FE2709 Maximum Likelihood Estimation (MLE)

Maximum likelihood estimation (MLE) is a technique used to estimate a parameter or set of parameters that determine the probability distribution function or probability density function of a random variable having observed a random sample $X_1, \ldots, X_N$ of size N, which has been drawn from the distribution. Knowing the value of the parameter gives a better understanding of the distribution and allows one to predict more effectively the future outcomes of the random variable. An example will serve to illustrate the application of MLE.

As a simple model for the behavior of the NASDAQ, suppose that it goes up on a given day with probability $0 \leq \theta \leq 1$, independent of its history. Observations over the past ten days have shown that the NASDAQ has gone up on the first two days and down on the following eight days. MLE provides an estimate of the probability θ given only these observations. Assuming that the probability of an up move is θ, the probability of observing the past ten moves is $L(\theta) = \theta^2(1-\theta)^8$. For this particular example, this function of θ is called the *likelihood function*, and is illustrated in Figure 1. Maximizing this function of θ in the range $[0, 1]$ furnishes the maximum likelihood estimate of θ given by $\hat{\theta} = 0.2$.

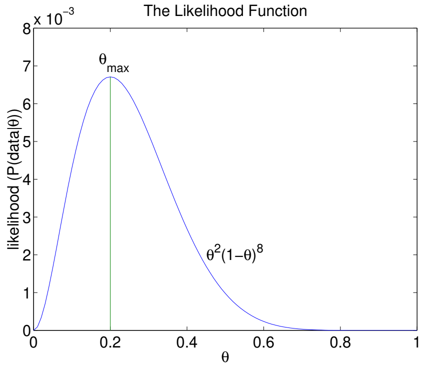

Figure 1. The likelihood function.

Definition of the Maximum Likelihood Estimator (MLE). Suppose that the random variables $X_1, \ldots, X_N$ form a random sample from a discrete or continuous distribution for which the probability distribution function or probability density function is given by $f(x|\theta)$, where θ belongs to some parameter space Θ. θ may be a real valued parameter or a vector. For any observed vector $\mathbf{x} = (x_1, \ldots, x_N)$ in the sample, the value of the joint distribution or density function, denoted $f_N(\mathbf{x}|\theta)$, when viewed as a function of θ for a fixed $\mathbf{x}$ is called the *likelihood function*. In the special case that the X's form an *independent* random sample, the likelihood is given by the product $f(x_1|\theta)f(x_2|\theta)\cdots f(x_N|\theta)$.

For each possible observed vector $\mathbf{x}$, let $\delta(\mathbf{x}) \in \Theta$ denote a value of $\theta \in \Theta$ for which the likelihood function $f_N(\mathbf{x}|\theta)$ is a maximum,

$$\delta(\mathbf{x}) = \operatorname*{argmax}_{\theta \in \Theta} f_N(\mathbf{x}|\theta). \tag{1}$$

The estimator $\hat{\theta} = \delta(\mathbf{X})$ defined in this way is called the *maximum likelihood estimate* or *maximum likelihood estimator* of θ, abbreviated MLE. It should be noted that maximizing any monotonically increasing function of the likelihood with respect to θ will yield an equivalent MLE.

Examples. In all of the examples we use $\overline{X_N}$ to denote the average value of a random sample $X_1, \ldots, X_N$ and $\overline{X_N^2}$ to denote the average squared value.

(i) *Sampling from a Bernoulli distribution.* Suppose the $X_1, \ldots, X_N$ form an independent random sample from a Bernoulli distribution with unknown parameter $0 \le \theta \le 1$. For any value of $\mathbf{x}$ where each x_i is either 0 or 1, the likelihood function is

$$f_N(\mathbf{x}|\theta) = \prod_{i=1}^{N} \theta^{x_i}(1-\theta)^{1-x_i}. \tag{2}$$

and the MLE of θ is $\hat{\theta} = \overline{X_N}$

(ii) *Sampling from a Normal distribution.* Suppose the $X_1, \ldots, X_N$ form an independent random sample from a Normal distribution with unknown mean μ and unknown variance $\sigma^2 > 0$. For any observed value of $\mathbf{x}$ where each x_i is a real number, the likelihood function is

$$f_N(\mathbf{x}|\mu, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2\right]. \tag{3}$$

and the MLE of μ is $\hat{\mu} = \overline{X}_N$, called the *sample mean*, and the MLE of σ^2 is $\widehat{\sigma^2} = \overline{X_N^2} - \overline{X_N}^2$, called the *sample variance.*

(iii) *Sampling from a uniform distribution.* If the $X_1, \ldots, X_N$ form an independent random sample from a uniform distribution on the interval $[0, \theta]$ with unknown parameter $\theta > 0$. For any observed value of $\mathbf{x}$ where each $x_i \ge 0$, the likelihood function is

$$f_N(\mathbf{x}|\theta) = \begin{cases} \dfrac{1}{\theta^N} & \text{for } 0 \le x_i \le \theta \ (i = 1, \ldots, N), \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

The MLE of θ is $\hat{\theta} = \max(X_1, \ldots, X_N)$.

Further details including various technical properties of MLE such as consistency may be found in the further reading section, (see for example DeGroot and Schervish, 2002, ch 6).

MALIK MAGDON-ISMAIL

Further Reading

DeGroot, M. H. and Schervish, M. J. (2002). *Probability and Statistics.* Addison–Wesley, Reading, Massachusetts.

Devore, J. L. (1999). *Probability and Statistics for Engineering and the Sciences.* Duxbury Press.

Larsen, R. and Marx, M. (1986). *An Introduction to Mathematical Statistics.* Prentice-Hall, New Jersey.

Ross, S. M. (2001). *A First Course in Probability.* Prentice Hall, New Jersey.

See also: Averages (computations), Bayes theorem and risk analysis, Method of generalized moments, Probability (fundamentals).