MTH211A: Theory of Statistics

Module 4: Methods of Point Estimation

In the previous modules, we have talked about properties of good estimators. Next, we will discuss some common methods of finding point estimators of θ. The two most popular frequentist methods of finding estimators are:

(A) method of moments (MoM), and

(B) maximum likelihood (ML) estimation.

We will now discuss these methods in details.

1 Method of Moments (MoM)

One of the simplest and oldest methods of finding estimators is the method of moments or substitution principle.

Let $X_1, \ldots, X_n$ be a random sample from a population with distribution function $\{F_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$. Method of moments estimators are found by equating the first k sample moments to the corresponding k population moments, and solving the resulting system of equations.

Let $\boldsymbol{\theta}$ be a k-dimensional parameter. Then usually we require a system of k-equations to get an estimate of $\boldsymbol{\theta}$. Let M'_r be the r-th sample raw moment, and $\mu'_r = E\left(X_1^r\right)$ be the r-th raw moment of $F_{\boldsymbol{\theta}}$, $r = 1, \ldots, k$. Of course, for each r, μ'_r will be a function of $\boldsymbol{\theta}$. Then we obtain estimates of $\boldsymbol{\theta}$ by solving the following equations:

$$M'_r = \mu'_r, \qquad r = 1, \ldots, k.$$

Example 1. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Normal}(\mu, \sigma^2)$. The MoM estimators of μ and σ^2 are $\bar{X}$ and S^2 respectively.

Remark 1. *From Weak (Strong) Law of Large Numbers, $M'_r \xrightarrow{P\ (a.s)} \mu'_r$. Therefore, if one is interested in the population moments, then MoM provides consistent (strongly consistent) estimators.*

Remark 2. *However, MoM may lead to estimators having sub-optimal sampling properties, and may lead to absurd estimators in some cases.*

Example 2. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Uniform}(\alpha, \beta)$. The MoM estimators of α and β are $T_1(\mathbf{X})$ and $T_2(\mathbf{X})$, respectively, where

$$T_1(\mathbf{X}) = \bar{X} - \sqrt{\frac{3\sum_{i=1}^n (X_i - \bar{X})^2}{n}}, \quad \text{and} \quad T_2(\mathbf{X}) = \bar{X} + \sqrt{\frac{3\sum_{i=1}^n (X_i - \bar{X})^2}{n}}.$$

Observe that, none of the estimates are based on minimal sufficient statistic.

2 Maximum Likelihood (ML) Estimation

Definition 1 (Likelihood Function)**.** *Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a random sample from a population with distribution function $\{F_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$. Suppose the distribution $\{F_{\boldsymbol{\theta}}; \boldsymbol{\theta} \in \Theta\}$ is characterized by a pdf (or, pmf) $f_{\mathbf{X}}(\cdot; \theta)$. Further, suppose $\mathbf{x}$ is a realization of $\mathbf{X}$. Then the function of θ defined as $L(\theta \mid \mathbf{x}) = f(\mathbf{x}; \theta)$ is called the likelihood function.*

Definition 2 (Maximum Likelihood Estimate, MLE)**.** *Given a realization $\mathbf{x}$, let $\widehat{\theta}$ be the value in Θ that maximizes the likelihood function $L(\theta \mid \mathbf{x})$ with respect to θ, then $\widehat{\theta}$ is called the MLE of the parameter θ.*

Note that, the maximizer $\widehat{\theta}$ is nothing but a function of the realization $\mathbf{x}$. Thus we can treat the maximizer of the likelihood function as a statistic or estimator of θ. This estimator is called Maximum Likelihood (ML) estimator. Notationally, we write $\widehat{\theta} = \widehat{\theta}_{\text{ML}}(\mathbf{X})$.

Example 3. Suppose there are n tosses of a coin, and we do not know the value of n, or the probability of head (p). However, we know that n is between 3 to 5 and one of the sides of the coin is twice as heavy as the other (i.e., either $p = 2(1-p)$ or $(1-p) = 2p$). Then what is the MLE of $\theta = (n, p)$?

Remark 3. *If the likelihood function is differentiable with respect to θ, the one may take the differentiation approach for finding the MLE. In case of a several value function $L(\boldsymbol{\theta})$, if the function is twice continuously differentiable with respect to each θ_j, then a critical point of $L(\boldsymbol{\theta} \mid \mathbf{x})$ can be obtained by equating $\frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta} \mid \mathbf{x}) = \mathbf{0}$. Then to verify, if the critical point is a maximizer, one can check if the Hessian matrix $\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L(\boldsymbol{\theta} \mid \mathbf{x})$ is negative definite at the critical point.*

Remark 4. *Often it is convenient to work with the log likelihood function, instead of the likelihood function. As logarithm is a monotone function, the maximizer of likelihood and the log likelihood are the same. The log likelihood is generally denoted by $l(\boldsymbol{\theta}; \mathbf{x})$.*

Example 4. Let $X_1, \ldots, X_n$ be a random sample from `normal`(μ, σ^2). Then the MLE of μ and σ^2 are $\bar{X}_n$ and S_n^2, respectively.

Example 5. Let $X_1, \ldots, X_n$ be a random sample from `uniform`(α, β). Then the MLE of α and β are $X_{(1)}$ and $X_{(n)}$, respectively.

In the above two examples, we have seen that MLE is a function of a sufficient statistic. This phenomenon is in general true, as stated by the following theorem.

Theorem 1 (Properties of MLE: 1)**.** *Let $X_1, \ldots, X_n$ be a random sample from some distribution with pdf (or pmf) $f_\theta; \theta \in \Theta$, and let $T(\mathbf{X})$ be a sufficient statistic for θ. Then an MLE, if exists and unique, is a function of T. If MLE exists, but is not unique then, one can find an MLE which is a function of T.* [Proof]

Remark 5. *Maximum likelihood estimate may not exist. See the following example.*

Example 6. Let X_1, X_2 be a random sample from `Bernoulli`(θ), $\theta \in (0, 1)$. Suppose the realization $(0, 0)$ is observed. Then the MLE does not exist.

Remark 6. *Even if MLE exists, it may not be unique. See the following example.*

Example 7. Let $X_1, \ldots, X_n$ be a random sample from `Double Exponential`(θ, σ) distribution, $\theta \in \mathbb{R}$. Then the $\widehat{\theta}_{\text{ML}}$ is the median $X_1, \ldots, X_n$, which is not unique.

Remark 7. *The method of maximum likelihood estimation may produce an absurd estimator.*

In spite of all the above drawbacks, MLE is by far the most popular and reasonable frequentist method of estimation. The reason is that MLE possesses a list of desirable properties. We will discuss some of them below.

Theorem 2 (Properties of MLE: 2)**.** *Suppose the regularity conditions of CRLB (see Theorem 5 of Module 3) are satisfied, the log-likelihood is twice differentiable, and there exists an unbiased estimator $\widehat{\theta}^\star$ of θ, the variance of which attains the CRLB, then $\widehat{\theta}^\star = \widehat{\theta}_{\text{ML}}(\mathbf{X})$.* [Proof]

Corollary 1. *Theorem 2 implies that if the CRLB is attained by any estimator, then it must be an MLE. However, the converse is not true, i.e., the variance of an MLE may not attain the CRLB.*

Theorem 3 (Properties of MLE: 3, Invariance Property)**.** *Let $\{f_{\mathbf{X}}(\cdot; \theta) : \theta \in \Theta\}$ be a family of PDFs (PMFs), and let $L(\theta \mid \mathbf{x})$ be the likelihood function. Suppose $\Theta \in \mathbb{R}^k, k \geq 1$. Let $h : \Theta \to \Lambda$ be a mapping of Θ onto Λ, where Λ is a subset of $\mathbb{R}^p$ $(1 \leq p \leq k)$. If $\widehat{\theta}_{\text{ML}}(\mathbf{X})$ is an MLE of θ, then $h\left(\widehat{\theta}_{\text{ML}}(\mathbf{X})\right)$ is an MLE of $h(\theta)$.* [Proof]

Example 8. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Gamma}(1, \theta)$ distribution, $\theta > 0$. Find an MLE of θ.

Example 9. Let $X_1, \ldots, X_n$ be a random sample from $\texttt{Poisson}(\theta)$ distribution, $\theta > 0$. Find an MLE of $P(X = 1) = \exp\{-\theta\}$.

Theorem 4 (Properties of MLE: 4, Large Sample Property)**.** *Under some regularity conditions, the following is true*

$$\sigma_\theta^{-1}\sqrt{n}\left\{\widehat{\theta}_{\mathrm{ML}}(\mathbf{X}) - \theta\right\} \xrightarrow{D} \mathrm{N}(0,1), \quad \textit{where} \quad \sigma_\theta^{-2} = E_\theta\left[\frac{\partial \log f_{\mathbf{X}}(\cdot;\theta)}{\partial\theta}\right]^2.$$

[Without Proof]

Corollary 2. *Theorem 4 implies that under appropriate regularity conditions, MLE is a consistent estimator of θ, i.e., $\widehat{\theta}_{\mathrm{ML}}(\mathbf{X}) \xrightarrow{p} \theta$. Further, it can be shown that MLE is an asymptotically efficient estimator of θ, i.e., the variance of MLE approaches the CRLB as $n \to \infty$.*