3. Asymptotics for Maximum Likelihood Estimation

In this section, we show consistency and asymptotic normality of maximum likelihood estimator (MLE). The tests based on MLE, such as *likelihood ratio* (LR), *Wald* (W) and *Lagrange multiplier* (LM) *tests*, are also introduced. Throughout the section, we let $X_1, \ldots, X_n$ be i.i.d. random variables with the common underlying distribution, which is given by a parametric family of densities $\mathcal{P} = \{p(\cdot, \theta) | \theta \in \Theta\}$ with respect to some base measure μ. The true value of θ is denoted by θ_0, which yields the true density $p(\cdot, \theta_0)$. Let P_0 be the true underlying distribution with density $p(\cdot, \theta_0)$, and let expectation E_0 be an integral operator on $\mathcal{R}$ with respect to P_0. Denote by $\hat{\theta}_n$ the MLE for θ_0.

Denote by $p(x_i, \theta), \ell(x_i, \theta), s(x_i, \theta), h(x_i, \theta)$ $(h(\theta))$ and $\iota(\theta)$, density, loglikelihood, score, Hessian (expected Hessian), and (Fisher) information of X_i, $i = 1, \ldots, n$. Moreover, define

$$\bar{\ell}(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} \ell(x_i, \theta),\ \ \bar{s}(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} s(x_i, \theta),\ \ \bar{h}(x, \theta) = \frac{1}{n} \sum_{i=1}^{n} h(x_i, \theta)$$

with $x = (x_1, \ldots, x_n)'$.

1. Consistency of MLE

First we consider

$$\mathrm{E}_0 \ell(\cdot, \theta) = \int \ell(x, \theta) p(x, \theta_0)\, d\mu(x)$$

as a function of θ. The following lemma shows that it is maximized at θ_0.

Lemma 1 *We have*

$$\mathrm{E}_0 \ell(\cdot, \theta_0) \geq \mathrm{E}_0 \ell(\cdot, \theta)$$

for all $\theta \in \Theta$.

Proof It follows from the Jensen's inequality that

$$\mathrm{E}_0 \ell(\cdot, \theta) - \mathrm{E}_0 \ell(\cdot, \theta_0) \ \ = \ \ \mathrm{E}_0 \log \frac{p(\cdot, \theta)}{p(\cdot, \theta_0)}$$

$$\begin{aligned} &\leq \log \mathrm{E}_0 \frac{p(\cdot,\theta)}{p(\cdot,\theta_0)} \\ &= \log \int \frac{p(x,\theta)}{p(x,\theta_0)} p(x,\theta_0) d\mu(x) = 0 \end{aligned}$$

which completes the the proof. ■

Theorem 2 *Under suitable regularity conditions, we have*

$$\hat{\theta}_n \stackrel{a.s. \mathrm{or}\, p}{\longrightarrow} \theta_0$$

Proof Since X_i's are i.i.d., so are $\ell(X_i,\theta)$'s for any $\theta \in \Theta$. We may thus invoke a LLN for i.i.d. random variables to deduce that

$$\frac{1}{n}\sum_{i=1}^n \ell(X_i,\theta) \stackrel{a.s. \mathrm{or}\, p}{\longrightarrow} \mathrm{E}_0 \ell(\cdot,\theta)$$

for all $\theta \in \Theta$. Under certain regularity conditions, the convergence is uniform on Θ, i.e., *uniform law of large numbers* (ULLN) holds. Therefore, we get

$$\mathrm{argmax}_\theta \frac{1}{n}\sum_{i=1}^n \ell(X_i,\theta) \stackrel{a.s. \mathrm{or}\, p}{\longrightarrow} \mathrm{argmax}_\theta \mathrm{E}_0 \ell(\cdot,\theta)$$

The stated result can now be easily deduced. ■

2. Asymptotic Normality of MLE

Theorem 3 *Under suitable regularity conditions, we have*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_d \mathbf{N}(0, \iota(\theta_0)^{-1})$$

Proof Under certain regularity conditions, we may expect to have

(a) $\displaystyle \frac{1}{\sqrt{n}}\sum_{i=1}^n s(X_i,\theta_0) \to_d \mathbf{N}(0,\iota(\theta_0))$

(b) $\displaystyle \frac{1}{n}\sum_{i=1}^n h(X_i,\theta_0) \to_p \mathrm{E}_0 h(\cdot,\theta_0) = h(\theta_0) = -\iota(\theta_0)$

(c) $\bar{s}(x,\theta)$ is differentiable at θ_0 for all x,

(d) $\hat{\theta}_n = \theta_0 + O_p\left(\frac{1}{\sqrt{n}}\right)$

It follows from

$$\bar{s}(x, \theta) = \bar{s}(x, \theta_0) + \bar{h}(x, \theta_0)(\theta - \theta_0) + o\left(\|\theta - \theta_0\|\right)$$

that

$$\frac{1}{n}\sum_{i=1}^{n} s(X_i, \hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} s(X_i, \theta_0) + \left(\frac{1}{n}\sum_{i=1}^{n} h(X_i, \theta_0)\right)(\hat{\theta}_n - \theta_0) + o_p\left(\frac{1}{\sqrt{n}}\right)$$

Therefore,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, \hat{\theta}_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, \theta_0) + \left(\frac{1}{n}\sum_{i=1}^{n} h(X_i, \theta_0)\right)\sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1)$$

from which we have

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_n - \theta_0) &= -\left(\frac{1}{n}\sum_{i=1}^{n} h(X_i, \theta_0)\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, \theta_0) + o_p(1) \\
&= -h(\theta_0)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} s(X_i, \theta_0) + o_p(1) \\
&\to_d \mathbf{N}(0, \iota(\theta_0)^{-1})
\end{aligned}$$

by CMT. ■

Remark The MLE achieves the Cramer-Rao lower bound asymptotically, and is therefore often said to be *efficient*. The bound, however, is a finite sample result, and not necessarily hold in asymptotics. We can indeed construct an estimator which has asymptotic variance smaller than the bound. Such an estimator is called *hyper-efficient*.

3. Asymptotic Tests based on MLE

We now introduce the likelihood ratio (LR), Lagrange multiplier (LM), and Wald (W) tests, which are based on MLE. It will be shown that their limiting distributions are all chi-squre, and that they are asymptotically equivalent. For simplicity, we consider the hypothesis

$$H_0 : \theta = \theta_0$$

which is to be tested against $\theta \neq \theta_0$. Assume θ is m-dimensional.

Define

Definition 1

$$\begin{aligned}
\mathrm{LR} &= 2\left(\sum_{i=1}^{n} \ell(x_i, \hat{\theta}_n) - \sum_{i=1}^{n} \ell(x_i, \theta_0)\right) \\
\mathrm{W} &= \sqrt{n}(\hat{\theta}_n - \theta_0)' \iota(\hat{\theta}_n) \sqrt{n}(\hat{\theta}_n - \theta_0) \\
\mathrm{LM} &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(x_i, \theta_0)\right)' \iota(\theta_0)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(x_i, \theta_0)\right)
\end{aligned}$$

Remarks

(a) For the LR statistic, notice that the likelihood ratio is given by

$$\lambda(x_1, \ldots, x_n) = \frac{\max_{\theta \in \Theta} p(x_1, \ldots, x_n, \theta)}{p(x_1, \ldots, x_n, \theta_0)}$$

and, therefore, we may write $\mathrm{LR} = 2 \log \lambda(x_1, \ldots, x_n)$.

(b) The LR, W and LM statistics are based, respectively, on the ratio of restricted and unrestricted maximum likelihoods, the difference between the estimated and hypothesized values of the parameter, and the first derivative of the likelihood function at the hypothesized value of the parameter. If the hypothesized value is equal to the true value, all three must be small.

We have

Theorem 4 *Under suitable regularity conditions,*

$$\mathrm{LR}, \mathrm{W}, \mathrm{LM} \quad \to_d \quad \chi^2_m$$

Proof Assume the conditions introduced in Theorem 3 hold. Since

$$\begin{aligned}
\bar{\ell}(x, \theta) &= \bar{\ell}(x, \theta_0) + \bar{s}(x, \theta_0)'(\theta - \theta_0) + \frac{1}{2}(\theta - \theta_0)' \bar{h}(x, \theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|^2) \\
\bar{s}(x, \theta_0) &= \bar{s}(x, \theta) - \bar{h}(x, \theta_0)(\theta - \theta_0) + o(\|\theta - \theta_0\|)
\end{aligned}$$

we have

$$\bar{\ell}(x,\theta) = \bar{\ell}(x,\theta_0) + \bar{s}(x,\theta)'(\theta-\theta_0) - \frac{1}{2}(\theta-\theta_0)'\bar{h}(x,\theta_0)(\theta-\theta_0) + o(\|\theta-\theta_0\|^2)$$

It follows that

$$\begin{aligned}
&\sum_{i=1}^n \ell(X_i,\hat{\theta}_n) - \sum_{i=1}^n \ell(X_i,\theta_0) \\
&= -\frac{1}{2}\sqrt{n}(\hat{\theta}_n-\theta_0)'\left(\frac{1}{n}\sum_{i=1}^n h(X_i,\theta_0)\right)\sqrt{n}(\hat{\theta}_n-\theta_0) + o_p(1) \\
&= \frac{1}{2}\sqrt{n}(\hat{\theta}_n-\theta_0)'\iota(\theta_0)\sqrt{n}(\hat{\theta}_n-\theta_0) + o_p(1)
\end{aligned}$$

which yields the result for the LR statistic.

To get the result for the Wald statistic, assume that $\iota(\theta)$ is continuous at θ_0 so that $\iota(\hat{\theta}_n) = \iota(\theta_0) + o_p(1)$, and notice that

$$\sqrt{n}(\hat{\theta}_n-\theta_0) \;\to_d\; \mathbf{N}(0,\iota(\theta_0)^{-1})$$

For the LM test (or score test), we consider the limiting distribution of the score, i.e.,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n s(X_i,\theta_0) \;\to_d\; \mathbf{N}(0,\iota(\theta_0))$$

from which the stated result follows directly. ■

Remark Different versions of W are possible with the replacement of $\iota(\hat{\theta}_n)$ by any one of the following:

(a) $\frac{1}{n}\sum_{i=1}^n s(X_i,\hat{\theta}_n)s(X_i,\hat{\theta}_n)'$

(b) $-h(\hat{\theta}_n)$

(c) $-\frac{1}{n}\sum_{i=1}^n h(X_i,\hat{\theta}_n)$

We have

Corollary 5 *The tests based on the statistics* LR, W *and* LM *are asymptotically equivalent.*

Proof Observe that

$$\begin{aligned}
&2\left(\sum_{i=1}^{n} \ell(X_i, \hat{\theta}_n) - \sum_{i=1}^{n} \ell(X_i, \theta_0)\right) \\
&\quad = \sqrt{n}(\hat{\theta}_n - \theta_0)' \iota(\hat{\theta}_n) \sqrt{n}(\hat{\theta}_n - \theta_0) + o_p(1) \\
&\quad = \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(X_i, \theta_0)\right)' \iota(\theta_0)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} s(X_i, \theta_0)\right) + o_p(1)
\end{aligned}$$

from which the stated result is immediate. ■