Maximum - Likelihood Estimator's

$X_1, X_2, X_3, \ldots, X_n \overset{iid}{\sim} f(x \mid \theta_1, \theta_2, \ldots \theta_k)$

$$L(\theta \mid x) = L(\theta_1, \theta_2, \ldots, \theta_k \mid x_1, x_2, \ldots, x_n)$$

$$= \prod_{i=1}^{n} f(x_i \mid \theta_1, \theta_2, \ldots, \theta_k)$$

Definition 7.2.4 :-

for each sample point x, let $\hat{\theta}(x)$ be a parameter value at which $L(\theta \mid x)$ attain's its maximum as a function of θ, with x held fixed. The Maximum Likelihood estimator (MLE) of the parameter θ based on a sample X is $\hat{\theta}(X)$.

two inherent drawbacks of finding MLE

(1) Actually finding the Global maximum and verifying that, indeed, a global maximum has been found.

(2) Numerical sensitivity. That is, how sensitive is the estimate to small changes in the data?

unfortunately it is sometimes the case that a slightly different sample will produce a vastly different MLE, making its use suspect.

if the likelihood function is differentiable (in θ_i), possible candidates for the MLE are the value of $(\theta_1, \theta_2, .. \theta_k)$ that solve

$$\frac{\partial}{\partial \theta_i} L(\theta | x) = 0 \qquad i = 1, 2, .. k.$$

Boundary must be checked seperately for extrema.

Example 7.2.5 (Normal Likelihood)

$$X_1, X_2, \ldots X_n \sim n(\theta, 1)$$

$$L(\theta|x) = \prod_{i=1}^{n} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}(x_i-\theta)^2}$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left\{ -\frac{1}{2}\left((n-1)s^2 + n(\bar{x}-\theta)^2 \right) \right\}$$

$$\frac{d}{d\theta} L(\theta|x) = L(\theta|x) \cdot \left\{ -\frac{1}{2} \cdot 2n(\bar{x}-\theta) - 1 \right\} = 0$$

$$\Rightarrow \quad n(\bar{x}-\theta) = 0$$

$$\Rightarrow \quad \boxed{\hat{\theta}_{MLE} = \bar{X}}$$

$$\frac{d^2}{d\theta} L(\theta|x) = -n < 0$$

$\Rightarrow$ Global maxima.

Example 7.2.7 (Bernoulli MLE)

$$X_1, X_2, \ldots, X_n \sim \text{Ber}(p)$$

$$L(p|x) = \prod_{i=1}^{n} p^{x_i} (1-p)^{1-x_i}$$

$$= p^{\sum x_i} (1-p)^{n - \sum x_i}$$

$$\log L(p|x) = \sum x_i \log p + \left(n - \sum x_i\right) \log(1-p)$$

$$\frac{d}{dp} \log L(p|x) = \sum x_i \cdot \frac{1}{p} + \frac{\left(n - \sum x_i\right)}{1-p} \cdot -1 = 0$$

$$\Rightarrow \quad \frac{\sum x_i}{p} = \frac{n - \sum x_i}{1-p}$$

$$\Rightarrow \quad \sum x_i - p \cancel{\sum x_i} = np - p \cancel{\sum x_i}$$

$$\boxed{\Rightarrow \quad \hat{p}_{MLE} = \frac{1}{n} \sum x_i}$$

Example 7.2.8 (Restricted range MLE)

$X_1, X_2, \ldots, X_n \sim N(\theta, 1)$

θ is nonnegative (given)

$\hat{\theta}_{MLE} = \bar{X}$ if $\bar{X}$ is nonnegative

$\hat{\theta}_{MLE} = 0$ if $\bar{X}$ is -ve.

Example 7.2.9 : (Binomial MLE, unknown number of trail's)

$X_1, X_2, \ldots, X_n \sim \text{binomial}(k, p)$

where p is known, k is unknown

$$f(x \mid k, p) = \binom{k}{x} p^{x} (1-p)^{k-x}$$

$$f(X \mid k, p) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}$$

$$L(k \mid x, p) = \prod_{i=1}^{n} \binom{k}{x_i} p^{x_i} (1-p)^{k-x_i}$$

what we know $k > \max_i x_i$

& MLE for k should be an integer,

& $$\frac{L(k \mid x, p)}{L(k-1 \mid x, p)} \geq 1$$

and $$\frac{L(k+1 \mid x, p)}{L(k \mid x, p)} < 1$$

$$\Rightarrow \frac{L(k \mid x, p)}{L(k-1 \mid x, p)} = \frac{\prod_{i=1}^{n} \binom{k}{x_i} \cancel{p^{x_i}} (1-p)^{k-x_i}}{\prod_{i=1}^{n} \binom{k-1}{x_i} \cancel{p^{x_i}} (1-p)^{k-1-x_i}}$$

$$= \frac{(1-p)^n \prod_{i=1}^{n} \binom{k}{x_i}}{\prod_{i=1}^{n} \binom{k-1}{x_i}}$$

$$\binom{k}{x_i} = \frac{k!}{(k-x_i)!\, x_i!}$$

$$= \frac{k \cdot (k-1)!}{(k-x_i)(k-1-x_i)!\, x_i!} = \frac{k}{(k-x_i)} \cdot \binom{k-1}{x_i}$$

$$\Rightarrow \frac{L(k|x,p)}{L(k-1|x,p)} = \frac{(1-p)^n \quad k^n}{\prod_{i=1}^{n} (k-x_i)}$$

$$= \frac{(k(1-p))^n}{\prod_{i=1}^{n} (k-x_i)} \geq 1$$

$$\Rightarrow (k(1-p))^n \geq \prod_{i=1}^{n} (k-x_i) \quad - 1$$

$$\{(k+1)(1-p)\}^n < \prod_{i=1}^{n} (k+1-x_i) \quad - 2$$

divide by k^n & $z = \frac{1}{k}$

$$(1-p)^n \geq \prod_{i=1}^{n} (1-x_i z)$$

Invariance Property of MLE:-

if $\hat{\theta}$ is the MLE for θ, then $\tau(\hat{\theta})$ will be the MLE for $\tau(\theta)$

(i) if the mapping $\theta \longmapsto \tau(\theta)$ is one-to-one then inverse exists

let $\eta = \tau(\theta) \Rightarrow \tau^{-1}(\eta) = \theta$

$$L^*(\eta|x) = \prod_{i=1}^{n} f(x_i \mid \tau^{-1}(\eta))$$

$$= L(\tau^{-1}(\eta)|x)$$

$$\sup_{\eta} L^*(\eta|x) = \sup_{\eta} L(\tau^{-1}(\eta)|x)$$

$$= \sup_{\theta} L(\theta|x)$$

Thus the maximum of $L^*(\eta|x)$ is attained at $\eta = \tau(\theta) = \tau(\hat{\theta})$

Induced likelihood function: L^*

$$(7.2.5) \quad - \quad L^*(\eta|x) = \sup_{\{\theta:\, \tau(\theta)=\eta\}} L(\theta|x)$$

The value of $\hat{\eta}$ that maximizes $L^*(\eta|x)$ will be called the MLE of $\eta = \tau(\theta)$; and it can be seen from (7.2.5) that the maxima of L^* and L coincide.

This mean's for each value of η, we maximize the likelihood function over all θ values that satisfy $\tau(\theta) = \eta$. This ensures that the likelihood function for η is as large as possible, leading to the best MLE of η.

Theorem 7.2.10 :- (Invariance Property of MLE)

if $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$

if $\hat{\theta}$ is MLE for parameter θ

then $\tau(\hat{\theta})$ is MLE for $\tau(\theta)$

$\Rightarrow$ $\hat{\theta}^2$ is MLE for θ^2

$\Rightarrow$ $\sqrt{\hat{p}(1-\hat{p})}$ is MLE for $\sqrt{p(1-p)}$

Example 7.2.11 (Normal MLE's μ and σ unknown)

$X_1, X_2, \ldots, X_n \overset{iid}{\sim} n(\theta, \sigma^2)$

$$L(\theta, \sigma^2 | x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{ \frac{-1}{2\sigma^2} \left((n-1)S + n(\bar{x}-\mu)^2 \right) \right\}$$

$$\log L(\theta, \sigma^2 | x)$$

$$= -\frac{n}{2}\log 2\pi \cdot -\frac{n}{2}\log\sigma^2$$

$$-\frac{1}{2\sigma^2}\left((n-1)s^2 + n(\bar{x}-\theta)^2\right)$$

$$\frac{\partial}{\partial\theta}\log L(\theta, \sigma^2 | x)$$

$$= -\frac{n}{2\sigma^2}\cdot 2(\bar{x}-\theta)\cdot -1$$

$$= \frac{n}{\sigma^2}(\bar{x}-\theta) = 0$$

$$\Rightarrow \hat{\theta} = \bar{x}$$

$$\frac{\partial}{\partial\sigma^2}\log L(\theta, \sigma^2 | x)$$

$$= -\frac{n}{2}\frac{1}{\sigma^2} - \frac{1}{2}\left((n-1)s^2 + n(\bar{x}-\theta)^2\right)\cdot -\frac{1}{\sigma^4}$$

$$= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\left((n-1)s^2 + n(\bar{x}-\theta)^2\right) = 0$$

$$\sigma^2 = \frac{1}{n}\left((n-1)s^2 + n(\bar{x}-\theta)^2\right)$$

$$\sigma^2 = \frac{1}{n}\left((n-1)s^2 + n(\bar{x}-\bar{x})^2\right)$$

$$\Rightarrow \quad \sigma^2 = \frac{1}{n}(n-1)s^2$$

$$\hat{\theta} = \bar{x}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$$

Example 7.2.12

To use two-variate calculus to verify that a function $H(\theta_1, \theta_2)$ has a local maximum at $(\hat{\theta}_1, \hat{\theta}_2)$, it must be

(a) The first-order partial derivatives are 0

$$\frac{\partial}{\partial \theta_1} H(\theta_1, \theta_2)\Big|_{\theta_1 = \hat{\theta}_1,\ \theta_2 = \hat{\theta}_2} = 0$$

$$\frac{\partial}{\partial \theta_2} H(\theta_1, \theta_2)\Big|_{\theta_1 = \hat{\theta}_1,\ \theta_2 = \hat{\theta}_2} = 0$$

(b) At least one second-order partial derivative is negative,

$$\frac{\partial^2}{\partial \theta_1^2} H(\theta_1, \theta_2)\Big|_{\theta_1=\hat{\theta}_1,\ \theta_2=\hat{\theta}_2} < 0$$

or

$$\frac{\partial^2}{\partial \theta_2^2} H(\theta_1, \theta_2)\Big|_{\theta_1=\hat{\theta}_1,\ \theta_2=\hat{\theta}_2} < 0$$

(c) The Jacobian of the second-order partial derivates is +ve.

$$\left|\begin{matrix} \frac{\partial^2}{\partial \theta_1^2} H(\theta_1,\theta_2) & \frac{\partial^2}{\partial \theta_1 \partial \theta_2} H(\theta_1,\theta_2) \\ \frac{\partial^2}{\partial \theta_2 \partial \theta_1} H(\theta_1,\theta_2) & \frac{\partial^2}{\partial \theta_2^2} H(\theta_1,\theta_2) \end{matrix}\right|_{\theta_1=\hat{\theta}_1,\ \theta_2=\hat{\theta}_2} > 0$$

log-normal :

$$L(\theta|x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \Sigma (x_i-\theta)^2\right\}$$

$$\log L(\theta|x) = -\frac{n}{2}\log\sigma^2 - \frac{n}{2}\log 2\pi - \frac{1}{2\sigma^2}\Sigma(x_i-\theta)^2$$

$(x_1-\theta)^2 + (x_2-\theta)^2 + \ldots$

$$\frac{\partial^2}{\partial\theta^2} \log L(\theta|x) = \frac{\partial}{\partial\theta}\left(-\frac{\cancel{2}}{\cancel{2}\sigma^2} \sum(x_i-\theta) - 1 \right)$$

$$= \frac{\partial}{\partial\theta}\left(\frac{1}{\sigma^2} \sum(x_i-\theta) \right)$$

$$= -\frac{n}{\sigma^2} < 0$$

$$\frac{\partial^2}{\partial(\sigma^2)^2} \log L(\theta,\sigma^2|x)$$

$$= \frac{\partial}{\partial\sigma^2}\left[-\frac{n}{2\sigma^2} + \frac{\sum(x_i-\theta)^2}{2} \cdot \frac{1}{\sigma^4} \right)$$

$$= \frac{\partial}{\partial\sigma^2}\left(-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum(x_i-\theta)^2 \right)$$

$$= \frac{n}{2\sigma^4} - \frac{2}{2\sigma^6} \sum(x_i-\theta)^2$$

$$= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum(x_i-\theta)^2$$

$$\frac{\partial^2}{\partial\sigma^2\,\partial\theta} \log L(\theta,\sigma^2)x)$$

$$= \frac{\partial}{\partial\sigma^2}\left(\frac{1}{\sigma^2} \sum(x_i-\theta) \right)$$

$$= -\frac{1}{\sigma^4} \sum(x_i-\theta)$$

$$J = \begin{vmatrix} \dfrac{-n}{\sigma^2} & -\dfrac{1}{\sigma^4}\sum(x_i-\theta) \\ -\dfrac{1}{\sigma^4}\sum(x_i-\theta) & \dfrac{n}{2\sigma^4} - \dfrac{1}{\sigma^6}\sum_{i=1}^{n}(x_i-\theta)^2 \end{vmatrix}$$

$$J = \frac{-n^2}{2\sigma^6} + \frac{n}{\sigma^8}\sum_{i=1}^{n}(x_i-\theta)^2 - \frac{1}{\sigma^8}\sum(x_i-\theta)\sum(x_i-\theta)$$

$$J = \frac{1}{\sigma^6}\left[\frac{-n^2}{2} + \frac{n}{\sigma^2}\sum_{i=1}^{n}(x_i-\theta)^2 - \frac{1}{\sigma^2}\sum(x_i-\theta)\sum(x_i-\theta)\right]$$

$$J = \frac{1}{\sigma^6}\left[\frac{-n^2}{2} + \frac{n}{\frac{1}{n}\cancel{\sum_{i=1}^{n}(x_i-\theta)^2}}\cancel{\sum_{i=1}^{n}(x_i-\theta)^2} - \frac{1\left(\sum(x_i-\theta)\right)^2}{\frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2}\right]$$

$$J = \frac{1}{\sigma^6}\left[\frac{-n^2}{2} + n^2 - n\frac{\Big[\overbrace{\sum_{i=1}^{n}(x_i-\bar{x})}^{0}\Big]^2}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right]$$

$$J = \frac{n^2}{2\sigma^6} > 0 \iff \text{its global maximum}$$

The amount of calculation's, even in this simple problem, is formidable, and things will only get worse in higher orders.

⇒ Finally, it was mentioned earlier that, since MLE's are found by a maximization process, they are susceptible to the problem's associated with that process, among them that of numerical instability.