MA 3140: Statistical Inference

Dr. Sameen Naqvi
Department of Mathematics, IIT Hyderabad
Email id: sameen@math.iith.ac.in

Properties of MLE

- **Non-uniqueness of MLE**

Example: Let $X_1, \ldots, X_n \sim U[\theta - a, \theta + a]$, $\theta \in \mathbb{R}$, $a > 0$ is a known constant.

The likelihood function is

$$L(\theta, \boldsymbol{x}) = \begin{cases} \frac{1}{(2a)^n}, & \theta - a \le x_{(1)} \le \ldots \le x_{(n)} \le \theta + a \\ 0, & \text{o.w.} \end{cases}$$

L is maximum when $\theta - a \le x_{(1)}$ (or, $\theta \le x_{(1)} + a$) and $\theta + a \ge x_{(n)}$ (or, $\theta \ge x_{(n)} - a$).

So, any value between $x_{(n)} - a$ and $x_{(1)} + a$ is a MLE of θ.

- **MLE need not be in a nice analytic form**

Example: Let $X_1, \ldots, X_n \sim N(\theta, \theta^2)$, $\theta > 0$.

The likelihood function is

$$L(\theta, \boldsymbol{x}) = \begin{cases} \frac{1}{(\theta\sqrt{2\pi})^n} e^{-\frac{1}{2\theta^2}\sum(x_i-\theta)^2}, & x_i \in \mathbb{R}, \theta > 0 \\ 0, & \text{o.w.} \end{cases}$$

$$l(\theta) = \log L(\theta, \boldsymbol{x}) = -n \log \theta - \frac{n}{2} \log 2\pi - \frac{\sum(x_i - \theta)^2}{2\theta^2}$$

$$\frac{dl}{d\theta} = -\frac{n}{\theta} + \frac{\sum(x_i - \theta)}{2\theta^2} + \frac{\sum(x_i - \theta)^2}{\theta^3}$$

$$= \frac{1}{\theta^3}\left[\sum x_i^2 - n\theta\overline{x} - n\theta^2\right] = -\frac{n}{\theta^3}\left[\theta^2 + \theta\overline{x} - \alpha\right]$$

Properties of MLE

$$\frac{dl}{d\theta} = 0 \Longrightarrow \theta = \frac{-\bar{x} \pm \sqrt{\bar{x}^2 + 4\alpha}}{2}$$

So, for $\hat{\theta}_1 = \frac{-\bar{x}-\sqrt{\bar{x}^2+4\alpha}}{2}$ and $\hat{\theta}_2 = \frac{-\bar{x}+\sqrt{\bar{x}^2+4\alpha}}{2}$,

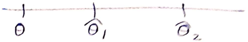
$$\begin{aligned} \frac{dl}{d\theta} = -\frac{n}{\theta^3}(\theta - \hat{\theta}_1)(\theta - \hat{\theta}_2) &< 0 \quad \text{if} \quad \theta < \hat{\theta}_1 \quad \text{or} \quad \theta > \hat{\theta}_2 \\ &> 0 \quad \text{if} \quad \hat{\theta}_1 < \theta < \hat{\theta}_2 \end{aligned}$$

Thus, we can see that $l(\theta)$ is maximized at $\theta = \hat{\theta}_2$, i.e.,

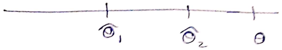
$$\hat{\theta}_{MLE} = \frac{-\overline{X} + \sqrt{\overline{X}^2 + 4\alpha}}{2}$$

Properties of MLE

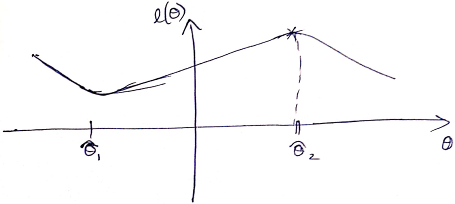
If $\theta < \hat{\theta}_1$,



If $\theta > \hat{\theta}_2$



If $\hat{\theta}_1 < \theta < \hat{\theta}_2$



Properties of MLE

- **MLE may not be in a closed form**

Example: Let $X_1, \ldots, X_n \sim Gamma(r, \lambda)$, where λ is known. Find MLE of r.

The likelihood function is

$$L(r, \boldsymbol{x}) = \prod_{i=1}^{n} \left[\frac{\lambda^r}{\Gamma r} e^{-\lambda x_i} x_i^{r-1} \right] = \frac{\lambda^{nr}}{(\Gamma r)^n} e^{-\lambda \sum x_i} (\prod x_i)^{r-1}$$

$$l(r) = \log L = nr \log \lambda - n \log \Gamma r - \lambda \sum x_i + (r-1) \log \prod x_i$$

$$= -n \log \Gamma r - \sum x_i + (r-1) \log \prod x_i \qquad (\lambda = 1)$$

$$\frac{dl}{dr} = -\frac{n}{\Gamma r} \Gamma'(r) + \sum \log x_i = 0$$

Properties of MLE

- **MLE may be absurd or may not exist**

Example

Let $X_1, \ldots, X_n \sim Bin(1, p)$, where $0 < p < 1$ is unknown.

Answer: If $(0, \ldots, 0)$ $((1, \ldots, 1))$ is observed, $\overline{X} = 0$ $(\overline{X} = 1)$ is the MLE, which is not admissible value of p. Hence, an MLE does not exist.

Properties of MLE

- **Invariance of MLE**

Theorem: Let $\hat{\theta}_{ML}$ denote the MLE of $\theta \in \Theta$. Consider $\phi = g(\theta)$ where $\phi \in \Phi = g(\Theta) = \{g(\theta) : \theta \in \Theta\}$ and g is one-to-one function. Then

$$\hat{\phi}_{ML} = g(\hat{\theta}_{ML}).$$

Proof: Let $L(\theta)$ and $L^*(\phi)$ denote the likelihood functions corresponding to θ and ϕ, respectively.
We have $L(\hat{\theta}_{ML}) \geq L(\theta), \forall\ \theta \in \Theta$. So for $\hat{\phi}_{ML} = g(\hat{\theta}_{ML})$, we get

$$\begin{aligned} L^*(\hat{\phi}_{ML}) = L^*(g(\hat{\theta}_{ML})) = L(g^{-1}g(\hat{\theta}_{ML})) \\ = L(\hat{\theta}_{ML}) \geq L(\theta) = L^*(\phi), \forall\ \phi \in \Phi \end{aligned}$$

Thus, $\hat{\phi}_{ML}$ is the MLE of ϕ.

Properties of MLE

- **Invariance of MLE**

Theorem (Zahna, 1967): If $\hat{\theta}$ is the MLE of θ, then for any function $g(\theta)$, the MLE of $g(\theta)$ is $g(\hat{\theta})$.

Properties of MLE

- **Examples**

(i) Let $X_1, \ldots, X_n$ be a random sample from a Poisson(λ) distribution, where $\lambda \geq 0$ is unknown. Find the MLE of $g(\lambda) = P(X_1 = 0) = e^{-\lambda}$.

Answer: Since MLE of λ is $\overline{X}$, MLE of $e^{-\lambda}$ is $e^{-\overline{X}}$.

(ii) Let $X \sim Bin(1, p)$, where $0 \leq p \leq 1$ is unknown. Find the MLE of $g(p) = p(1-p)$.

Answer: Since MLE of p is $\overline{X}$, MLE of $p(1-p)$ is $\overline{X}(1-\overline{X})$.

Properties of MME

(i) If $\hat{\boldsymbol{\theta}}$ is the MME of $\boldsymbol{\theta}$, then $g(\hat{\boldsymbol{\theta}})$ is the MME of $g(\boldsymbol{\theta})$.

(ii) The method of moments is not applicable when population moments do not exist (for instance, Cauchy distribution).

(iii) Modified MME can be obtained when first population moment does not exist (for instance, in $U(-\frac{\theta}{2}, \frac{\theta}{2})$, $E(X) = 0$ but $E(X^2) = \frac{\theta^2}{12}$).

(iv) M.M.E. may not exist when the underlying equations do not have a solution. Also, the M.M.E. may not be unique as the underlying equations may have more than one solution.

How to be sure that this estimator is good? Is it close to the actual true parameter?

Desirable criteria for estimators

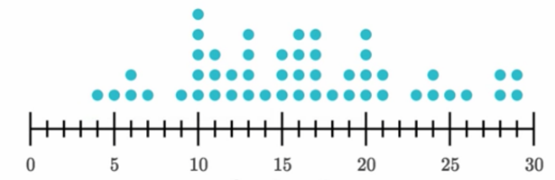
- Unbiasedness
- Consistency

Unbiasedness

- Let $\underline{X} = (X_1, \ldots, X_n)$ be a random sample from a population with the probability distribution $F(x, \theta)$, $\theta \in \Theta$. An estimator $T(\underline{X})$ is said to be **unbiased** for estimating $g(\theta)$, if

$$E_\theta T(\underline{X}) = g(\theta), \quad \forall\ \theta \in \Theta.$$

In other words, unbiasedness means that the sampling distribution of the estimator is centered around the parameter.

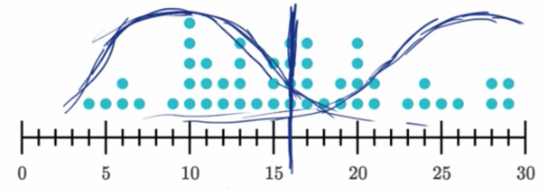


Unbiasedness

- If $E_\theta T(\underline{X}) = g(\theta) + b(\theta)$, then $b(\theta)$ is called the **bias** of T.

If $b(\theta) > 0$, $\forall\ \theta$ then T is said to **over-estimate** $g(\theta)$.

If $b(\theta) < 0$, $\forall\ \theta$ then T is said to **under-estimate** $g(\theta)$.



Example 1: Binomial Distribution

Let $X \sim B(n, p)$, n is known and $0 \leq p \leq 1$.

We know that MLE of p is $\frac{X}{n}$.

(i) **Is it also an UE of p?**

$$E\left(\frac{X}{n}\right) = p, \forall\ p.$$

Thus, $\frac{X}{n}$ is an unbiased estimator of p.

Example 1: Binomial Distribution cont'd

(ii) **Is $\frac{X+1}{n+2}$ an UE of p?:**

$$
\begin{aligned}
E\left(\frac{X+1}{n+2}\right) &= \frac{1}{n+2}E(X+1) \\
&= \frac{1}{n+2}[E(X)+1] \\
&= \frac{1}{n+2}[np+1] \\
&\neq p
\end{aligned}
$$

Thus, $\frac{X+1}{n+2}$ is an not an unbiased estimator of p.

Example 1: Binomial Distribution cont'd

(iii) **What will be an UE of p^2?**

We know that $E(X) = np$ and $E(X(X-1)) = n(n-1)p^2$.

$$
\begin{aligned}
E\ X(X-1) &= n(n-1)p^2 \\
\Longrightarrow E\left\{\frac{X(X-1)}{n(n-1)}\right\} &= p^2
\end{aligned}
$$

Thus, $\frac{X(X-1)}{n(n-1)}$ is unbiased for p^2.

Example 1: Binomial Distribution cont'd

(iv) **What will be an UE of Variance?**

Here, $V(X) = np(1-p) = n(p-p^2)$.

Consider $\frac{X}{n} - \frac{X(X-1)}{n(n-1)}$.

$$E\left[\frac{X}{n} - \frac{X(X-1)}{n(n-1)}\right] = E\left[\frac{X}{n}\right] - E\left[\frac{X(X-1)}{n(n-1)}\right] = p - p^2$$

$$\implies n\ E\left[\frac{X}{n} - \frac{X(X-1)}{n(n-1)}\right] = n(p-p^2)$$

Thus, $n\left[\frac{X}{n} - \frac{X(X-1)}{n(n-1)}\right] = \frac{X(n-X)}{n-1}$ is an UE of $V(X)$.

Example 2: Poisson Distribution

Let $X_1, \ldots, X_n \sim P(\lambda)$, $\lambda > 0$.

We know that MoM of λ is $\frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$.

(i) **Is it also an UE of λ?**

$$E(\overline{X}) = \frac{1}{n} E\Big(\sum_{i=1}^{n} X_i\Big) = \lambda.$$

Thus, $T_1(\underline{X}) = \overline{X}$ is an unbiased estimator of λ.

(ii) **Check whether the following estimators are unbiased for λ.**

$$T_2(\underline{X}) = X_i, \;\; i = 1, 2, \ldots, n$$

$$T_3(\underline{X}) = \frac{X_1 + 2X_2}{3}$$

$$T_4(\underline{X}) = \frac{1}{n-1} \sum (X_i - \overline{X})^2$$

Remark

- If $E(X)$ exists, then the sample mean $\overline{X}$ is an UE of the population mean μ.
- Let $E(X^2)$ exist, i.e., $V(X) = \sigma^2$ exists. Then, sample variance
$$S^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$$
is an UE of population variance σ^2.

Remark cont'd

- **What will be an UE of μ^2?**

  We know $E(\overline{X}) = \mu$ and $E(S^2) = \sigma^2$.

$$
\begin{aligned}
Var(\overline{X}) &= E(\overline{X}^2) - (E(\overline{X}))^2 \\
\implies \quad E(\overline{X}^2) &= \frac{\sigma^2}{n} + \mu^2 \\
\implies E\left(\overline{X}^2 - \frac{S^2}{n}\right) &= \mu^2
\end{aligned}
$$

So, $\overline{X}^2 - \frac{S^2}{n}$ is unbiased for μ^2.

Example 3: Exponential Distribution

Let $X_1, \ldots, X_n$ be a random sample from an exponential distribution with parameter λ.

We know that MoM of λ is $\frac{1}{\overline{X}}$.

(i) **What is an UE of $\frac{1}{\lambda}$?**

$$\because E(X_i) = \frac{1}{\lambda} \Longrightarrow E(\overline{X}) = \frac{1}{\lambda}$$

Thus, $\overline{X}$ is an unbiased estimator of $\frac{1}{\lambda}$.

Example 3: Exponential Distribution cont'd

Let $X_1, \ldots, X_n$ *be a random sample from an exponential distribution with mean* $\frac{1}{\lambda}$*.* The p.d.f. is

$$f_{X_i}(x) = \lambda e^{-\lambda x}, \; x > 0, \lambda > 0.$$

- MLE of λ is $\frac{1}{\overline{X}}$.
- $\overline{X}$ is an UE of $\frac{1}{\lambda}$.

Let $X_1, \ldots, X_n$ *be a random sample from an exponential distribution with mean* λ*.* The p.d.f. is

$$f_{X_i}(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}}, \; x > 0, \lambda > 0.$$

- MLE of λ is $\overline{X}$.
- $\overline{X}$ is an UE of λ.

Example 4: Revisiting Poisson

Let $X \sim P(\lambda)$ with p.m.f.

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x = 0, 1, \ldots$$

Suppose we want to estimate the probability of no occurrence, i.e., $P(X = 0) = e^{-\lambda}$.
Define an indicator function

$$I(X) = \begin{cases} 1, & \text{if } X = 0 \\ 0, & \text{if } X \neq 0. \end{cases}$$

Then

$$E[I(X)] = 1.P(X = 0) + 0.P(X \neq 0) = P(X = 0) = e^{-\lambda}.$$

So, $I(X)$ is an unbiased estimator of $e^{-\lambda}$.

Thanks for your patience!