Section 6.2 Maximum Likelihood Estimation

Tonglin Zhang
Definition

The MLE, which attempts to maximize \( L(\theta) \) to estimate \( \theta \), is the most important approach in statistics. Let \( L(\theta) \) The maximum likelihood estimator (MLE) \( \hat{\theta} \) is the maximum of \( L(\theta) \), i.e,

\[
\hat{\theta} = \arg \max_\theta L(\theta).
\]
Properties

- The MLE \(\hat{\theta}\) is a function of data. Thus, it is random.
- For any continuous function \(g(\theta)\), the MLE of \(g(\theta)\) is \(g(\hat{\theta})\), which means it is transformation invariant.
- The choice of distributions is important in maximum likelihood estimation.
As estimator $\tilde{\theta}$ of $\theta$ is unbiased if

$$E\tilde{\theta} = \theta.$$ 

An unbiased estimator is not invariant under transformations. For example, Suppose that $\tilde{\theta}$ is an unbiased estimator of $\theta$. In general, $\tilde{\theta}^2$ is not an unbiased estimator of $\theta^2$:

$$E(\tilde{\theta}^2) \neq \theta^2.$$
6.2 Maximum Likelihood Estimation

**Computation**

Let $\ell(\theta) = \log L(\theta)$ is the loglikelihood function, where

$$\theta = (\theta_1, \ldots, \theta_k)^\top.$$

Then, $\theta$ is one of the solutions to

$$\nabla \ell(\theta) = \left( \frac{\partial \ell(\theta)}{\partial \theta_1}, \ldots, \frac{\partial \ell(\theta)}{\partial \theta_k} \right)^\top = 0.$$

We need to make sure the solution is a global maximum (this is a hard topic in research). If the solution is unique, then we guarantee it is the global maximum.
Example for SS, MSS, MLE, and Unbiasedness

Example: Let $X_1, \cdots, X_n$ be iid $Bernoulli(\theta)$. The PMF is $P(X = 1) = \theta; P(X = 0) = 1 - \theta$.

Solution: The joint PMF is

$$f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^{N} \theta^{X_i} (1 - \theta)^{1-X_i}$$

$$= \theta^{\sum_{i=1}^{n} X_i} (1 - \theta)^{\sum_{i=1}^{n} (1-X_i)}$$

$$= \theta^{\bar{X}} (1 - \theta)^{n(1-\bar{X})}.$$  

Be factorization theorem, we have that $SS = \{\bar{X}\}$. Because $\theta$ is one-dimensional, it is also the MSS.
As a function of $\theta$, the likelihood functions is

$$L(\theta) = \theta^n\bar{X} (1 - \theta)^{n(1-\bar{X})}.$$ 

The loglikelihood function is

$$\ell(\theta) = \log L(\theta) = n\bar{X} \log \theta + n(1 - \bar{X}) \log(1 - \theta).$$
Taking derivative, we have

\[ \ell'(\theta) = \frac{n\bar{X}}{\theta} + \frac{n(1 - \bar{X})}{1 - \theta}, \]

which is called the score function. Solving \( \ell'(\theta) = 0 \), we obtain the MLE

\[ \hat{\theta} = \bar{X}. \]

Because

\[ E(\hat{\theta}) = \theta, \]

we conclude that it is also an unbiased estimator. Moreover, \( \bar{X}^2 \) is the MLE but not an unbiased estimator of \( \theta^2 \).
Example: Let $X_1, \cdots, X_n$ be iid $\text{Poisson}(\theta)$.
Solution: The joint PMF is

$$f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^{n} \frac{\theta^{X_i}}{X_i!} e^{-\theta}$$

$$= \frac{\theta \sum_{i=1}^{n} X_i}{\prod_{i=1}^{N} X_i!} e^{-n\theta}$$

$$= \left( \prod_{i=1}^{n} X_i! \right)^{-1} \left( \theta^n \bar{X} e^{-n\theta} \right).$$

Note that only the second term contains both $\theta$ and data. We have $SS = \{ \bar{X} \}$. It is also the MSS because the size is 1.
6.2 Maximum Likelihood Estimation

Treating it as a function of $\theta$, we obtain the likelihood function as

$$L(\theta) = \left( \prod_{i=1}^{n} X_i! \right)^{-1} \left( \theta^n \bar{X} e^{-n\theta} \right).$$

The log-likelihood function is

$$\ell(\theta) = \log L(\theta) = - \log \left( \prod_{i=1}^{n} X_i! \right) + n\bar{X} \log \theta - n\theta.$$ 

Then,

$$\ell'(\theta) = \frac{n\bar{X}}{\theta} - n = 0 \Rightarrow \hat{\theta} = \bar{X}.$$ 

It is also unbiased because

$$\mathbb{E}(\hat{\theta}) = \theta.$$
Example: Example 6.2.3: Let $X_1, \cdots, X_n$ be iid $\text{Exp}(\theta)$. The PDF is $f(x) = \theta e^{-\theta x}$.

Solution: The joint PDF is

$$f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^{n} \theta e^{-\theta X_i}$$

$$= \theta^n e^{-n \bar{X}}$$

We have $SS = \{ \bar{X} \}$ and it is also the MSS because the size is 1.
Treating it as a function of \( \theta \), we obtain the likelihood function as

\[
L(\theta) = \theta^n e^{-n\bar{X}}.
\]

The log-likelihood function is

\[
\ell(\theta) = \log L(\theta) = n \log \theta - n\bar{X}.
\]

Then,

\[
\ell'(\theta) = \frac{n}{\theta} - \bar{X} = 0 \implies \hat{\theta} = \frac{1}{\bar{X}}.
\]

It is not unbiased (omitted).
Example: Example 6.2.4: Let $X_1, \cdots, X_n$ be iid from PMF $p_1 = P(X = 1) = \theta$, $p_2 = P(X_2) = \theta^2$ and $p_3 = P(X = 3) = 1 - \theta - \theta^2$. Check only the SS and MSS problem.

Solution: We express the PMF of

$$f_{\theta}(X_i) = \theta I(X_i=1)\theta^2 I(X_i=2)(1 - \theta - \theta^2) I(X_i=3).$$

Thus, the likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \left\{ \theta I(X_i=1)\theta^2 I(X_i=2)(1 - \theta - \theta^2) I(X_i=3) \right\}$$

$$=\theta^{n_1}\theta^{2n_2}(1 - \theta - \theta^2)^{1-n_1-n_2},$$

where $n_1$ is the total number of 1 and $n_2$ is the total number of 2 in the data. We have $SS = \{n_1, n_2\}$. Further, we can show it is MSS. Note that the size of $\theta$ is 1. The proof (omitted) is not easy.
Example: Example 6.2.5: Let $X_1, \cdots, X_n$ be iid $Uniform(\theta)$.

Solution: Let $X_{(1)} = \min(X_i)$ and $X_{(n)} = \max(X_i)$. We express the PDF as

$$f(x) = \frac{1}{\theta} I(0 \leq x \leq \theta) = \frac{1}{\theta} I(0 \leq x) I(x \leq \theta).$$

The likelihood function is

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I(0 \leq X_i) I(X_i \leq \theta) = \frac{1}{\theta^n} I(X_{(1)} \geq 0) I(X_{(n)} \leq \theta).$$

Thus, $SS = \{X_{(n)}\} = \{\max(X_i)\}$, which is also an MSS. Observe the above, we have the MLE

$$\hat{\theta} = X_{(n)}.$$
Next, we want to compute the PDF of $X_{(n)}$. For any $x \in [0, \theta]$,

$$P(X_{(n)} \leq x) = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x) = \prod_{i=1}^{n} P(X_i \leq x) = \frac{x^n}{\theta^n}.$$ 

Thus, the PDF of $X_{(n)}$ is

$$f_n(x) = \frac{d}{dx} \frac{x^n}{\theta^n} = \frac{n x^{n-1}}{\theta^n}.$$
Further, we have

\[ E(X_{(n)}) = \int_0^\theta x f_n(x) \, dx = \frac{n}{\theta^n} \int_0^\theta x^n \, dx = \frac{n\theta}{n + 1}. \]

By

\[ E(X_{(n)}^2) = \int_0^\theta x^2 f_n(x) \, dx = \frac{n}{\theta^n} \int_0^\theta x^{n+1} \, dx = \frac{n\theta^2}{n + 2}, \]

we have

\[ V(X_{(n)}) = \frac{n\theta^2}{n + 2} - \left( \frac{n\theta}{n + 1} \right)^2 = \frac{n\theta^2}{(n + 1)^2(n + 2)}. \]

Thus, the MLE is not unbiased. The bias is

\[ \text{Bias}(X_{(n)}) = E(X_{(n)}) - \theta = -\frac{\theta}{n + 1}. \]
Example: Example 6.2.2: Let $X_1, \cdots, X_n$ be iid $N(\mu, \sigma_0^2)$ with known $\sigma_0^2$.

Solution: Let $\theta = \mu$. The joint PDF is

$$f_{\theta}(X_1, \cdots, X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \mu)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \sigma_0^{-n} e^{-\frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \mu)^2}$$

$$= \left\{ (2\pi)^{-\frac{n}{2}} \theta_0^{-n} e^{-\frac{1}{2\theta_0^2} \sum_{i=1}^{n} (X_i - \bar{X})^2} \right\} \left\{ e^{-\frac{n}{2\sigma^2} (\bar{X} - \mu)^2} \right\}$$

Note that only the second term contains both $\mu$ and data. We have $SS = \{ \bar{X} \}$. It is also an MSS.
The log-likelihood function is

\[
\ell(\mu) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma_0^2 - \frac{1}{2\sigma_0^2} \sum_{i=1}^{n} (X_i - \bar{X})^2 - \frac{n}{2\sigma_0^2} (\bar{X} - \mu)^2.
\]

Then,

\[
\ell'(\mu) = \frac{n}{\sigma_0^2} (\bar{X} - \mu) \Rightarrow \hat{\mu} = \bar{X}.
\]

Clearly, it is unbiased.
Example: Example 6.2.6: Let $X_1, \cdots, X_n$ be iid $N(\mu, \sigma^2)$.
Solution: Let $\theta = (\theta_1, \theta_2)^\top = (\mu, \sigma^2)^\top$. The joint PDF is

$$f_\theta(X_1, \cdots, X_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(X_i - \mu)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \theta_2^{-\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{i=1}^{n}(X_i - \theta_1)^2}$$

$$= (2\pi)^{-\frac{n}{2}} \theta_2^{-\frac{n}{2}} e^{-\frac{1}{2\theta_2} [\sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \theta_1)^2]}$$

Note that only the second term contains both $\theta$ and data. We have $SS = \{ \bar{X}, \sum_{i=1}^{n}(X_i - \bar{X})^2 \}$. Using

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n}(X_i - \bar{X})^2,$$

we also have $SS = \{ \bar{X}, S^2 \}$ (SS is not unique). It is also the MSS because the size is 2, equal to the size of $\theta$. 
Treating it as a function of $\theta$, we obtain the likelihood function as

$$L(\theta) = (2\pi)^{-\frac{n}{2}} \theta_2^{-\frac{n}{2}} e^{-\frac{1}{2\theta_2} \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \theta_1)^2}.$$

The log-likelihood function is

$$\ell(\theta) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \theta_2 - \frac{1}{2\theta_2} \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \theta_1)^2.$$
Then,

\[
\frac{\partial \ell(\theta)}{\partial \theta_1} = \frac{n}{\theta_2}(\bar{X} - \mu)
\]

\[
\frac{\partial \ell(\theta)}{\partial \theta_2} = -\frac{n}{2\theta_2} + \frac{1}{2\theta_2^2} \left[ \sum_{i=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \theta_1)^2 \right].
\]

Thus, we have

\[
\hat{\mu} = \hat{\theta}_1 = \bar{X}
\]

\[
\hat{\sigma}^2 = \hat{\theta}_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2.
\]

Clearly \( \hat{\mu} \) is unbiased. Using \( E(S^2) = \sigma^2 \), we conclude that \( \hat{\sigma}^2 \) is not unbiased. It is biased. The bias is

\[
E(\hat{\sigma}^2) - \sigma^2 = E\left( \frac{n-1}{n} S^2 \right) - \sigma^2 = \frac{n-1}{n} \sigma^2 - \sigma^2 = -\frac{\sigma^2}{n}.
\]