MLE ~~[illegible]~~

Maximum Likelihood Estimation (MLE)

Let $X_1, X_2, \cdots, X_n$ be a random sample (iid rvs) with common pdf $f_X(x, \theta)$, $\theta \in \Theta$. Also assume that $(x_1, x_2, \cdots, x_n)$ be a observed value of $(X_1, X_2, \cdots, X_n)$. Then likelihood function of θ given the data $x = (x_1, x_2, \cdots, x_n)$ is defined as

$$L(\theta \mid x) = \prod_{i=1}^{n} f_X(x_i, \theta). \quad \text{—— (1)}$$

A value of θ that maximizes $L(\theta|x)$ is called the Maximum Likelihood Estimator (MLE) of θ. Normally MLE of θ is denoted as $\hat{\theta} = \hat{\theta}(x)$. Thus

$$L(\hat{\theta}|x) = \max_{\theta \in \Theta} \{L(\theta|x)\}$$

⊛ Maximum likelihood estimation method is most widely used estimation method in statistical literature.

Before we discuss specific example, we must stress that when ever a maximum is sought by differentiation, the 2nd order derivatives must also be examined in search of a maximum.

It should also be mentioned that maximization of the likelihood function ① is equivalent to maximization of its logarithm function.

<u>Note</u>: Let us recall that a function $y = f(x)$ attains a maximum at a point $x = x_0$ if $\frac{d}{dx} f(x)\Big|_{x=x_0} = 0$ and $\frac{d^2}{dx^2} f(x)\Big|_{x=x_0} < 0$.

<u>Ex</u>: Let $X_1, X_2, \cdots, X_n$ be iid rvs from Bernoulli Ber(θ) dist^n. Find MLE of θ.

<u>Sol^n</u>: The likelihood func^n of θ given the observed data $x = (x_1, x_2, \cdots, x_n)$ is

$$L(\theta|x) = \prod_{i=1}^{n} f_X(x_i, \theta) = \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}$$

The log-likelihood function of θ is given by

$$l_n(\theta) = \log(L(\theta|x)) = \log\left(\theta^{\sum_{i=1}^{n} x_i} (1-\theta)^{n-\sum_{i=1}^{n} x_i}\right)$$

$$l(\theta) = \left(\sum_{i=1}^{n} x_i\right) \log\theta + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta)$$

$$\frac{d\,l(\theta)}{d\theta} = \frac{\sum_{i=1}^{n} x_i}{\theta} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-\theta}$$

$$\therefore \frac{d\,l(\theta)}{d\theta} = 0 \Rightarrow \theta = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{———} (*)$$

$\therefore$ Now let us ~~look~~ check the 2nd derivate

$$\frac{d^2 l(\theta)}{d\theta^2} = -\frac{\sum_{i=1}^{n} x_i}{\theta^2} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-\theta)^2}$$

$< 0 \quad \forall \ \theta$ such that $0 \leq \theta \leq 1$.

In particular if $\theta = \frac{1}{n}\sum_{i=1}^{n} x_i$ then also

$$\frac{d^2 l(\theta)}{d\theta^2} < 0.$$

Thus MLE of θ is

$$\boxed{\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}}$$

Ex: Determine the MLE of λ in Poisson $P(\lambda)$ distribution based on ~~a~~ random sample $X_1, X_2, \cdots, X_n$ taken from $P(\lambda)$ distn.

The likelihood function of λ is given by

$$L(\lambda|x) = \prod_{i=1}^{n} f(x_i, \lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

the log-likelihood function of λ is given by

$$l(\lambda) = \log(L(\lambda|x)) = \log\left[\frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} \right]$$

$$= -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \log\lambda - \log\left(\prod_{i=1}^{n} x_i!\right)$$

now

$$\frac{d l(\lambda)}{d\lambda} = -n + \frac{\sum_{i=1}^{n}(x_i)}{\lambda}$$

Thus $\frac{d l(\lambda)}{d\lambda} = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$.

Also $\frac{d^2 l(\lambda)}{d\lambda^2} = -\frac{\left(\sum_{i=1}^{n} x_i\right)}{\lambda^2} < 0 \ \forall \ \lambda > 0$.

Therefor MLE of λ is

$$\boxed{\hat{\lambda} = \bar{x}}$$

Ex: Let $x_1, x_2, \cdots, x_n$ be iid rvs from a $N(\mu, \sigma^2)$ distribution. Find MLE when following conditions hold

(i) μ is unknown, σ^2 known

(ii) μ is known, σ^2 unknown

(iii) μ unknown, σ^2 unknown.

Ans: (i) $\hat{\mu} = \bar{x}$

(ii) $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2$

(iii) $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Result: Invariance Property of MLE

Maximum likelihood Estimators satisfy a very useful property known as invariance property.

Suppose MLE of a parameter θ is $\hat{\theta}$. You are interested to know what is the MLE of a function of θ, say $g(\theta)$. Then MLE of $g(\theta)$ is $g(\hat{\theta})$. This is known as invariance property of MLE.

In the previous example suppose we ask:

(i) in part:

(ii) Find MLE of σ then your answer will be $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2}$

Similarly in part

(iii) $\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$.

In part (i) suppose we ask what is the MLE of e^{μ}. Your answer should be it is $e^{\bar{X}}$.

[In part (iii) what is MLE of $\mu + k\sigma$ where k is given constant.]

Ex: Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim} Ber(\theta)$. Find MLE of $\sqrt{\theta(1-\theta)}$.

Ex: Let $X_1, X_2, \cdots, X_n \overset{iid}{\sim}$ Poisson $P(\theta)$. Find MLE of $e^{-2\theta}$.

Ex: Let $x_1, x_2, \cdots, x_n$ be a random sample from a exponential $\exp(\lambda)$ dist^n. Find MLE of λ.

Ex: Let $x_1, x_2, \cdots x_n$ be a random sample from gamma $G(\alpha, \beta)$ dist^n. Find MLE
(i) of α when β known (ii) of β when α known
(iii) of α and β when both unknown.

Many more examples you can try yourself.