THE MAXIMUM LIKELIHOOD METHOD(MLM)

Suppose we have a sample . . .

. . . sample — **x**: $X_1 = x_1, X_2 = x_2, \ldots, X_N = x_N$ (N random variables) and we compute the *a posteriori* Probability of obtaining such a sequence

$$d\mathcal{P} = f(\mathbf{x};\boldsymbol{\lambda})d\mathbf{x} = f(x_1,\ldots,x_N;\lambda_1,\ldots,\lambda_p)dx_1\ldots dx_N \quad \text{or}$$

$$d\mathcal{P} = \prod_{j=1}^{N} f(x_j;\boldsymbol{\lambda})dx_j; \qquad \boldsymbol{\lambda} = \lambda_1,\ldots,\lambda_p$$

We introduce the Likelihood Function

$$L = \prod_{j=1}^{N} f(x_j;\boldsymbol{\lambda})$$

and the Likelihood Quotient with numerator and denominator being L calculated for 2 $\boldsymbol{\lambda}$'s:

$$Q = \frac{L(\boldsymbol{\lambda}_1)}{L(\boldsymbol{\lambda}_2)} = \frac{\prod_{j=1}^{N} f(x_j;\boldsymbol{\lambda}_1)}{\prod_{j=1}^{N} f(x_j;\boldsymbol{\lambda}_2)}$$

parameters $\boldsymbol{\lambda}_1$ are Q times more likely to occur (or: more plausible) than the parameters $\boldsymbol{\lambda}_2$)

Example:

Suppose we have a coin which – as we happen to know – is not a fair one. Namely – one side is likely to happen twice as frequently as the second one but ... we do not know which one.
we perform an experiment: flipping the coin 6 times and we get 4 heads (and 2 tails). We have two possibilities:

First case	Second case
(1) $\mathcal{P}(H) = 2/3; \quad \mathcal{P}(T) = 1/3$ $W_4^6 = \binom{6}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2$	(2) $\mathcal{P}(H) = 1/3; \quad \mathcal{P}(T) = 2/3$ $W_4^6 = \binom{6}{4}\left(\frac{1}{3}\right)^4\left(\frac{2}{3}\right)^2$

$$\frac{L_1}{L_2} = \frac{\binom{6}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2}{\binom{6}{4}\left(\frac{2}{3}\right)^4\left(\frac{1}{3}\right)^2} = \ldots = 4.$$

Obviously, the first hypothesis about the $\mathcal{P}$'s is the better one.

The "best" parameters $\boldsymbol{\lambda}$

will be the ones that maximise the L: $L = L_{max}$.
For the sake of convenience it is practical to use the logarithmic likelihood function:

$$l = \ln L = \ln \left\{ \prod_{j=1}^{N} f(x_j; \boldsymbol{\lambda}) \right\} = \sum_{j=1}^{N} \ln f(x_j; \boldsymbol{\lambda})$$

The maximum of L (or l) will be attained if:

$$\frac{\partial l}{\partial \lambda_i} = 0; \quad i = 1, 2, \ldots, p$$

the derivative:

$$\frac{\partial l}{\partial \lambda_i} = \sum_{j=1}^{N} \frac{\partial}{\partial \lambda_i} \left[\ln f(x_j; \boldsymbol{\lambda}) \right] = \sum_{j=1}^{N} \frac{\partial f / \partial \lambda_i}{f} \equiv \sum_{j=1}^{N} \phi(x_j; \boldsymbol{\lambda})$$

is called the information (of the sample) with respect to the estimated parameter λ_i.

Example:

$x_1, x_2, \ldots, x_n$ is drawn from the Poisson population with the distribution

$$f(x) = e^{-\lambda} \frac{\lambda^x}{x!}$$

(λ unknown). The L function is:

$$L(x_1, \ldots, x_n; \lambda) = \frac{e^{-n\lambda} \ \lambda^{x_1 + x_2 + \ldots + x_n}}{x_1! \, x_2! \ \ldots \ x_n!}.$$

We determine $\hat{\lambda}$ from the condition $d \ln L / d\lambda = 0$. It gives
$\hat{\lambda} = (x_1 + x_2 + \ldots + x_n)/n$.
The λ parameter is simply the arithmetic average.

Suppose we have an n-element sample:

$x_1, x_2, \ldots, x_n$ drawn from a normal distribution $N(\mu, \sigma)$.

$$L(x_1, \ldots, x_n; \mu, \sigma) = \frac{1}{\sigma^n \left(\sqrt{2\pi}\right)^n} \exp\left[-\frac{(x_1-\mu)^2 + \ldots + (x_n-\mu)^2}{2\sigma^2}\right]$$

and its logarithm

$$\ln L = l = -n \ln \sigma - \frac{n}{2} \ln 2\pi - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2.$$

Now, we want to adjust μ and σ that maximise l. Thus

$$\frac{\partial \ln L}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

and we obtain the MLM estimators:

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

A very interesting and important case

is when the values of x_i which constitute the sample are drawn with different variances (accuracies). We have:

$$\begin{aligned} X_1 &\hat{=} N(\mu, \sigma_1) \\ X_2 &\hat{=} N(\mu, \sigma_2) \\ &\vdots \\ X_n &\hat{=} N(\mu, \sigma_n) \end{aligned}$$

The modified formulae are:

$$L(x_1, \ldots, x_n; \mu, \sigma) = \frac{1}{\sigma_1 \sigma_2 \ldots \sigma_n} \frac{1}{(\sqrt{2\pi})^n} \exp\left[-\frac{(x_1 - \mu)^2}{2\sigma_1^2} - \ldots - \frac{(x_n - \mu)^2}{2\sigma_N^2} \right]$$

and its logarithm

$$\ln L = l = -\sum_{i=1}^{n} \ln \sigma_i - \frac{n}{2} \ln 2\pi - \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{2\sigma_i^2}.$$

The MLM estimator for μ is a weighted mean:

A very interesting and important case

The MLM estimator for μ is a weighted mean:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \times x_i}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}},$$

with the weights being equal to the reciprocal of variances. (The more "accurate" value the bigger is its contribution to the $\hat{\mu}$.)