MA204: Mathematical Statistics

Tutorial 4

T4.1 Maximum Likelihood Estimator (MLE)

Step 1: Calculate the log-likelihood function

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(x_i; \boldsymbol{\theta}).$$

Step 2: The MLE of $\boldsymbol{\theta}$ is obtained through

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta}).$$

Example T4.1 (Unrestricted MLEs of two parameters). Let $X_1, \ldots, X_n$ be a random sample from the distribution function

$$F(x; \theta_1, \theta_2) = \begin{cases} 1 - (\theta_1/x)^{\theta_2}, & \text{if } x \geqslant \theta_1, \\ 0, & \text{otherwise,} \end{cases}$$

where $\theta_1 > 0$ and $\theta_2 > 0$. Find the MLEs of θ_1 and θ_2.

Solution: The density function is given by

$$f(x; \theta_1, \theta_2) = \frac{\mathrm{d}}{\mathrm{d}x} F(x; \theta_1, \theta_2) = \begin{cases} \theta_1^{\theta_2} \theta_2 x^{-\theta_2 - 1}, & \text{if } x \geqslant \theta_1, \\ 0, & \text{otherwise.} \end{cases}$$

The joint density of $X_1, \ldots, X_n$ is

$$f(x_1, \ldots, x_n; \theta_1, \theta_2) = \begin{cases} \theta_1^{n\theta_2} \theta_2^n (x_1 \cdots x_n)^{-\theta_2 - 1}, & \text{if } x_i \geqslant \theta_1, \ \forall i = 1, \ldots, n, \\ 0, & \text{otherwise.} \end{cases}$$

So that the likelihood function is given by

$$L(\theta_1, \theta_2) = \begin{cases} \theta_1^{n\theta_2}\theta_2^n (x_1 \cdots x_n)^{-\theta_2 - 1}, & \text{if } 0 < \theta_1 \leqslant x_{(1)} = \min\{x_1, \ldots, x_n\} \text{ and } \theta_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Then the log-likelihood function is

$$\ell(\theta_1, \theta_2) = \begin{cases} (n\theta_2)\log\theta_1 + n\log\theta_2 - (\theta_2 + 1)\sum_{i=1}^n \log x_i, & \text{if } 0 < \theta_1 \leqslant x_{(1)} \text{ and } \theta_2 > 0, \\ 0, & \text{otherwise.} \end{cases}$$

By partially differentiating $\ell(\theta_1, \theta_2)$ with respect to θ_1, we have

$$\frac{\partial \ell(\theta_1, \theta_2)}{\partial \theta_1} = \frac{n\theta_2}{\theta_1} > 0, \qquad \text{since } n > 0,\ \theta_1 > 0 \text{ and } \theta_2 > 0.$$

That means $\ell(\theta_1, \theta_2)$ is an increasing function with respect to θ_1 when θ_2 is fixed, and since $0 < \theta_1 \leqslant x_{(1)}$, $\ell(\theta_1, \theta_2)$ is maximized at $\theta_1 = x_{(1)}$. Thus, the MLE of θ_1 is $\hat{\theta}_1 = X_{(1)}$. By partially differentiating $\ell(\theta_1, \theta_2)$ with respect to θ_2 and letting it equal zero, i.e.,

$$\frac{\partial \ell(\theta_1, \theta_2)}{\partial \theta_2} = n\log\theta_1 + \frac{n}{\theta_2} - \sum_{i=1}^n \log x_i = 0,$$

we obtain

$$\theta_2 = \frac{n}{\sum_{i=1}^n \log x_i - n\log\theta_1}.$$

Thus, the MLE of θ_2 is

$$\hat{\theta}_2 = \frac{n}{\sum_{i=1}^n \log X_i - n\log X_{(1)}}. \qquad \|$$

Example T4.2 (Restricted MLE of the parameter in Bernoulli distribution). Let $X_1, \ldots, X_n$ be a random sample from the Bernoulli distribution

$$f(x;\theta) = \theta^x(1-\theta)^{1-x}, \quad x = 0,\ 1,$$

where $0 < \theta \leqslant \frac{1}{2}$, i.e., the parameter space is $\Theta = \{\theta\colon\ 0 < \theta \leqslant \frac{1}{2}\}$. Find the MLE of θ.

Solution: The log-likelihood function is

$$\ell(\theta) = \sum_{i=1}^{n} \log f(x_i;\theta) = \left(\sum_{i=1}^{n} x_i\right) \log\theta + \left(n - \sum_{i=1}^{n} x_i\right) \log(1-\theta), \quad 0 < \theta \leqslant \frac{1}{2}.$$

Let $\bar{x} = \sum_{i=1}^{n} x_i/n$ and we have

$$\ell'(\theta) = \frac{\sum_{i=1}^{n} x_i}{\theta} + \frac{n - \sum_{i=1}^{n} x_i}{\theta - 1} = \frac{n(\bar{x}-\theta)}{\theta(1-\theta)}, \quad 0 < \theta \leqslant \frac{1}{2}.$$

Since x_i $(i = 1, \ldots, n)$ is either 0 or 1, $0 \leqslant \bar{x} \leqslant 1$.

If $0 < \bar{x} \leqslant \frac{1}{2}$, the solution to the equation $\ell'(\theta) = 0$ is $\theta = \bar{x}$.

If $\bar{x} > \frac{1}{2}$, the fact that $\ell'(\theta) > 0$ implies $\ell(\theta)$ is a strictly increasing function of θ. In this case, $\ell(\theta)$ is maximized at $\theta = \frac{1}{2}$.

Thus, the MLE of θ is $\hat{\theta} = \min(1/2, \bar{X})$. ‖

Example T4.3 (Quadratic maximization with a single parameter). Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} N(\mu, 1)$ with $a \leqslant \mu \leqslant b$, where a and b are two fixed constants. According to Example 3.7 of the Textbook "Mathematical Statistics" on pages 113–115, the restricted MLE of μ is $\hat{\mu} = \text{median}(a, \bar{X}, b)$. Find $E(\hat{\mu})$ and $\text{Var}(\hat{\mu})$

Solution: As $\bar{X} \sim N(\mu, 1/n)$, we have $E(\bar{X}^2) = \mu^2 + 1/n$. Note that

$$\begin{aligned} \hat{\mu} &= \text{median}(a, \bar{X}, b) \\ &= \begin{cases} a, & \text{if } \bar{X} < a, \\ \bar{X}, & \text{if } a \leqslant \bar{X} \leqslant b, \\ b, & \text{if } \bar{X} > b, \end{cases} \end{aligned}$$

which can be rewritten as

$\hat{\mu}$	a	$\bar{X}$	b
Probability	p_1	p_2	p_3

where

$$\begin{aligned}
p_1 &\triangleq \Pr(\bar{X} < a) = \Pr\{\sqrt{n}(\bar{X} - \mu) < \sqrt{n}(a - \mu)\} = \Phi(\sqrt{n}(a - \mu)),\\
p_2 &\triangleq \Pr(a \leqslant \bar{X} \leqslant b) = \Phi(\sqrt{n}(b - \mu)) - \Phi(\sqrt{n}(a - \mu)),\\
p_3 &\triangleq \Pr(\bar{X} > b) = 1 - \Phi(\sqrt{n}(b - \mu)).
\end{aligned}$$

Thus

$$\begin{aligned}
E(\hat{\mu}|\bar{X}) &= a \times p_1 + \bar{X} \times p_2 + b \times p_3,\\
E(\hat{\mu}^2|\bar{X}) &= a^2 \times p_1 + \bar{X}^2 \times p_2 + b^2 \times p_3,\\
E(\hat{\mu}) &= E\{E(\hat{\mu}|\bar{X})\} = ap_1 + \mu p_2 + bp_3,\\
E(\hat{\mu}^2) &= E\{E(\hat{\mu}^2|\bar{X})\} = a^2 p_1 + (\mu^2 + 1/n)p_2 + b^2 p_3,
\end{aligned}$$

and $\mathrm{Var}(\hat{\mu}) = E(\hat{\mu}^2) - \{E(\hat{\mu})\}^2$. ‖

T4.2 Newton's Method

4.2.1 Newton's method for root finding and optimization

(a) Root finding: For a given differentiable function $f(x)$, Newton's method is an iterative root finding technique to solve $f(x) = 0$, defined by

$$x^{(t+1)} = x^{(t)} - \frac{f(x^{(t)})}{f'(x^{(t)})},$$

where $x^{(0)}$ is an initial value.

(b) Optimization: For a twice differentiable function $g(x)$, under some conditions, an optimum $x^{(\infty)}$ satisfies $g'(x^{(\infty)}) = 0$. Then Newton's method for finding the maximizer or the minimizor of $g(x)$ is derived as

$$x^{(t+1)} = x^{(t)} - \frac{g'(x^{(t)})}{g''(x^{(t)})}.$$

4.2.2 Remarks

(a) Newton's method is highly sensitive to the initial value. Inappropriate initial values may lead to divergence or a local optimum.

(b) Besides, there is no assurance that all $x^{(t)}$ will locate in the support.

Example T4.4 (Maximizor of a function). Let

$$f(x) = \left(\frac{x}{2}\right)^{1/2} + 2\left(\frac{1-x}{3}\right)^{1/2}.$$

(a) Find the accurate x maximizing $f(x)$.

(b) Use Newton's method to calculate the numerical solution x^*. The initial value is set as $x^{(0)} = 0.1$. The stopping rule is: $|x^{(t+1)} - x^{(t)}| < 10^{-6}$.

Solution: (a) On the one hand, let

$$f'(x) = \frac{1}{4}\left(\frac{x}{2}\right)^{-1/2} - \frac{1}{3}\left(\frac{1-x}{3}\right)^{-1/2} = 0,$$

we obtain $x = 3/11$. On the other hand, since

$$\begin{aligned} f''(x) &= -\frac{1}{4}\left(\frac{1}{4}\right)\left(\frac{x}{2}\right)^{-3/2} - \left(\frac{1}{3}\right)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\left(\frac{1-x}{3}\right)^{-3/2} \\ &= -\frac{1}{16}\left(\frac{x}{2}\right)^{-3/2} - \frac{1}{18}\left(\frac{1-x}{3}\right)^{-3/2}, \end{aligned}$$

we have $f''(3/11) = -1.7066 < 0$, indicating that $f(x)$ has the strictly local maximum at $x = 3/11 \approx 0.2727273$ with $f(3/11) = 1.3540064$.

(b) Let $x^{(0)} = 0.1$, Newton's method shows that

$$\begin{aligned}
x^{(1)} &= x^{(0)} - \frac{f'(x^{(0)})}{f''(x^{(0)})} = 0.1859363, \\
x^{(2)} &= x^{(1)} - \frac{f'(x^{(1)})}{f''(x^{(1)})} = 0.2552335, \\
x^{(3)} &= x^{(2)} - \frac{f'(x^{(2)})}{f''(x^{(2)})} = 0.2721640, \\
x^{(4)} &= x^{(3)} - \frac{f'(x^{(3)})}{f''(x^{(3)})} = 0.2727267, \\
x^{(5)} &= x^{(4)} - \frac{f'(x^{(4)})}{f''(x^{(4)})} = 0.2727273.
\end{aligned}$$

Note that $|x^{(5)} - x^{(4)}| = 6 \times 10^{-7} < 10^{-6}$, thus the maximum of the $f(x)$ is gotten when $x = x^{(5)} = 0.2727273$ and $f(0.2727273) = 1.3540064$. $\|$

Example T4.5 (Newton'method for solving a quadratic/cubic equation). Let the log-likelihood function of θ be

$$\ell(\theta) = a\log(\theta) + b\log(1-\theta) + c\log(\theta+2), \quad \theta \in (0,1),$$

where $a = 34$, $b = 38$ and $c = 125$. Use the Newton method to calculate the mle of θ.

Solution: Although a closed–form solution to the following quadratic equation

$$0 = \ell'(\theta) = \frac{a}{\theta} - \frac{b}{1-\theta} + \frac{c}{\theta+2}$$

is available, for the illustration purpose, we apply the Newton method to obtain

$$\theta^{(t+1)} = \theta^{(t)} + [-\ell''(\theta^{(t)})]^{-1}\ell'(\theta^{(t)}),$$

where

$$-\ell''(\theta|Y_{\text{obs}}) = \frac{a}{\theta^2} + \frac{b}{(1-\theta)^2} + \frac{c}{(\theta+2)^2}.$$

Take $\theta^{(0)} = 0.5$, we obtain

$$\theta^{(1)} = 0.6363636, \quad \theta^{(2)} = 0.6269687, \quad \theta^{(3)} = 0.6268215, \quad \theta^{(4)} = 0.6268215.$$

R codes: The corresponding R code is as follows:

```
function(th0, NumNR)
{   # Function name: Linkage.model.NR(th0, NumNR)
    # -------------- Input ------------------------
    # th0   = initial value of \theta, th0 = 0.5
    # NumNR = the number of iterations in the
    #         Newton algorithm
    # -------------- Output -----------------------
    # TH    = approximates of the MLE of \theta
    # ---------------------------------------------
    a = 34; b = 38; cc = 125
    th <- th0
    TH <- matrix(0, NumNR, 1)
    for (tt in 1:NumNR) {
       Lp <- a/th - b/(1-th) + cc/(th + 2)
       nLpp <- a/th^2 + b/(1-th)^2 + cc/(th+2)^2
       th <- th + Lp/nLpp
       TH[tt] <- th
    }
    return(TH)
}
```

T4.3 Moment Estimator

Equate the sample moments to the corresponding population moments, and then solve the system of equations.

Example T4.6 (Moment estimators of parameters in uniform distribution). Let $X_1, \ldots, X_n$ be a random sample from a uniform distribution on the interval $[a, b]$. Find the moment estimators of a and b.

Solution: The pdf of a uniform distribution is

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{if } a \leqslant x \leqslant b, \\ 0, & \text{otherwise.} \end{cases}$$

Hence, the first two population moments are

$$E(X) = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2} \quad \text{and} \quad E(X^2) = \int_a^b \frac{x^2}{b-a} dx = \frac{a^2+ab+b^2}{3}.$$

Denote the first two sample moments as $\hat{\mu}_1$ and $\hat{\mu}_2$ respectively, we have

$$\hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{and} \quad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Equating the first two sample moments to the corresponding population moments, we obtain

$$\hat{\mu}_1 = \frac{a+b}{2} \quad \text{and} \quad \hat{\mu}_2 = \frac{a^2+ab+b^2}{3}$$

which, solving for a and b, results in the moment estimators of a and b,

$$\hat{a}^M = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)} \quad \text{and} \quad \hat{b}^M = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1^2)}. \qquad \|$$