# 06_Maximum_Likelihood_Estimation

**Bayesian:**

$$P(D_x|\theta_x) = \frac{P(\theta_x|D_x)P(D_x)}{P(\theta_x)}$$

$$Likelihood = \frac{Posterior \times \ Evidence}{Prior}$$

$D = \{x_1, x_2 \ldots, x_n\}$
$D_x$ : $xth$ sampling data set in training set D,
Samples in each data set ($x \in D_x$) are $Independent \ \ and \ Identically \ Distributed$
$\theta_x$: model parameter to $D_x$

- **Prior**
  $P(\theta_x)$ represents the knowledge on hypothesis before any observation.
- **Likelihood**
  $P(D_x|\theta_x)$ represents the probability of observing data given the hypothesis
- **Posterior**
  $P(\theta_x|D_x) = \frac{P(D_x|\theta_x)P(\theta_x)}{P(d_x)}$ represents the probability of hypothesis after the data has been observed
- **Evidence** $P(D_x)$
  encodes the quality of the underlying model

$$P(D_x) = \begin{cases} \sum_\theta P(\ D_x|\ \theta_x)P(\theta_x)\ , & Classification \\ \int_\theta P(\ D_x|\ \theta_x)P(\theta_x)\ , & Regression \end{cases}$$

**MLE**

Given a random sample x complies a certain probability distribution with uncertain parameter $\theta$. Parameter is estimated by observing the result of several trials. Assuming $parameter\ \theta\ maximaizes\ the\ probability\ of\ the\ sample\ D_x$ then use $\theta$ as the value of estimation.

1. Likelihood function:

$$P(D_x|\theta_x) = \prod_{x \in D_x} P(x|\theta_x)$$

2. Logarithm Likelihood function:

$$LL(\theta_x) = log\ P(D_x|\theta_x) = \sum_{x \in D_x} log\ P(x|\theta_x)$$

$$Maximal\ Likelihood\ Estimation : \hat{\theta}_x = arg_{\theta_x}\ max\ LL(\theta_x) = arg_{\theta_x}\ max \sum_{x \in D_x} log\ P(x|\theta_x)$$

**Example: Bernoulli MLE**

1. Likelihood Function:

$$P(D_x) = \begin{cases} \theta, & D_x = yes \\ 1 - \theta, & D_x = no \end{cases}$$

$$P(D_x|\theta_x) = \prod_{x \in D_x} P(x|\theta_x) = \theta^n (1-\theta)^{N-n}$$

2. Logarithm:

$$logP(D_x|\theta_x) = log(\theta^n (1-\theta)^{N-n}) = nlog\theta + (N-n)log(1-\theta)$$

3. Derivation=0:

$$\frac{d}{d\theta_x} log(D_x|\theta_x) = \frac{n}{\theta} - \frac{N-n}{1-\theta} = \frac{n-N\theta}{\theta(1-\theta)} = 0$$
$$\theta = \frac{n}{N}$$

**Example: Gaussian MLE**

$$N(\vec{x} \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} exp[-\frac{(x-\mu)^2}{2\sigma^2}]$$
$$Log - Likelihood$$
$$Maximize\ Log - Likelihood$$
$$differential = 0$$
$$\begin{cases} \mu_{ML} = \frac{1}{N} \sum_i^N x_i \\ \sigma_{ML}^2 = \frac{1}{N} \sum_i^N (x_i - \mu_{ml})^2 \end{cases}$$