III

Maximum Likelihood Estimators (MLE)

$\mathbf{X} \sim f(\underline{x} \mid \underline{\theta})$

Likelihood function $L(\underline{\theta} \mid \mathbf{x}) \equiv \prod_{i=1}^{n} f(x_i \mid \underline{\theta})$

Def: Maximum likelihood estimator (MLE) of the parameter $\underline{\theta}$ based on a sample $\underline{X}$ is $\underline{\hat{\theta}}(\underline{X})$, where

$$L\left(\underline{\hat{\theta}}(\underline{x}) \mid \underline{x}\right) = \sup_{\underline{\theta}} L(\underline{\theta} \mid \underline{x}) \quad , \forall \text{ sample point } \underline{x}.$$

Note: 1° The MLE is the parameter point for which the observed sample is most likely.

2° Inherent drawbacks of finding the maximum of a fn:

1) finding and verifying the global maximum
2) numerical sensitivity of the maximum.

Def: Induced likelihood function L^* for $\eta = \tau(\theta)$ is

$$L^*(\eta \mid \underline{x}) \equiv \sup_{\{\theta : \tau(\theta) = \eta\}} L(\theta \mid \underline{x})$$

The value $\hat{\eta}$ that maximizes $L^*(\eta \mid \underline{x})$ will be called the MLE of $\eta = \tau(\theta)$.

Thm: (Invariance property of MLEs)

If $\hat{\theta}$ is MLE of θ, then $\forall$ fn. $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Proof: 1° The maxima of L and L^* coincide.

$$L^*(\hat{\eta} \mid \underline{x}) \equiv \sup_{\eta} \sup_{\{\theta : \tau(\theta) = \eta\}} L(\theta \mid \underline{x})$$
$$= \sup_{\theta} L(\theta \mid \underline{x})$$
$$= L(\hat{\theta} \mid \underline{x})$$

2° $L(\hat{\theta} \mid \underline{x}) = \sup_{\{\theta : \tau(\theta) = \tau(\hat{\theta})\}} L(\theta \mid \underline{x}) \qquad \left(\hat{\theta} \in \{\theta : \tau(\theta) = \tau(\hat{\theta})\} \subseteq \Theta\right)$

$$= L^*(\tau(\hat{\theta}) \mid \underline{x})$$

$\therefore L^*(\tau(\hat{\theta}) \mid \underline{x}) = L^*(\hat{\eta} \mid \underline{x}) \quad \square$

Note: 1° There's nothing in the proof that precludes θ from being a vector, so the invariance property of MLE also holds in the multivariate case.

Properties:

1. In most cases, MLEs are consistent

1° MLE's are asymptotically efficient. (Asymptotic property)

In most cases,

2° MLE is always fn. of sufficient stat.

Techniques for finding/verifying MLE's:

1° differentiation

2° direct maximization (Unique attainable global upper bound).

3° log likelihood ($u(\hat{\theta}) = \frac{\partial}{\partial\theta} \log p(x;\theta) = 0$; MLE solves the score fn.)

4° successive maximizations

5° the EM algorithm

Note: 1° Using second derivative condition to check for maximum likelihood requires negative definiteness of the Hessian matrix, which is formidable.

2° It is always important to analyse the likelihood fn. as much as possible, to find the number and nature of its local maxima, before using numeric maximization.

The EM Algorithm:

This is an algorithm suited to find MLE for missing data problems, by constructing a sequence that is guaranteed to converge to the MLE.

$\underline{Y} = (Y_1, \cdots, Y_n)$ incomplete data $\sim g(\underline{y} \mid \theta)$

$\underline{X} = (X_1, \cdots, X_m)$ augmented data.

$(\underline{Y}, \underline{X})$ complete data $\sim f(\underline{y}, \underline{x} \mid \theta)$