Large Number theorey of MLE

Asymptotic Normality of MLE

In essense, we approximate sampling distribution of MLE estimator by using limiting argument as sample size increases.

Definition. ***Asymptotically Normal*** *Let $X_1, \cdots, X_n \sim f_\theta$ We say that $\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n)$ is asymptotically normal with mean θ and variance $\dfrac{\sigma^2}{n}$ if for all $z \in \mathbb{R}$*

$$F_{Z_n}(z) \overset{n\to\infty}{\to} \Phi(z)$$

where F_{Z_n} is the cdf of $Z_n = \dfrac{\hat{\theta}_n - \theta}{\sigma/\sqrt{n}}$. Equivalent to convergence in distribution

$$\sqrt{n}(\hat{\theta} - \theta) \overset{D}{\to} \mathcal{N}(0, \sigma^2)$$

Definition. *Let $X_1, \cdots, X_n \sim f_\theta$ with $l(\theta) = \log f(x_1, \cdots, c_n|\theta)$*

1. *the **Score** with respect to θ is*
$$u(\theta) := l'(\theta)$$
Under regularity conditions, the first moment $\mathbb{E}[u(\theta)] = \displaystyle\int \frac{\partial \log f(x|\theta)}{\partial \theta} f(x|\theta)dx = 0$; the second moment is the Fisher information $\mathbb{E}[u^2(\theta)] = \mathcal{I}(\theta)$

2. *the **Fisher Information** for θ is*
$$\mathcal{I}(\theta) := -\mathbb{E}[l''(\theta)]$$
where $l(\theta) = \displaystyle\sum_{i=1}^{n} \log f(x_i|\theta)$

3. *Fisher information of θ based on a single observation*
$$\mathcal{I}^* := -\mathbb{E}[\frac{\partial^2 \log f(x|\theta)}{\partial \theta^2}]$$
We notice that $\mathcal{I}(\theta) = n\mathcal{I}^*(\theta)$

Proposition. *Under some regularity conditions*

1. $\mathcal{I}(\theta) = \mathbb{E}[u^2(\theta)]$

2. $\dfrac{u(\theta)}{\sqrt{n}} \overset{D}{\to} \mathcal{N}(0, \mathcal{I}^*(\theta))$ *(asymptotically normal)*

3. $-\frac{1}{n}\frac{\partial^2 l(\theta)}{\partial\theta^2} \xrightarrow{p} \mathcal{I}^*(\theta)$

Theorem. ***Slutsky's Theorem*** *Let $\{X_n\}$ and $\{Y_n\}$ be two sequences of RV such that $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{P} c$ for some $c \in \mathbb{R}$ then for any continuous function g we have*

$$g(X_n, Y_n) \xrightarrow{D} g(X, c)$$

1. $X_n + Y_n \xrightarrow{D} X + c$
2. $X_n Y_n \xrightarrow{D} cX$
3. $X_n / Y_n \xrightarrow{D} X/c$

In particular if $X_n \xrightarrow{D} \mathcal{N}(0, \sigma^2)$ and $Y_n \xrightarrow{P} c$ we have

$$X_n Y_n \xrightarrow{D} \mathcal{N}(0, c^2\sigma^2)$$

Theorem. ***Asymptotic Normality of MLEs*** *Let $X_1, \cdots, X_n$ be random sample from f_θ and let $\hat{\theta}_n$ denote the maximum likelihood estimator of θ. Under some regularity conditions, $\hat{\theta}_n$ is asymptotically normal with mean θ and variance $\mathcal{I}^{-1}(\theta)$. in other words,*

$$\sqrt{n}(\hat{\theta}_{MLE} - \theta) \xrightarrow{D} \mathcal{N}(0, \frac{1}{\mathcal{I}^*(\theta)})$$

$$\hat{\theta}_{MLE} \sim AN(\theta, \mathcal{I}^{-1}(\theta))$$

Theorem. ***Invariance of MLE and transformation*** *Let $X_1, \cdots, X_n$ be a sample from f_θ and let $\eta = g(\theta)$ for some transform g. Then*

1. $\hat{\eta}_{MLE} = g(\hat{\theta}_{MLE})$
2. *If g is differentiable then $\hat{\eta}_{MLE} \sim AN(\eta, [g'(\theta)]^2\mathcal{I}^{-1}(\theta))$*

Remark. We can derive MLE for function of θ_{MLE} with this theorem. For example, we can find MLE for log-odds

$$\psi = \log\frac{p}{1-p}$$

where $X_1, \cdots, X_n \overset{i.i.d}{\sim} Binom(1, p)$ by first computing MLE for Bernoulli distribution, i.e. $\hat{p}_{MLE} = \overline{X}$ with Fisher information $\mathcal{I}(p) = \frac{n}{p(1-p)}$ then,

$$\hat{\psi}_{MLE} = \log\frac{\overline{X}}{1-\overline{X}}$$

To compute asymptotic sampling distribution, let $g(p) = \log \frac{p}{1-p}$. we calculate the asymptotic variance given by

$$[g'(p)]^2\mathcal{I}^{-1}(p) = (\frac{1-p}{p}\frac{1-p+p}{(1-p)^2})^2\frac{p(1-p)}{n} = \frac{1}{np(1-p)}$$

Since g is differentiable

$$\log \frac{\overline{X}}{1-\overline{X}} \sim AN(\log \frac{p}{1-p}, \frac{1}{np(1-p)})$$

Theorem. ***Consistency in MLE*** *If the regularity condition for asymptotic normality is satisfied, the MLE is consistent (in probability)*

Proof. Given MLE is asymptotically normal $(\hat{\theta}_{MLE} - \theta) \sim AN(0, \mathcal{I}^{-1}(\theta))$, then

$$F_{Z_n}(z) \sim \Phi(z) \quad \text{for} \quad Z_n = \sqrt{n\mathcal{I}^*(\theta)}(\hat{\theta}_{MLE} - \theta)$$

Let $\epsilon > 0$ be given, then

$$P(|\hat{\theta}_{MLE} - \theta| \leq \epsilon) = P(|\sqrt{n\mathcal{I}^*(\theta)}(\hat{\theta}_{MLE} - \theta)| \leq \epsilon\sqrt{n\mathcal{I}(\theta)}) = 2\Phi(\sqrt{n\mathcal{I}^*(\theta)}\epsilon) - 1 + \delta_n$$

where deviation from normality $\delta_n \overset{n\to\infty}{\to} 0$. Hence

$$P(|\hat{\theta}_{MLE} - \theta| \leq \epsilon) = 2 - 1 + 0 = 1$$

$\hat{\theta}_{MLE} \overset{p}{\to} \theta$ so then MLE is a consistent estimator

Remark. The proof is basically proving convergence in probability given convergence in distribution (although this is not true in all cases)

□

Definition. ***Standard Error*** *The standard deviation of an estimator $\hat{\theta}$ of a parameter θ is called the standard error of $\hat{\theta}$; In other word, it is the standard deviation of the sampling distribution of a statistic.*

Theorem. ***The Plug-in Principle*** *Let $\hat{\theta}$ be the MLE of θ satisfying the regularity condition for asymptotic normality, then by Slutsky's Theorem,*

$$\hat{\theta} \sim AN(\theta, \mathcal{I}^{-1}(\theta))$$

*Then for any **consistent estimator** $\hat{\theta}$ of θ*

$$\hat{\theta} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}))$$

In particular,

$$\hat{\theta} \sim AN(\theta, \mathcal{I}^{-1}(\hat{\theta}_{MLE})$$

*Then the **estimated standard error** is*

$$\hat{\sigma}_{\hat{\theta}_{MLE}} = \mathcal{I}^{-1/2}(\hat{\theta}_{MLE})$$

Remark. We are motivated to do this because $\mathcal{I}^{-1}(\theta)$ is a function of population parameter which we do not know. Instead we approximate it with $\hat{\theta}_{MLE}$ instead. And this theorem ensures that if we do so, asymptotic normality of MLE is preserved

Definition. ***MLE estimate of multinomial cell probability***