Likelihood Function :
=================

Likelihood Function

Likelihood is the probability of the observed data given the parameters that generate the underlying data. Let's suppose we observe n observations $x_1, x_2, \ldots x_n$ and assume that the observations are independent and identically normally distributed with mean μ and variance σ^2.

The likelihood function in this case would be as follows:

$$P(Data/Model\ parameters) = P(x_1, x_2, \ldots x_n / \mu, \sigma^2)$$

Since the observations are independent, we can factorize the likelihood as follows:

$$P(Data/Model\ parameters) = \prod_{i=1}^{n} P(x_i / \mu, \sigma^2)$$

Each of the $x_i \sim Normal(\mu, \sigma^2)$, hence the likelihood can be further expanded as follows:

$$P(Data/Modelparameters) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x_i - \mu)^2}{2\sigma^2}}$$

Maximum Likelihood Estimate

Maximum likelihood estimate (MLE) is a technique for estimating the parameters of a distribution or model. This is achieved by deriving the parameters that would maximize the likelihood function—i.e., maximize the probability of observing the data given the parameters of the model. Let's work through an example to understand maximum likelihood estimates.

Suppose Adam tosses a coin 10 times and observes 7 heads and 3 3tails. Also, assume that the tosses are independent and identical. What would be the maximum likelihood estimate for the probability of heads for the given coin?

Each toss of a coin is a Bernoulli trial, with the probability of heads being, let's say, p, which is an unknown parameter that we want to estimate. Also, let the event that a toss produces heads be denoted by 1 1and tails by 0.

The likelihood function can be represented as follows:

$$P(Data/parameter) = L(p) = P(x_1, x_2, \ldots\ldots x_{10}/p)$$

$$= \prod_{i=1}^{10} P(x_i/p)$$

$$= p^7 (1-p)^3$$

Just for clarification, let's see how the likelihood L came to be $p^7(1 - p)^3$.

For each heads, the probability from the Bernoulli distribution is $P(x_i = 1/p) = p^1(1-p)^0 = p$. Similarly, for each tails the probability is $P(x_i = 0/p) = p^0(1-p)^1 = 1-p$. As we have 7 heads and 3 tails, we get the likelihood $L(p)$ to be $p^7(1 - p)^3$.

To maximize the likelihood L, we need to take the derivate of L with respect to p and set it to 0.

Now, instead of maximizing the likelihood $L(p)$ we can maximize the logarithm of the likelihood—i.e., $logL(p)$. Since logarithmic is a monotonically increasing function, the parameter value that maximizes $L(p)$ would also maximize $logL(p)$. Taking the derivative of the log of the likelihood is mathematically more convenient than taking the derivative of the product form of the original likelihood.

$$logL(p) = 7logp + 3log(1-p)$$

Taking the derivative of both sides and setting it to zero looks as follows:

$$\frac{dLog(L(p))}{dp} = \frac{7}{p} - \frac{3}{1-p} = 0$$

$$\Rightarrow p = 7/10$$

Interested readers can compute the second derivative $\frac{d^2 \log(L)}{dp^2}$ at $p = \frac{7}{10}$; you will for sure get a negative value, confirming that $p = \frac{7}{10}$ is indeed the point of maxima.