Chebyshev's Inequality: If a distribution follows ~~B~~ the Gaussian / Normal distribution, then by the emperical rule, we can say – Probability of a point x, lying in some region –

$$P_r(\mu - \sigma < x < \mu + \sigma) \approx 68\%$$

$$P_r(\mu - 2\sigma < x < \mu + 2\sigma) \approx 95\%$$

$$P_r(\mu - 3\sigma < x < \mu + 3\sigma) \approx 99.7\%$$

But if the distribution does not follow any Gaussian / Normal distribution, then we can find out what percentage of our population lies within what range of standard deviation using the Chebyshev's inequality. It's relationship ~~can~~ be given as —

$$\boxed{P_r(\mu - K\sigma < x < \mu + K\sigma) > 1 - \frac{1}{K^2}}$$

Let's say, $K = 2$, then

$$P_r(\mu - 2\sigma < x < \mu + 2\sigma) > 1 - \frac{1}{4}$$

$$P_r(\mu - 2\sigma < x < \mu + 2\sigma) > \frac{3}{4}$$

i.e., 75% or more population lies within 2 standard deviation from the mean.

Q-Q Plot (Quantile-Quantile Plot):

If a set of observations is approximately normally distributed a normal quantile-quantile plot of the observations will result in an approximately straight line.

It can be used to check whether a sample coming from a distribution is normal or not.

Suppose, we have a data sample. Check if it is normally distributed or not.

3.89, 4.75, 6.33, 4.75, 7.21, 5.78, 5.80, 5.20, 6.64

Step 1: Rearrange the data in ascending order.

3.89, 4.75, 4.75, 5.20, 5.78, 5.80, 6.33, 7.21, 7.90

$\mu = \frac{51.61}{9} = 5.73$, $\sigma^2 = \frac{12.854}{9} = 1.4282$ | $\sigma = 1.195$

$Z_{(3.89)} = \frac{3.89-5.73}{1.195} = -1.53$

$Z_{(4.75)} = \frac{4.75-5.73}{1.195} = -0.82$

$Z_{(5.20)} = -0.44$

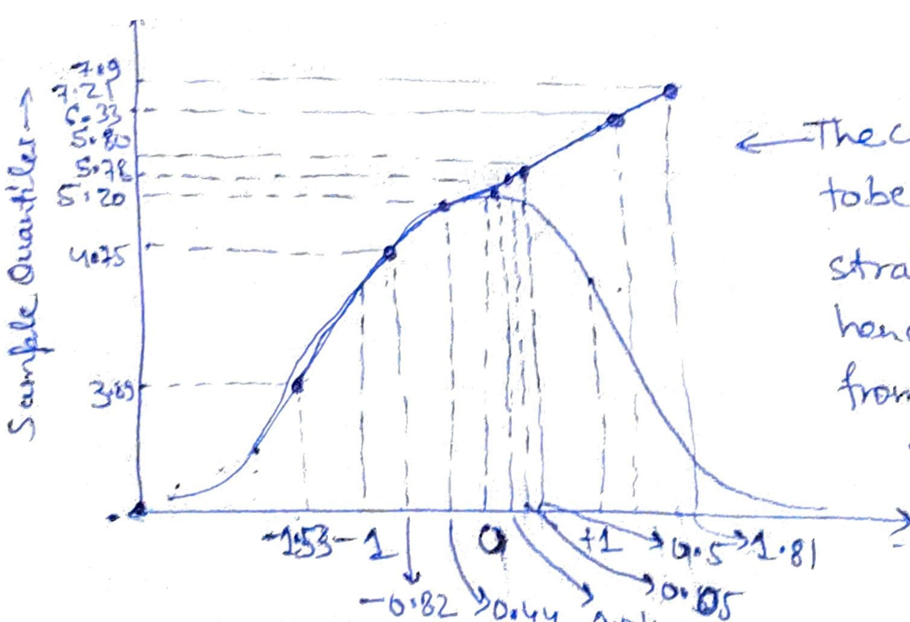
$Z_{(5.78)} = 0.04$

$Z_{(5.80)} = 0.05$

$Z_{(6.33)} = 0.50$

$Z_{(7.21)} = 1.23$

$Z_{(7.90)} = 1.81$



← The curve seems to be forming a straight line hence, it comes from normal distribution.

Log Normal Distribution : A distribution is said to be following a log-normal distribution when, the log of sample point in the distribution follows a normal/ gaussian distribution.

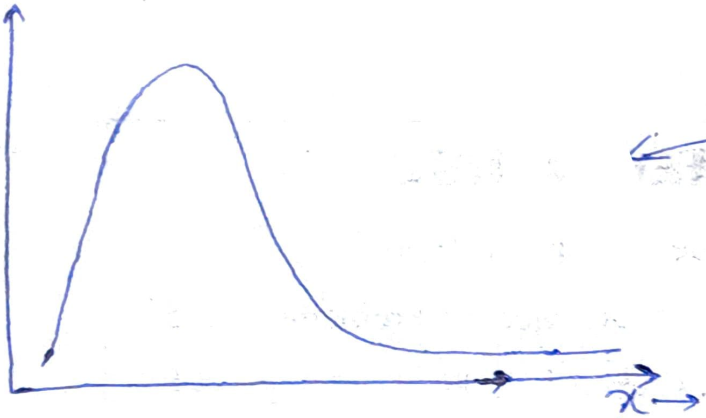
$S = \{x_1, x_2, x_3, --- x_n\}$

$$S \rightarrow \underbrace{\log(x_1), \log(x_2), \log(x_3) --- \log(x_n)}$$

$\Downarrow$

follows Gaussian Distribution where

$GD(\mu, \sigma)$



Graph of log-normal distribution.

It is right-skewed normal distribution

We can convert this log-normal distribution ~~into~~ follows - by calculating log of each sample point and then it will start following the normal distribution.