Lecture 19: Maximum Likelihood Estimator

11 April, 2018

Sunil Kumar Gauttam

Department of Mathematics, LNMIIT

In many cases, there will be an obvious or natural candidate for a point estimator of a particular parameter. For example, the sample mean is a natural candidate for a point estimator of the population mean. Sample variance is a natural candidate for a point estimator of the population variance and sample standard deviation is a natural candidate for a point estimator of the population standard deviation. In general, estimating a population parameter with its sample analogue is usually reasonable.

In more complicated models, ones that often arise in practice, we need a more methodical way of estimating parameters, which give us some reasonable candidates for consideration. We study the method of maximum likelihood which is by far, the most popular technique for deriving estimators.

Definition 19.1 *Let $X_1, \cdots, X_n$ be random sample from a population with pdf or pmf $f(x|\theta)$. The likelihood function is defined by*

$$L(\theta|x_1, x_2, \cdots, x_n) = f(x_1|\theta)f(x_2|\theta)\cdots f(x_n|\theta),$$

where $x_1, x_2, \cdots, x_n$ is the realized value of the random sample $X_1, \cdots, X_n$.

Note that likelihood function $L(\cdot|x_1, x_2, \cdots, x_n)$ is a function of the unknown population parameter θ given the sample values $x_1, x_2, \cdots, x_n$. So the domain of L is the set of all admissible values of population parameter.

Example 19.2 *Let $X_1, \cdots, X_4$ be a random sample from Bernoulli(p) population. For the sample value $(1, 0, 1, 1)$, the likelihood function of the random sample is*

$$L(p|1, 0, 1, 1) = p\ (1-p)\ p\ p = p^3(1-p),\ \textit{for } p \in [0, 1].$$

Similarly for the sample value $(1, 0, 0, 1)$, the likelihood function of the random sample is

$$L(p|1, 0, 0, 1) = p^2(1-p)^2,\ \textit{for } p \in [0, 1].$$

Example 19.3 *Let $X_1, \cdots, X_4$ be a random sample from Exponential(λ) population. For the sample value $(1.23, 3.32, 1.98, 2.12)$, the likelihood function of the random sample is*

$$L(\lambda|(1.23, 3.32, 1.98, 2.12)) = \lambda^4 e^{-\lambda(1.23+3.32+1.98+2.12)},\ \textit{for } \lambda \in (0, \infty).$$

Definition 19.4 *For each sample point* $\mathbf{x} = (x_1, x_2, \cdots, x_n)$*, let* $\hat{\theta}(\mathbf{x})$ *be a parameter value at which* $L(\theta|\mathbf{x})$ *attains its maximum as a function of* θ*, with* $\mathbf{x}$ *held fixed. A maximum likelihood estimator (MLE) of the parameter* θ *based on a random sample* $\mathbf{X} = (X_1, \cdots, X_n)$ *is* $\hat{\theta}(\mathbf{X})$.

Notice that, by its construction, the range of the MLE coincides with the range of the parameter. We also use the abbreviation MLE to stand for maximum likelihood estimate when we are talking of the realized value of the estimator.

Intuitively, the MLE is a reasonable choice for an estimator. The MLE is the parameter point for which the observed sample is most likely (In case $f(x|\theta)$ is pmf then it is very clear the case. In case of pdf, since pdf at a point represents the rate at which probability accumulate near the point, it again convey the similar message). If the likelihood function is differentiable (in θ), possible candidates for the MLE are the values of θ that solve

$$\frac{dL}{d\theta} = 0 \tag{19.1}$$

Note that the solutions to (19.1) are only possible candidates for the MLE since the first derivative being 0 is only a necessary condition for a maximum, not a sufficient condition. Furthermore, the zeros of the first derivative locate only extreme points in the interior of the domain of a function. If the extrema occur on the boundary the first derivative may not be 0. Thus, the boundary must be checked separately for extrema. Points at which the first derivatives are 0 may be local or global minima, local or global maxima, or inflection points. Our job is to find a global maximum.

Example 19.5 *Let* X_1, X_2, X_3, X_4 *be a random sample from the Bernoulli(p) population. For which value of p is the probability of the observed sample value* $(1, 0, 1, 1)$ *is the largest?*

Solution: The likelihood function of the given sample value is

$$L(p|1,0,1,1) = p\ (1-p)\ p\ p = p^3(1-p), \text{ for } p \in [0,1].$$

$$L'(p) = 3p^2 - 4p^3 = 0 \implies p = 0, \frac{3}{4}$$

So only critical point is $p = \frac{3}{4}$. Since L is a continuous function over a closed and bounded interval hence it attains its global maximum on the interval. At both boundary points 0 and 1 the value of function is zero and at $p = \frac{3}{4}$ it is positive therefore L attains it's supremum at $p = \frac{3}{4}$. ■

In most cases, especially when differentiation is to be used, it is easier to work with the natural logarithm of $L(\theta|\mathbf{x}), \log L(\theta|\mathbf{x})$ (known as the log likelihood function), than it is to work with $L(\theta|\mathbf{x})$ directly. This is possible because the log function is strictly increasing on $(0, \infty)$ and it's inverse exp function is also strictly increasing, which implies that the extrema of $L(\theta|\mathbf{x})$ and $\log L(\theta|\mathbf{x})$ coincide.

Example 19.6 *Let X_1, X_2, X_3, X_4 be a random sample from* $\exp(\lambda)$ *distribution. Find the MLE of λ for the sample value* $(1.23, 3.32, 1.98, 2.12)$.

Solution: The likelihood function of the given sample value is

$$L(\lambda|1.23, 3.32, 1.98, 2.12) = \lambda^4 e^{-(8.65)\lambda}, \text{ for } \lambda \in (0, \infty).$$

Taking logarithm of likelihood function we obtain

$$f(\lambda) = \log L = 4\log\lambda - (8.65)\lambda.$$

$$f'(\lambda) = \frac{4}{\lambda} - 8.65 = 0 \implies \lambda = \frac{4}{8.65} = 0.462427746$$
$$f''(\lambda) = \frac{-4}{\lambda^2} < 0, \ \forall\lambda \in (0, \infty)$$

There for $\lambda_0 = 0.462427746$ is a point of local maximum for f. But how do we ensure that it is the global maximum? We present two ways:

1. Since λ_0 is the only extreme point in the interior of the domain $(0, \infty)$ and it is the maximum. So there can not be maximum at boundary point 0, otherwise there has to be a local minimum between 0 and λ_0, which is a contradiction to the unique extreme point.

2. $f'(\lambda) > 0$ on the interval $(0, \lambda_0)$ and $f'(\lambda) > 0$ on the interval (λ_0, ∞). Therefore f is strictly increasing on $(0, \lambda_0)$ and strictly decreasing on (λ_0, ∞). So λ_0 is the point of global maximum.

Hence λ_0 the MLE of λ for the given sample value. ∎