# Other Terms

Z Value :

The Z value or Z score is the measure of, how much standard deviations far a datapoint is from its mean. The Z score is calculated for each datapoint. Consider the normal distribution, here the Z score at point on

$\mu - 3\sigma$ will be -3, similarly the Z score at point

$\mu + \sigma$ will be 1. Basically, Z score of a points tells us that, "how much standard deviations a point if far from its mean". Here its mean means the mean of distribution. Here is the formula,

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

Where, the x is data point, μ is mean and σ is standard deviation.
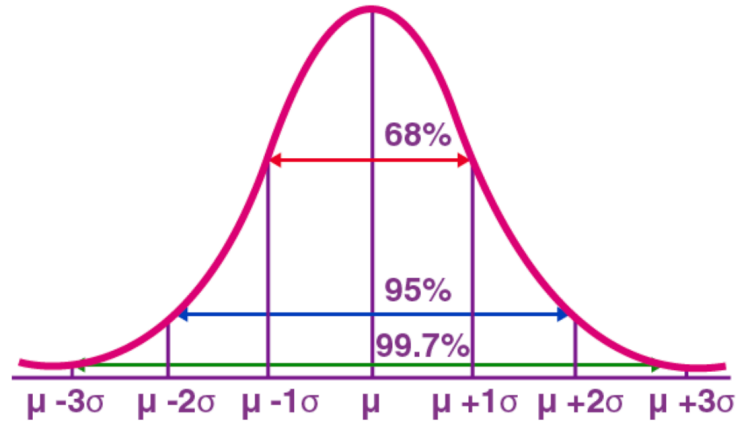
KL Divergence :

KL divergence is an asymmetric metric used to get difference between two probability distributions. By asymmetric I mean that, KLD of distribution P referencing Q will be different from KLD of distribution Q referencing to P. Here is the formula,

$$D_{\mathsf{KL}}(P\|Q) = \sum_{x} P(x)\log\left(\frac{P(x)}{Q(x)}\right)$$

Central Limit Theorem :

The central limit theorem states that, from a population the sample variable of same sized samples follows normal distribution as the sample size increases regardless of the distribution of population.

Lets break down the definition. The population means whole dataset, the sample means subset of dataset, sample variable means the sum or average of sample, sample size means number of datapoints in a sample. Now we revise the definition, it says that from a population, if we take n samples (subset from population) then the mean or sum of these samples follows the normal distribution as the sample size becomes larger irrespective of the distribution of population. Furthermore mean and variance of these sample variables will be equal to mean and variance of population. Here are some characteristics of CLT :

1. Sampling must be random - each datapoint must have equal probability of being selected

By Karan Hadiyal

2. Samples should be independent - results from previous or future sample should not affect the current sample
3. Sample size should be limited to 10% of the population
4. Each sample must have same size
5. As number of samples increase the CLT is more applicable

Generally sample size of 30-50 is preferred in CLT.

Maximum Likelihood Estimation :

MLE is about finding parameters of distribution that best fits the given data points. Let's say we are given 10 datapoints, we intuitively decided that it follows the Gaussian distribution. Note that the first step is to decide that which distribution is perfect for the given datapoints. Now we have found that the normal distribution fits best, now our job is to find parameters of this distribution. In case of normal distribution these parameters are mean (μ) and standard deviation (σ). These parameters are called maximum likelihood estimates and the method we use is called maximum likelihood estimation. Sometimes, MLE is difficult to compute so we use expected MLE.

Here I have not defined the maths behind MLE, you can refer to Jonny Brooks-Bartlett's article on MLE on towards datascience.

By Karan Hadiyal