Lecture 05: Maximum Likelihood Estimation

Mathematical Statistics II, MATH 60062/70062

Thursday February 3, 2022

Reference: Casella & Berger, 7.2.2

Recap: Maximum likelihood estimator

For each sample $\boldsymbol{x}$, let $\hat{\boldsymbol{\theta}}(\boldsymbol{x})$ be a parameter value at which $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ attains its maximum as a function of $\boldsymbol{\theta}$, with $\boldsymbol{x}$ held fixed. A **maximum likelihood estimator (MLE)** of the parameter $\boldsymbol{\theta}$ based on a sample $\boldsymbol{X}$ is $\hat{\boldsymbol{\theta}}(\boldsymbol{X})$.

- The range of the MLE coincides with the range of the parameter.
- The MLE is the parameter point for which the observed sample is most likely.
- In general, the MLE is a good point estimator, possessing some of the optimality properties.

It is often easier to work with the **log-likelihood function**, for both analytical and computational reasons.

Normal MLE

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, 1)$, where $-\infty < \theta < \infty$. The log-likelihood function is

$$\log L(\theta \mid \boldsymbol{x}) = \log \left(\frac{1}{\sqrt{2\pi}} \right)^n - \frac{1}{2} \sum_{i=1}^{n} (x_i - \theta)^2.$$

Setting the first derivative

$$\frac{\partial}{\partial \theta} \log L(\theta \mid \boldsymbol{x}) = \sum_{i=1}^{n} (x_i - \theta)$$

equal to zero gives the solution $\hat{\theta} = \bar{x}$. Compared to the likelihood function, it is much easier to verify the concavity of $\log L(\theta \mid \boldsymbol{x})$,

$$\frac{\partial^2}{\partial \theta^2} \log L(\theta \mid \boldsymbol{x}) \Big|_{\theta = \bar{x}} = -n < 0.$$

Restricted range MLE

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\theta, 1)$, where the parameter θ is nonnegative. The MLE of θ is

$$\hat{\theta}(\boldsymbol{x}) = \arg\max_{\theta \in \Theta_0} L(\theta \mid \boldsymbol{x}),$$

where $\Theta_0 = \{\theta : \theta \geq 0\}$.

With no restriction on θ, the MLE of θ is $\bar{X}$. If $\bar{X}$ is negative, it will be outside the range of the parameter, and the (log-)likelihood function is decreasing in θ for $\theta \geq 0$ and is maximized at $\hat{\theta} = 0$. Hence, the restricted MLE is

$$\hat{\theta} = \begin{cases} \bar{X}, & \text{if } \bar{X} \geq 0 \\ 0, & \text{if } \bar{X} < 0. \end{cases}$$

Multivariate MLE

Suppose that $X_1, \ldots, X_n$ are iid $\mathcal{N}(\mu, \sigma^2)$, where $-\infty < \mu < \infty$ and $\sigma^2 > 0$. Both parameters are unknown. Let $\boldsymbol{\theta} = (\mu, \sigma^2)$. The likelihood function is

$$L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}.$$

The log-likelihood function is

$$\log L(\boldsymbol{\theta} \mid \boldsymbol{x}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

The score equations are

$$\frac{\partial}{\partial\mu}\log L(\boldsymbol{\theta} \mid \boldsymbol{x}) = \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu) = 0$$

$$\frac{\partial}{\partial\sigma^2}\log L(\boldsymbol{\theta} \mid \boldsymbol{x}) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - \mu)^2 = 0.$$

Solving the score equations gives the first-order critical point $(\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2/n)$.

Multivariate MLE, second-order condition

The second-order condition is verified with the **Hessian matrix**,

$$\begin{aligned}\boldsymbol{H} &= \frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\log L(\boldsymbol{\theta} \mid \boldsymbol{x}) \\ &= \begin{pmatrix} -\frac{n}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^n (x_i-\mu) \\ -\frac{1}{\sigma^4}\sum_{i=1}^n (x_i-\mu) & \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^n (x_i-\mu)^2 \end{pmatrix}.\end{aligned}$$

At a critical point, if the Hessian matrix is **negative definite**, then the function has a maximum. With $\boldsymbol{z} = (z_1, z_2)'$,

$$\boldsymbol{z}'\boldsymbol{H}\boldsymbol{z}|_{\mu=\hat{\mu},\sigma^2=\hat{\sigma}^2} = -\frac{nz_1^2}{\hat{\sigma}^2} - \frac{nz_2^2}{\hat{\sigma}^4} < 0.$$

Hence, $\hat{\boldsymbol{\theta}} = (\bar{x}, \sum_{i=1}^n (x_i - \bar{x})^2/n)'$ is the MLE of $\boldsymbol{\theta}$.

Invariance property of MLE

If $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Invariance property of MLE

If $\hat{\theta}$ is the MLE of θ, then for any function $\tau(\theta)$, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

It is easy to show this result if $\theta \to \tau(\theta)$ is one-to-one (see CB Theorem 7.2.10 for general case). Let $\eta = \tau(\theta)$, then $\tau^{-1}(\eta) = \theta$ is well defined and the likelihood function of $\tau(\theta)$ is given by

$$L^*(\eta \mid \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i \mid \tau^{-1}(\eta)) = L(\tau^{-1}(\eta) \mid \boldsymbol{x}).$$

The maximum of $L^*(\eta \mid \boldsymbol{x})$ is attained at $\eta = \tau(\hat{\theta})$ as

$$\sup_{\eta} L^*(\eta \mid \boldsymbol{x}) = \sup_{\eta} L(\tau^{-1}(\eta) \mid \boldsymbol{x}) = \sup_{\theta} L(\theta \mid \boldsymbol{x}).$$

Thus, $\hat{\eta} = \tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

Issues with MLE

Since MLEs are found by a maximization process, they are susceptible to the problems associated with that process.

- *Has a global maximum indeed been found?*
- *How sensitive is the estimate to small change in the data?* Sometimes, a slight different sample will produce a vastly different MLE. Such occurrences happen when the likelihood function is very flat in the neighborhood of its maximum. It is often wise to spend a little extra time investigating the stability of the solution.

See this great blog post by Radford Neal for other issues with MLE: https://radfordneal.wordpress.com/2008/08/09/inconsistent-maximum-likelihood-estimation-an-ordinary-example/