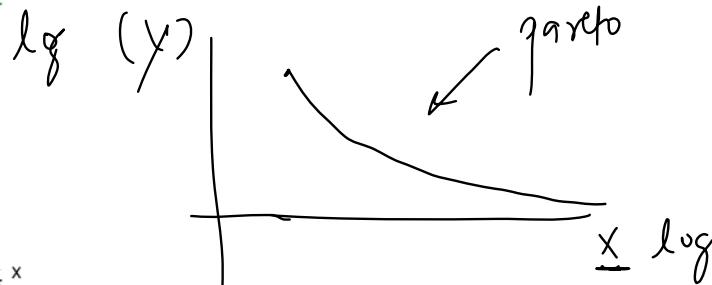
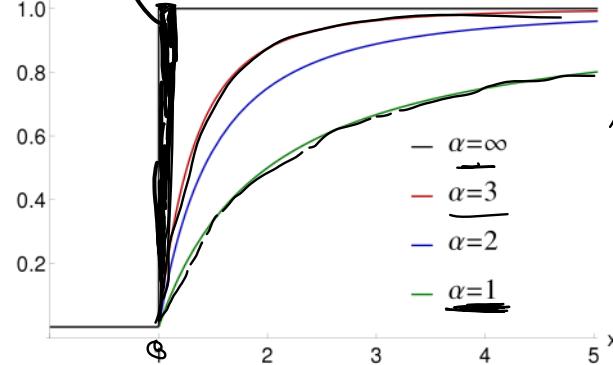
Graph & Parameters



CDF

$\Pr(X < x)$

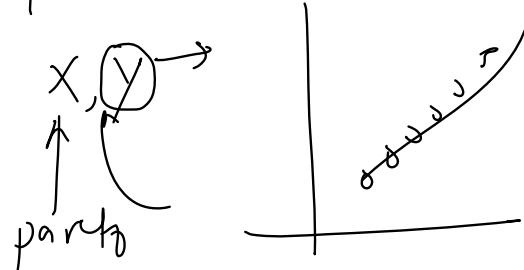
$$\text{CDF} \rightarrow \Pr(X \leq x) \rightarrow \text{PDF } \Pr(x = \alpha) \rightarrow \text{CDF } \Pr(x \leq \alpha)$$



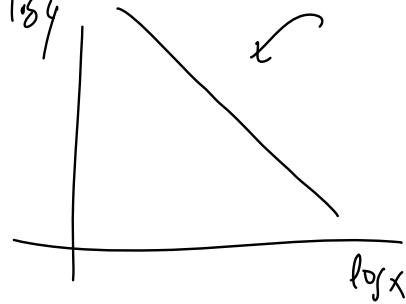
~~x~~ Skewness → [Skewed R8]

~~x~~ How to detect if a distribution is Pareto Distribution?

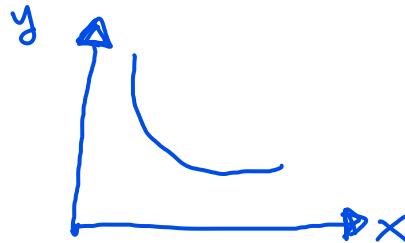
[Q-Q] Pdt



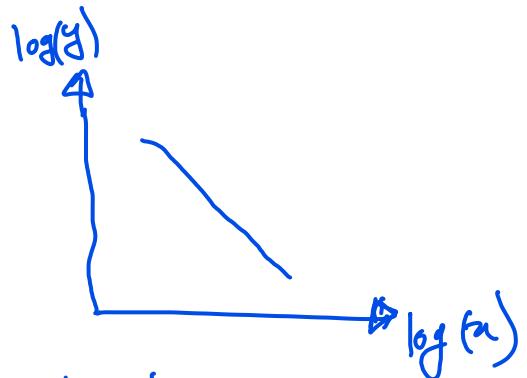
$$Y = \frac{\alpha x_m}{x^{\alpha+1}}$$



* [अन्तिम distribution pareto को देखते कीलके check करते]



if



⊗ (then, we can call it pareto distribution) ⊗

Transformations

23 March 2023 13:24

