Maximum Likelihood Estimation (MLE)

The Maximum Likelihood Estimation (MLE) provides a way to determines the natural parameters (The parameters on which the algorithm's performance governs) optimal values and other.

The provides the following things like:

1. It provides the point estimate i.e., if the algorithm is having more than 1 natural parameters it can provide best natural parameter for the respective algorithm.
2. It can provide the best estimator of the respective algorithm i.e., Mean/ Mode/ Meadian.

It's more like works as the Probability Density Function (PDF) but it can give the probability of the whole sample (where, sample size>1).

It is also called as **Deterministic Approximate Inference Method.**

What is Likelihhood Function ?

Likelihood Function is also a Probability Density Function (PDF) which tells the probability of occurance of particular values in a sample.

Likelihood Function denoted as the **L()** .

Let's find out the Minimum Variance and Unbaised Estimator (MVU) using the Maximum Likelihood Estimate (MLE) Method for the Normal/Gaussian Distribution.

$$L() = P(X_1 = x_1 \cap X_2 = x_2 \cap X_3 = x_3 \cap X_4 = x_4 \cap X_5 = x_5 \cap X_6 = x_6)$$

In the generalized way,

$$L() = P(X_1 = x_1 \cap X_2 = x_2 \cap X_3 = x_3 \cap \ldots X_n = x_n)$$

$$L() = P(X = x_1).\,P(X = x_2).\ldots.P(X = x_n)$$

So,

$$L() = \frac{1}{\sqrt{2\pi}(\sigma^\wedge)}\mathrm{e}^{\frac{-(x_1-(\mu^\wedge))^2}{2(\sigma^\wedge)^2}}\ \frac{1}{\sqrt{2\pi}(\sigma^\wedge)}\mathrm{e}^{\frac{-(x_2-(\mu^\wedge))^2}{2(\sigma^\wedge)^2}}\ \ldots\ \frac{1}{\sqrt{2\pi}(\sigma^\wedge)}\mathrm{e}^{\frac{-(x_n-(\mu^\wedge))^2}{2(\sigma^\wedge)^2}}$$

σ, μ= are unknown and we will estimate then so we will use cap $(\wedge)$ on above natural parameters so, $\sigma^\wedge, \mu^\wedge$

It is independent samples so we have multiplied them and it is distributed identically so, $\sigma^\wedge, \mu^\wedge$ are same.

$$L(\sigma^\wedge, \mu^\wedge) = \frac{1}{\sqrt{2\pi}(\sigma^\wedge)}\mathrm{e}^{\frac{-\sum_{i=1}^{N}(x_i-(\mu^\wedge))^2}{2(\sigma^\wedge)^2}}$$

It is a Likelihood Function for the two variables ($\sigma^\wedge, \mu^\wedge$).
Now, we want to find such values of σ and μ such that $L(\sigma^\wedge, \mu^\wedge)$ reaches its maximum values.

We need to perform following things before applying it:

1. For the smoothing of the function, we will use Log function.
2. For the reversal of the non-convex function graph, we will use the negative (-) sign. It will provide the global maximum.

After all the alteration in the whole function the final equation of the function is:

$$-log\ L(\sigma^\wedge, \mu^\wedge)] = -log\frac{1}{\sqrt{2\pi}(\sigma^\wedge)}\mathrm{e}^{\frac{-\sum_{i=1}^{N}(x_i-(\mu^\wedge))^2}{2(\sigma^\wedge)^2}}$$

We will simplify the equative to find the natural parameters functions.

$$= -log_e \frac{1}{\sqrt{2\pi}(\sigma^\wedge)}^N + log_e \mathrm{e}^{\frac{-\sum_{i=1}^N (x_i - (\mu^\wedge))^2}{2(\sigma^\wedge)^2}}$$

$$= -(-N \log_e(\sqrt{2\pi}(\sigma^\wedge)) - \frac{-\sum_{i=1}^N (x_i - (\mu^\wedge))^2}{2(\sigma^\wedge)^2})$$

$$-log_e(L(\sigma^\wedge, \mu^\wedge)) = (N \log_e(\sqrt{2\pi}(\sigma^\wedge))) + \frac{-\sum_{i=1}^N (x_i - (\mu^\wedge))^2}{2(\sigma^\wedge)^2}$$

we will use the partial derivation to solve the equation further on.

$$= \frac{\delta}{\delta\mu^\wedge}(N \log_e(\sqrt{2\pi}(\sigma^\wedge))) + \frac{-\sum_{i=1}^N (x_i - (\mu^\wedge))^2}{2(\sigma^\wedge)^2}$$

$$\mu_{ML}^\wedge = \frac{\sum_{i=1}^N (x_i)}{N}$$

$$\sigma_{ML}^{\wedge^2} = \frac{\sum_{i=1}^N (x_i - \mu^\wedge)^2}{N}$$