單元 21: 最大概似法
(課本 §9.7)

最大概似法的構想說明: 設一盒中有 3 球, 某些爲紅球, 某些爲白球, 如圖示.

問. 若以不放回方式取 2 球, 如何估計盒中的紅球數 θ?

設取得 2 紅球 (以 A 表示), 問如何估計 θ?

答. 1. $\theta \geq 2$, 故 $\widehat{\theta}$ 爲 2 或 3 (不夠明確).

2. 若 $\theta = 2$, 則

$$P(A) = \frac{\binom{2}{2}\binom{1}{0}}{\binom{3}{2}} = \frac{1}{3}$$

若 $\theta = 3$, 則

$$P(A) = \frac{\binom{3}{2}}{\binom{3}{2}} = 1$$

此乃最大概似值. 因爲 A 發生, 故 $P(A)$ 應爲最大才合理 (因爲小的機率值表示不容易發生), 此乃相當於 $\theta = 3$, 得較合理的估計爲 $\widehat{\theta} = 3$.

最大概似法 (method of maximum likelihood)

設 $Y_1, \ldots, Y_n$ 爲一由含未知參數 $\theta_1, \ldots, \theta_k$ 的母體所取得的隨機樣本, 則最大概似法的步驟爲

1. 設定概似函數

$$L(y_1, \ldots, y_n | \theta_1, \ldots, \theta_k) \stackrel{\text{簡記}}{=} L(\theta_1, \ldots, \theta_k)$$

乃一 $\theta_1, \ldots, \theta_k$ 的函數 (因爲 $y_1, \ldots, y_n$ 爲已知的觀察值).

2. 求

$$\begin{aligned} \theta_1 &= \theta_1(y_1, \ldots, y_n) \\ \theta_2 &= \theta_2(y_1, \ldots, y_n) \\ &\vdots \\ \theta_k &= \theta_k(y_1, \ldots, y_n) \end{aligned}$$

使得 $L(\theta_1, \ldots, \theta_k)$ 爲最大. 一般而言, 此乃相當於針對方程組

$$\begin{aligned} \frac{\partial \ln L(\theta_1, \ldots, \theta_k)}{\partial \theta_1} &= 0 \\ \frac{\partial \ln L(\theta_1, \ldots, \theta_k)}{\partial \theta_2} &= 0 \\ &\vdots \\ \frac{\partial \ln L(\theta_1, \ldots, \theta_k)}{\partial \theta_k} &= 0 \end{aligned}$$

解 $\theta_1, \ldots, \theta_k$, 因為 ln(·) 為遞增函數,

$$L(\theta_1, \ldots, \theta_k)$$

與

$$\ln L(\theta_1, \ldots, \theta_k)$$

於相同的 $\theta_1, \ldots, \theta_k$ 產生極值.

3. 對於 $i = 1, \ldots, k$, θ_i 的最大概似估計量 (maximum likelihood estimator, MLE)

$$\widehat{\theta}_i = \text{步驟 2 求得的 } \theta_i$$

例 1. 設 $Y_1, \ldots, Y_n$ 是成功機率為 p 的 Bernoulli 試驗. 試求 p 的 MLE.

<解> 首先, 因為 Y_i 的 pmf

$$p(y_i|p) = p^{y_i}(1-p)^{1-y_i}, \ y_i = 0, 1$$

得概似函數

$$\begin{aligned} L(p) &= L(y_1, \ldots, y_n|p) = \prod_{i=1}^{n} \left[p^{y_i}(1-p)^{1-y_i}\right] \\ &= p^{\sum_{i=1}^n y_i}(1-p)^{n-\sum_{i=1}^n y_i} \\ &= p^y(1-p)^{n-y} \end{aligned}$$

其中 $y = \sum_{i=1}^{n} y_i$, $y_i = 0, 1$.

接著, 最大化 $L(p)$. (1) 若 $y = 0$, 則

$$L(p) = (1-p)^n$$

且在 $p = 0$ 有最大值.

(2) 若 $y = n$, 則

$$L(p) = p^n$$

且在 $p = 1$ 有最大值.

(3) 若 $y = 1, \ldots, n-1$, 則

$$L(p) = p^y (1-p)^{n-y}$$

在 $p = 0$ 與 $p = 1$ 時爲 0. 又當 $0 < p < 1$ 時, 最大化 $L(p)$ 等價於最大化

$$\ln L(p) = y \ln p + (n-y) \ln(1-p)$$

因爲 $\ln(\cdot)$ 爲遞增. 經由

$$\begin{aligned}\frac{d \ln L(p)}{dp} &= \frac{y}{p} + \frac{n-y}{1-p}(-1) \\ &= \frac{y - yp - np + yp}{p(1-p)} = \frac{y - np}{p(1-p)} = 0\end{aligned}$$

得臨界數

$$p = \frac{y}{n} \in (0, 1)$$

又根據 $\frac{d \ln L(p)}{dp}$ 在臨界數分割出的子區間上的符號, 如圖示, 可驗證出 $\ln L(p)$, 也就是 $L(p)$, 在 $p = \frac{y}{n}$ 有最大值.

因此, 綜合 (1)-(3), 得 p 的 MLE

$$\hat{p} = \frac{Y}{n}$$

因爲

$$y = 0,\ p = 0 = \frac{y}{n}$$

與

$$y = n,\ p = 1 = \frac{y}{n}$$

以及

$$y = 1, \ldots, n-1,\ p = \frac{y}{n}$$

均可表成 $\frac{y}{n}$.

例 2. 設隨機變數

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$$

其中 μ 與 σ^2 均未知. 試求 μ 與 σ^2 的 MLE.

<解> 首先, 因爲 Y_i 的 pdf

$$f(y_i|\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-(y_i-\mu)^2/2\sigma^2}$$

得概似函數

$$\begin{aligned}L(\mu,\sigma^2)&=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-(y_i-\mu)^2/2\sigma^2}\\ &= (2\pi)^{-\frac{n}{2}}(\sigma^2)^{-\frac{n}{2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right],\\ &\qquad -\infty<y_i<\infty, i=1,\ldots,n,\\ &\qquad -\infty<\mu<\infty,\sigma^2>0\end{aligned}$$

接著, 最大化 $L(\mu,\sigma^2)$, 此乃相當於解

$$\begin{aligned}\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}&=-\frac{1}{2\sigma^2}\sum_{i=1}^{n}2(y_i-\mu)(-1)\\ &=\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)=0 \qquad (1)\\ \frac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}&=-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu)^2\\ &=-\frac{1}{2\sigma^2}\left[n-\frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]\\ &=0 \qquad (2)\end{aligned}$$

由 (1) 式, 得

$$\sum_{i=1}^{n}(y_i-\mu)=\sum_{i=1}^{n}y_i-n\mu=0$$

故

$$\mu=\frac{1}{n}\sum_{i=1}^{n}y_i=\overline{y}$$

代入 (2) 式, 得

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\mu)^2=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2$$

最後, 經由驗證 (略, 自行練習), 得 $L(\mu,\sigma^2)$ 在臨界點

$$\left(\overline{y},\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2\right)$$

有最大值. 因此, μ 與 σ^2 的 MLE 分別爲

$$\widehat{\mu}=\overline{Y}$$

不偏的, 與

$$\widehat{\sigma^2}=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

偏差的, 但調整後,

$$\frac{n\widehat{\sigma^2}}{n-1}=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2=S^2$$

就是不偏的.

例 3. 設隨機變數

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \text{unif}(0, \theta)$$

試求 θ 的 MLE.

<解> 首先, 因為 Y_i 的 pdf

$$f(y_i|\theta) = \begin{cases} \frac{1}{\theta}, & 0 < y_i < \theta \\ 0, & \text{其它} \end{cases} = \frac{1}{\theta} I_{(0,\theta)}(y_i)$$

得概似函數

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{n} \frac{1}{\theta} I_{(0,\theta)}(y_i) \\ &= \frac{1}{\theta^n} I_{(-\infty,\theta)}(y_{(n)}) \cdot I_{(0,\infty)}(y_{(1)}) \end{aligned}$$

其中第一項僅是 $y_{(n)}$ 與 θ 的函數, 且第二項不含 θ 僅是 $y_1, \ldots, y_n$ 的函數.

接著, 最大化 $L(\theta)$. 給定觀察值 $y_1, \ldots, y_n$,

(1) 若 $y_{(1)} \leq 0$, 則

$$L(\theta) = 0$$

且取任何 θ 都產生最大值 0, 無意義.

(2) 若 $y_{(1)} > 0$, 則

$$L(\theta) = \frac{1}{\theta^n} I_{(y_{(n)},\infty)}(\theta)$$

且

(i) $\theta > y_{(n)}$ 時,

$$L(\theta) = \frac{1}{\theta^n}$$

乃一 θ 的遞減函數.

(ii) $\theta \le y_{(n)}$ 時,

$$L(\theta) = 0$$

如圖示.

故取 $\theta = y_{(n)}$ 得最大值.

因此, 綜合 (1) 與 (2), θ 的 MLE 為

$$\widehat{\theta} = Y_{(n)}$$

此外, 在 §9.6, 例 1 與 §9.2 中, 已分別證明 $Y_{(n)}$ 是一最小充分統計量以及 $\left(\frac{n+1}{n}\right)Y_{(n)}$ 是不偏的, 故

$$\left(\frac{n+1}{n}\right)Y_{(n)}$$

是 θ 的一 MVUE.

註 1. 若 U 是 θ 的任一充分統計量, 則 θ 的 MLE 會是 U 的某函數, 即 MLE 僅經由充分統計量的值而因觀察值而定.

<證> 由分解準則, 概似函數

$$\begin{aligned} L(\theta) &= L(y_1, \ldots, y_n|\theta) \\ &= g(u, \theta)h(y_1, \ldots, y_n) \end{aligned}$$

其中 $g(u, \theta)$ 僅是 u 與 θ 的函數且 $h(y_1, \ldots, y_n)$ 不因 θ 而定. 取對數, 得

$$\ln L(\theta) = \ln g(u, \theta) + \ln h(y_1, \ldots, y_n)$$

其中 $\ln h(y_1, \ldots, y_n)$ 不因 θ 而定. 故針對 θ, 最大化

$$\ln L(\theta)$$

乃相當於針對 θ, 最大化

$$\ln g(u, \theta)$$

又因 $\ln g(u,\theta)$ 僅經由充分統計量 U 的值 u 而定, 故 MLE 會是 U 的某一函數.

註 2. 若 U 是 θ 的充分統計量, 則根據註 1, θ 的 MLE $\widehat{\theta}$ 會是 U 的某函數.

透過調整, 得 $f(\widehat{\theta})$ 是不偏的.

一般而言, $f(\widehat{\theta})$ 就是一 MVUE, 如前面的例子.

MLE 的不變性 (invariance property). 若 $t(\theta)$ 是一 θ 的函數, 則 $t(\theta)$ 的 MLE

$$\widehat{t(\theta)} = t(\widehat{\theta})$$

意即, 將 $\widehat{\theta}$ 代入 $t(\cdot)$ 即可.

<證> 略, 自行參考習題 9.94, $t(\cdot)$ 是 1-1 的特例.

例 4. 在例 1, 已求得二項比率 p 的 MLE

$$\widehat{p} = \frac{Y}{n},\ Y \sim \text{binomial}(n,p)$$

試求 $\text{Var}(Y)$ 的 MLE.

<解> 因爲

$$\mathrm{Var}(Y) = np(1-p)$$

乃一 p 的函數, 故由 MLE 的不變性, 得

$$\begin{aligned}\widehat{\mathrm{Var}(Y)} &= np(1-p)|_{p=\widehat{p}=Y/n} \\ &= n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\end{aligned}$$

註. 因爲

$$\begin{aligned}E\left[n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\right] &= n\left[E\left(\frac{Y}{n}\right) - E\left[\left(\frac{Y}{n}\right)^2\right]\right] \\ &= n\left[p - \mathrm{Var}\left(\frac{Y}{n}\right) - \left(E\left(\frac{Y}{n}\right)\right)^2\right] \\ &= n\left[p - \frac{p(1-p)}{n} - p^2\right] = np(1-p)\left(1-\frac{1}{n}\right)\end{aligned}$$

故

$$n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)$$

不是 $np(1-p)$ 的不偏估計量, 但調整後, 得

$$\begin{aligned}&n\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\left(1-\frac{1}{n}\right)^{-1} \\ &\qquad = \frac{n^2}{n-1}\left(\frac{Y}{n}\right)\left(1-\frac{Y}{n}\right)\end{aligned}$$

是不偏的, 且是一 MVUE.