Maximum Likelihood Estimation

- Different populations generate different samples, and that any sample is more likely to have come from some populations than from others.

Define maximum likelihood estimators as follows:

If a random variable X has a probability distribution f(x) characterised by parameters $\theta_1, \theta_2, \theta_3, \ldots.., \theta_k$ *and if we observe a sample* $x_1, x_2, x_3 \ldots\ldots.. x_n$*, then the maximum likelihood estimators of* $\theta_1, \theta_2, \theta_3, \ldots.., \theta_k$ *are those values of the parameters that would generate the sample most often.*

In other words, the maximum likelihood estimators of $\theta_1, \theta_2, \theta_3, \ldots.., \theta_k$ are those values for which the probability of a given set of sample values is at a maximum.

The Likelihood Function

The concept of a likelihood function is a name given to the formula of the joint probability distribution of the sample. That is, if x,y,z are independent then we have:

$$f(x, y, z \ldots\ldots) = f(x) f(y) f(z) \ldots.$$

Then the likelihood function is described by the formula for the joint density:

$$l = f(x_1, x_2, x_3, \ldots\ldots. x_n)$$

$$l = f(x_1) f(x_2) f(x_3) \ldots\ldots. f(x_n)$$

The interpretation of the joint density and the likelihood function is different however:

- In the case of the joint probability distribution

$$\theta_1, \theta_2, \ldots\ldots. \theta_n$$

are considered as fixed and the x's (sample observation) vary

- In the case of the likelihood function the parameters

$$\theta_1, \theta_2, \ldots\ldots. \theta_n$$

may vary whilst the observations are considered fixed. The maximum likelihood is found by maximising the likelihood function with respect to the parameters. That is a necessary condition for a maximum is that the following first order conditions be satisfied:

$$\frac{\partial l}{\partial \theta_1} = 0, \frac{\partial l}{\partial \theta_2} = 0, \frac{\partial l}{\partial \theta_3} = 0, \ldots\ldots\ldots \frac{\partial l}{\partial \theta_k} = 0$$

Notes

- if we have k unknowns we have to solves k equations.
- certain second order conditions also need to be fulfilled.
- taking logs is a monotonic transformation of the likelihood. Finding the maximum of the log-likelihood is therefore equivalent to finding the maximum of the likelihood. Since in practice the solution of the the first order conditions generally turns out to be easier using the log-likelihood, we put

$$L = \ln l$$

and find the parameter values that satisfy

$$\frac{\partial L}{\partial \theta_1} = 0, \frac{\partial L}{\partial \theta_2} = 0, \frac{\partial L}{\partial \theta_3} = 0, \ldots\ldots\ldots \frac{\partial L}{\partial \theta_k} = 0$$

Example: Suppose that X is a binary variable that can take the value one with probability π, and otherwise takes the value zero i.e.

$$f(0) = 1 - \pi$$
$$f(1) = \pi$$

Then the probability distribution of X can be described by:

$$f(x) = (1-\pi)^{1-x}\pi^x$$

Check this by substitution:

$$f(0) = (1-\pi)^{1-0}\pi^0 = 1 - \pi$$

$$f(1) = (1-\pi)^{1-1}\pi^1 = \pi$$

which agrees with our earlier specification

$$f(x) = (1-\pi)^{1-x}\pi^x$$

$$l = f(x_1)f(x_2)\ldots f(x_n)$$

$$l = [(1-\pi)^{1-x_1}\pi^{x_1}][(1-\pi)^{1-x_2}\pi^{x_2}]\ldots.[(1-\pi)^{1-x_n}\pi^{x_n}]$$

$$l = [(1-\pi)^{(1-x_1)+(1-x_2)+\ldots..+(1-x_n)}\pi^{x_1+x_2+\ldots\ldots.+x_n}$$

$$l = [(1-\pi)^{n-\sum x_i}\pi^{\sum x_i}]$$

Now, if we take logarithms of this function

$$L = \left(n - \sum x_i\right)\log(1-\pi) + \left(\sum x_i\right)log\pi$$

The only unknown parameter is π. Therefore differentiating with respect to π gives us

$$\frac{dL}{d\pi} = \left(n - \sum x_i\right)\frac{1}{(1-\pi)}(-1) + \left(\sum x_i\right)\frac{1}{\pi}$$

Equating this to zero and solving for the estimator of π, which we shall call $\hat{\pi}$, we obtain:

$$\frac{\left(\sum x_i\right) - n}{(1-\hat{\pi})} + \frac{\sum x_i}{\hat{\pi}} = 0$$

Multiplying both sides by $\hat{\pi}(1-\hat{\pi})$ gives us

$$\left[\left(\sum x_i\right) - n\right]\hat{\pi} + \left(\sum x_i\right)(1-\hat{\pi}) = 0$$

which gives;

$$\hat{\pi} = \frac{1}{n}\sum x_i$$

Since $\sum x_i$ is the number of successes and n is the number of all observations in the sample, the MLE of π is simply the proportion of successes found in the sample. In the case that we considered earlier, in which the sample was $\{1,1,0\}$, the MLE of π is $\hat{\pi} = 2/3 \cong 0.7$