Some properties of MLE (without proof)

- MLE may be biased or unbiased
- MLE is consistent estimator.
- If an efficient estimator exists, it is the MLE estimator.

An efficient estimator $\hat{\theta}$ exists =>

$$\frac{\partial}{\partial\theta} L(\mathbf{x}/\theta) = c(\hat{\theta} - \theta)$$

at $\theta = \hat{\theta}$,

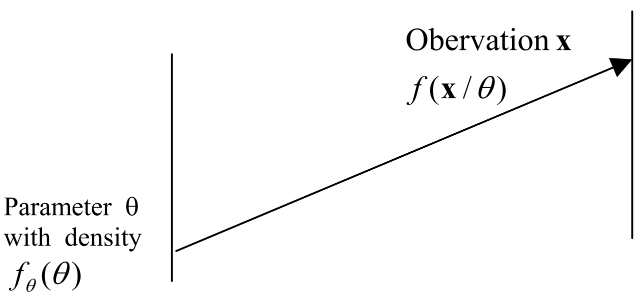
$$\frac{\partial L(\mathbf{x}/\theta)}{\partial\theta}\Big|_{\hat{\theta}} = c(\hat{\theta} - \hat{\theta}) = 0$$

$\Rightarrow \hat{\theta}$ is the MLE estimator.

Bayescan Estimators :

We may have some prior information about θ in a sense that some values of θ are more likely.(*a priori* information). We can represent this prior information in the form of a prior density function.

In the folowing we omit the suffix in density functions just for notational simplicity.



The likelihood function will now be the conditional density $f(\mathbf{x}/\theta)$.

$$f(\mathbf{x},\theta) = f(\theta) f(\mathbf{x}\,|\,\theta)$$

Also we have the Bayes rule

$$f(\theta\,|\,\mathbf{x}) = \frac{f(\theta) f(\mathbf{x}\,|\,\theta)}{f(\mathbf{x})}$$

where $f(\theta\,|\,\mathbf{x})$ is the *a posteriori* density function

The parameter θ is a random variable and the estimator $\hat{\theta}(\mathbf{x})$ is another random variable.

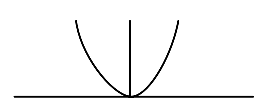
Estimation error $\varepsilon = \hat{\theta} - \theta$.

We associate a cost function $C(\hat{\theta}, \theta)$ with every estimator $\hat{\theta}$. It represents postive penalty with each wrong estimation.

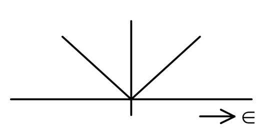
Thus $C(\hat{\theta}, \theta)$ is a non negative function.

The three most popular cost functions are:

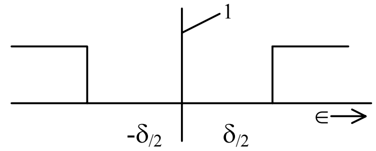
Quadratic cost function $(\hat{\theta} - \theta)^2$



Absolute cost function $\left|\hat{\theta} - \theta\right|$



Hit or miss cost function (also called uniform cost function
(minimising means minimising on an average)



Bayesean Risk function or average cost

$$\overline{C} = EC(\theta,\, \hat{\theta}) \;=\; \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(\theta,\hat{\theta})\, f(\mathbf{x},\, \theta) d\mathbf{x}\, d\theta$$

The estimator seeks to minimize the Bayescan Risk.

Case I. QUADRATIC COST FUNCTION

$$C = (\theta\,,\ \hat{\theta}) = (\hat{\theta} - \theta)^2$$

Estimation problem is

Minimize $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\theta-\hat{\theta})^2 f(\mathbf{x},\, \theta\,) d\mathbf{x}\, d\theta$

with respect to $\hat{\theta}$.

This is equivalent to minimizing

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\theta-\hat{\theta})^2 f(\theta\,|\,\mathbf{x}) f(\mathbf{x}) d\theta\, d\mathbf{x}$$

$$= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\infty} (\theta-\hat{\theta})^2 f(\theta\,|\,\mathbf{x}) d\theta \Big) f(\mathbf{x})\, d\mathbf{x}$$

Since $f(\mathbf{x})$ is always +ve, the above integral will be minimum if the inner integral is minimum. This results in the problem:

Minimize $\displaystyle\int_{-\infty}^{\infty} (\theta-\hat{\theta})^2 f(\theta\,|\,\mathbf{x}) d\theta)$

with respect to $\hat{\theta}$.

$$\Rightarrow \frac{\partial}{\partial\hat{\theta}} \int_{-\infty}^{\infty} (\hat{\theta}-\theta)^2 f(\theta\,|\,\mathbf{x})\, d\theta \;=\; 0$$

$$\Rightarrow -2\int_{-\infty}^{\infty} (\hat{\theta}-\theta)\, f(\theta\,|\,\mathbf{x})\, d\theta \;=\; 0$$

$$\Rightarrow \hat{\theta} \int_{-\infty}^{\infty} f(\theta\,|\,\mathbf{x})\, d\theta \;=\; \int_{-\infty}^{\infty} \theta\, f(\theta\,|\,\mathbf{x})\, d\theta$$

$$\Rightarrow \hat{\theta} = \int_{-\infty}^{\infty} \theta\, f(\theta\,|\,\mathbf{x})\, d\theta$$

$\therefore \hat{\theta}$ is the conditional mean or mean of the a posteriori density. Since we are minimizing quadratic cost it is also called *minimum mean square error estimator* (MMSE).

Salient Points

- Information about distribution of θ available.
- *a priori* density function $f(\mathbf{x}\,|\,\theta)$ is available. This denotes how observed data depend on θ
- We have to determine a posteriori density $f(\theta\,|\,\mathbf{x})$. This is determined form the

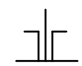
Estimated density of the observed data

$$f(\theta \mid \mathbf{x}) = \frac{f(\theta) f(\mathbf{x} \mid \theta)}{f(\mathbf{x})}$$

Case II HIT OR MISS COST FUNCTION

Risk $\overline{C} = EC(\theta, \hat{\theta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(\theta,\hat{\theta})\, f(\mathbf{x}, \theta) d\mathbf{x}\, d\theta$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} c(\theta,\hat{\theta})\;\; f(\theta \mid \mathbf{x}) f(\mathbf{x}) d\theta\, d\mathbf{x}$$



$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} c(\theta,\hat{\theta})\, f(\theta \mid \mathbf{x}) d\theta \right) f(\mathbf{x})\, d\mathbf{x}$$

We have to minimize

$$\int_{-\infty}^{\infty} C\,(\theta\; \hat{\theta})\, f(\theta\; \mathbf{x})\; d\theta \qquad \text{with respect to } \hat{\theta}.$$

This is equivalent to minimizing

$$= 1 - \int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} f\,(\theta|\mathbf{x})\; d\theta$$

This minimization is equivalent to maximization of

$$\int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} f\,(\theta|\mathbf{x})\; d\theta \cong \Delta f\,(\hat{\theta}|\mathbf{x}) \qquad \text{when } \Delta \text{ is very small}$$

This will be maximum of $f\,(\hat{\theta}|\mathbf{x}$ is maximum. That means select that value of $\hat{\theta}$ that maximizes the a posteriori density. So this is known as maximum a posteriori estimation (MAP) princide.

This estimator is denoted by $\hat{\theta}_{MAP}$.

Example

Let $X_1, X_2 ...., X_N$ be an *iid* Gaussian sequence with unity Variance and unknown mean θ. Further θ is known to be a 0-mean Gaussian with Unity Variance. Find the MAP estimator for θ.

Solution We are given

$$f_{\theta}(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\theta^2}$$

$$f(\mathbf{x} / \theta) = \frac{1}{(\sqrt{2\pi})^N} e^{-\sum_{i=1}^{N} \frac{(x_i - \theta)^2}{2}}$$

There fore $f(\theta \mid \mathbf{x}) = \dfrac{f(\theta) f(\mathbf{x} \mid \theta)}{f(\mathbf{x})}$

We have to find θ, such that

$f(\theta \mid \mathbf{x})$ is maximum.

Now $f(\theta \mid \mathbf{x})$ is maximum when $f(\theta) f(\mathbf{x} \mid \theta)$ is maximum.

$\Rightarrow \ln f(\theta) f(\mathbf{x} \mid \theta)$ is maximum

$\Rightarrow -\dfrac{1}{2}\theta^2 - \sum_{i=1}^{N} \dfrac{(x_i - \theta)^2}{2}$ is maximum

$$\Rightarrow \theta - \sum_{i=1}^{N} (x_i - \theta) \Bigg]_{\theta = \hat{\theta}_{MAP}} = 0$$

$$\Rightarrow \hat{\theta}_{MAP} = \frac{1}{N+1} \sum_{i=1} x_i$$