單元 33: 概似比值檢定
(課本 §10.11)

此法適用於單一或複合假設以及有或沒有含干擾參數 (nuisance parameters) 的情況.

設隨機變數

$$Y_1, \ldots, Y_n$$

爲一由含參數 $\theta_1, \ldots, \theta_k$ 的母體所取得的隨機樣本, 其中 k 個參數中, 一爲欲檢測的, 其餘爲已知的或干擾的. 爲簡化起見, 簡記

$$\Theta = (\theta_1, \ldots, \theta_k)$$

則此樣本的概似函數

$$L(\Theta) \overset{\text{記成}}{=} L(y_1, \ldots, y_n | \theta_1, \ldots, \theta_k)$$

乃一 k 個參數的函數. 例如,

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

其中 μ 與 σ^2 均未知, 則

$$\Theta = (\mu, \sigma^2)$$

若僅檢測 μ, 則 $L(\Theta)$ 爲 μ 與干擾參數 σ^2 的函數.

設 Ω_0 爲虛無假設所確定的 Θ 的可能值集合且 Ω_a 爲對立假設所確定的 Θ 的另一可能值集合而且

$$\Omega_0 \cap \Omega_a = \emptyset$$

又令 $\Omega = \Omega_0 \cup \Omega_a$, 則檢定

$$H_0 : \Theta \in \Omega_0 \text{ 對立於 } \Theta \in \Omega_a$$

時, H_0 與 H_a 可爲單一或複合的, 因爲可能含有欲估計參數的多個值或其它干擾參數. 例如,

$$Y_1, \ldots, Y_n \overset{\text{iid}}{\sim} \exp(\lambda)$$

且檢定

$$H_0 : \lambda = \lambda_0 \text{ (單一)} \quad \text{對立於 } H_a : \lambda \neq \lambda_0 \text{ (複合)}$$

則

$$\Theta = \lambda, \ \Omega_0 = \{\lambda_0\}, \ \Omega_a = \{\lambda > 0 : \lambda \neq \lambda_0\}$$

以及

$$\begin{aligned} \Omega &= \Omega_0 \cup \Omega_a = \{\lambda_0\} \cup \{\lambda > 0 : \lambda \neq \lambda_0\} \\ &= \{\lambda : \lambda > 0\} \end{aligned}$$

接著, 令

$$L(\hat{\Omega}_0) = \max_{\Theta \in \Omega_0} L(\Theta)$$

事實上, 爲上界值, supremum, 表示對所有 Ω_0 中的 Θ, 對觀察資料的最佳解釋 (best explanation, 以 Ω_0 中所對應的最大機率說明出現或發生的理由), 可由求 MLE 的方法求得.

同理, 令

$$L(\hat{\Omega}) = \max_{\Theta \in \Omega} L(\Theta)$$

表示對所有 Ω 中的 Θ, 對觀察資料的最佳解釋.

直觀上,

1. 若 $L(\hat{\Omega}_0) = L(\hat{\Omega})$, 則表示觀察資料的最佳解釋可在 Ω_0 中求得, 故不拒絕 H_0.

2. 若 $L(\hat{\Omega}_0) < L(\hat{\Omega})$, 則表示觀察資料的最佳解釋可在 Ω_a 中求得, 故拒絕 H_0 而支持 H_a

由此, 得概似比值檢定 (a likelihood ratio test, LRT): 定義

$$\lambda = \frac{L(\hat{\Omega}_0)}{L(\hat{\Omega})}$$

乃一 $y_1, \ldots, y_n$ 的函數, 故爲一統計量. 則在意義水準 α 下, 檢定

$$H_0 : \Theta \in \Omega_0 \text{ 對立於 } H_a : \Theta \in \Omega_a$$

的概似比值檢定乃是以 λ 作爲檢定統計量且拒絕域

$$\text{RR} = \{\lambda \le k\}$$

其中 k 由 α 所決定.

註. 因爲 $\Omega_0 \subset \Omega$, 故 $0 \le \lambda \le 1$; $\lambda \downarrow 0$ 表示

$$\text{在 } H_0 \text{ 的概似值 } \ll \text{ 在 } H_a \text{ 的概似值}$$

故資料愈顯示支持 H_a, 而勝於支持 H_0, 故拒絕 H_0.

例 1. 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. 試求在意義水準 α 下, 檢定

$$H_0 : \mu = \mu_0 \text{ 對立於 } H_a : \mu > \mu_0$$

的 LRT.

註. 直觀上, 可採用統計檢定量

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$$

當 H_0 眞時且拒絕域

$$\text{RR} = \{t > t_\alpha\}$$

此處將證明此 t-檢定就是一 LRT.

<解> 因爲 σ^2 是干擾參數, 故 Neyman-Pearson 引理不適用. 首先,

$$\Theta = (\mu, \sigma^2)$$

且

$$\Omega_0 = \{(\mu_0, \sigma^2) : \sigma^2 > 0\}$$

與

$$\Omega_a = \{(\mu, \sigma^2) : \mu > \mu_0, \sigma^2 > 0\}$$

以及

$$\Omega = \Omega_0 \cup \Omega_a$$

接著, 由常態分布, 得

$$\begin{aligned} L(\Theta) &= L(\mu, \sigma^2) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma^2}\right)^{n/2} \cdot \\ &\qquad \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right] \end{aligned}$$

且

1. 在 Ω_0 下, $\mu = \mu_0$, 故最大化

$$L(\Theta) = L(\mu_0, \sigma^2),\ \sigma^2 > 0$$

乃相當於解

$$\begin{aligned}\frac{\partial \ln L(\mu_0,\sigma^2)}{\partial \sigma^2} &= \frac{\partial}{\partial \sigma^2}\left[-n\ln\sqrt{2\pi} - \frac{n}{2}\ln\sigma^2 \right.\\ &\qquad \left. -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu_0)^2\right]\\ &= -\frac{n}{2\sigma^2} + \\ &\qquad \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(y_i-\mu_0)^2 = 0\end{aligned}$$

得最大值發生在

$$\hat\sigma_0^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i-\mu_0)^2$$

自行驗證. 故, 代入 μ_0 與 $\hat\sigma_0^2$, 得

$$\begin{aligned}L(\hat\Omega_0) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat\sigma_0^2}\right)^{n/2} \cdot \\ &\qquad \exp\left[-\frac{1}{2\hat\sigma_0^2}\sum_{i=0}^{n}(y_i-\mu_0)^2\right]\\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\hat\sigma_0^2}\right)^{n/2} e^{-n/2}\end{aligned}$$

2. 在 Ω 下, 最大化

$$L(\Theta) = L(\mu, \sigma^2),\ \ \mu \geq \mu_0,\ \sigma^2 > 0$$

乃相當於解

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

及

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

且 $\mu \geq \mu_0$, $\sigma^2 > 0$, 並驗證. 首先, 由 (1) 式, 得

$$\frac{n}{\sigma^2}(\overline{y}-\mu) = 0$$

即 $\mu = \overline{y}$ 且在二子區間上

$$(-\infty, \overline{y}):\ \frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = (+)$$

$$(\overline{y}, \infty):\ \frac{\partial \ln L(\mu,\sigma^2)}{\partial \mu} = (-)$$

所以, 最大值發生在

$$\widehat{\mu} = \begin{cases} \overline{y}, & \text{若 } \overline{y} > \mu_0 \\ \mu_0, & \text{若 } \overline{y} \le \mu_0 \end{cases}$$

如圖示. 代入 (2) 式, 得最大值發生在

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

自行驗證.

故代入 $\widehat{\mu}$ 與 $\widehat{\sigma}^2$, 得

$$\begin{aligned} L(\widehat{\Omega}) &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\widehat{\sigma}^2}\right)^{n/2} \cdot \\ &\qquad \exp\left[-\frac{1}{2\widehat{\sigma}^2}\sum_{i=1}^{n}(y_i - \widehat{\mu})^2\right] \\ &= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\widehat{\sigma}^2}\right)^{n/2} e^{-n/2} \end{aligned}$$

因此,

$$\begin{aligned} \lambda &= \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}\right)^{\frac{n}{2}} = \left[\frac{\sum_{i=1}^{n}(y_i - \widehat{\mu})^2}{\sum_{i=1}^{n}(y_i - \mu_0)^2}\right]^{n/2} \\ &= \begin{cases} \left[\frac{\sum_{i=1}^{n}(y_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \mu_0)^2}\right]^{n/2}, & \overline{y} > \mu_0 \\ 1, & \overline{y} \le \mu_0 \end{cases} \end{aligned}$$

且 RR $=\{\lambda\le k\}$, 某常數 $k<1$ (因爲若 $k=1$, 則 RR 等於整個 R^n, 無意義), 以及 $\overline{y}>\mu_0$. 因爲

$$\begin{aligned}
&\sum_{i=1}^n (y_i-\mu_0)^2\\
&= \sum_{i=1}^n (y_i-\overline{y}+\overline{y}-\mu_0)^2\\
&= \sum_{i=1}^n (y_i-\overline{y})^2+2(\overline{y}-\mu_0)\sum_{i=1}^n (y_i-\overline{y})\\
&\qquad +n(\overline{y}-\mu_0)^2\\
&= \sum_{i=1}^n (y_i-\overline{y})^2+n(\overline{y}-\mu_0)^2
\end{aligned}$$

其中最後一個等號成立乃因

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

所致, 以及兩邊同取 $2/n$ 次方, 故得 $\lambda\le k$ 等價於

$$\frac{\sum_{i=1}^n (y_i-\overline{y})^2}{\sum_{i=1}^n (y_i-\overline{y})^2+n(\overline{y}-\mu_0)^2}\le k^{2/n}=k_1$$

另一常數, 亦相當於

$$\frac{1}{1+\frac{n(\overline{y}-\mu_0)^2}{\sum_{i=1}^n (y_i-\overline{y})^2}}\le k_1$$

顚倒, 整理, 得

$$\frac{n(\overline{y}-\mu_0)^2}{\sum_{i=1}^n (y_i-\overline{y})^2} \geq \frac{1}{k_1}-1=k_2$$

另一常數. 同乘 $n-1$, 得

$$\begin{aligned}
\frac{n(\overline{y}-\mu_0)^2}{\frac{1}{n-1}\sum_{i=1}^n (y_i-\overline{y})^2} &= \frac{n(\overline{y}-\mu_0)^2}{S^2} \\
&= \frac{(\overline{y}-\mu_0)^2}{S^2/n} \geq (n-1)k_2
\end{aligned}$$

最後, 因爲 $\overline{y}>\mu_0$, 故兩邊開方, 得

$$\frac{\overline{y}-\mu_0}{S/\sqrt{n}} \geq \sqrt{(n-1)k_2}=k_3$$

另一常數. 又由母體的常態性假設, 在 $H_0: \mu=\mu_0$ 下, 得

$$\frac{\overline{Y}-\mu_0}{S/\sqrt{n}} \sim t(n-1)$$

並由

$$\begin{aligned}
\alpha &= P\left(\frac{\overline{Y}-\mu_0}{S/\sqrt{n}} \geq k_3, \text{ 當 } \mu=\mu_0\right) \\
&= P(t(n-1)\geq k_3)
\end{aligned}$$

以及

$$\alpha = P(t(n-1)>t_\alpha)$$

得 $k_3 = t_\alpha$. 因此, 在意義水準 α 下, LRT 的檢定統記量爲

$$T = \frac{\overline{Y} - \mu_0}{S/\sqrt{n}}$$

且 RR $= \{t > t_\alpha\}$, 即 t-檢定.

註 1. 如例 1, LRT 是熟知型式的現象並非不尋常; 事實上, 所有 §10.8 與 §10.9 的檢定均可由概似比值法求得.

註 2. 對大多數的實務問題, 由概似比值法可得最好檢定力的檢定.

註 3. 不幸, 概似比值法所得出的檢定統計量不常有已知的機率分布, 如例 1; 但當樣本夠大且母體 (或多個母體們) 符合正規條件 (regularity condition), 主要如 (1) 概似函數對參數的導函數存在與 (2) 概似函數爲正的區域不因未知參數值而定, 則可得 λ 的近似分布, 如下述的定理 2.

定理 2. 令隨機變數 $Y_1, \ldots, Y_n$ 的聯合概似函數爲 $L(\Theta)$ 且 r_0 爲 $H_0 : \Theta \in \Omega_0$ 所確定的自由參數

(free parameters) 的個數以及 r 爲 $\Theta \in \Omega$ 所確定的自由參數的個數, 則對於大的 n,

$$-2\ln\lambda \overset{\mathcal{D}}{\approx} \chi^2(r_0 - r)$$

註 1. LRT 的拒絕域

$$\text{RR} = \{\lambda \le k\} = \{-2\ln\lambda \ge -2\ln k = k^*\}$$

故, 當 n 夠大時, 由

$$\alpha = P(-2\ln\lambda \ge k^*) \approx P(\chi^2(r_0 - r) \ge k^*)$$

以及

$$\alpha = P(\chi^2(r_0 - r) \ge \chi^2_\alpha(r_0 - r))$$

得, 在意義水準 α 下, 大樣本概似比值檢定的拒絕域

$$\text{RR} = \{-2\ln\lambda \ge \chi^2_\alpha(r_0 - r)\}$$

註 2. 樣本多大才得 "好的" 近似是隨應用問題而異的.

例 2. 設一工程師欲比較二輪班工頭每週提出的申訴案件數. 今各取 100 個獨立申訴案件數, 得輪班 1 的平均值 $\overline{x} = 20$, 輪班 2 的平均值 $\overline{y} = 22$. 又假設輪班 i 的申

訴案件數爲 Poisson(θ_i), $i = 1, 2$. 試在意義水準 $\alpha \approx 0.01$ 下, 以概似比值法檢定

$$H_0 : \theta_1 = \theta_2 \text{ 對立於 } H_a : \theta_1 \neq \theta_2$$

<解> 因爲對於 $i = 1, \ldots, 100$,

$$f(x_i|\theta_1) = \frac{\theta_1^{x_i}}{x_i!} e^{-\theta_1}$$

且

$$f(y_i|\theta_2) = \frac{\theta_2^{y_i}}{y_i!} e^{-\theta_2}$$

以及獨立性, 得 $x_1, \ldots, x_{100}$ 與 $y_1, \ldots, y_{100}$ 的聯合概似函數

$$L(\theta_1, \theta_2) = \frac{1}{k} \theta_1^{\sum x_i} \theta_2^{\sum y_i} e^{-n(\theta_1 + \theta_2)}$$

其中

$$k = x_1! \cdots x_n! y_1! \cdots y_n!, \;\; n = 100$$

又

$$\Theta = (\theta_1, \theta_2)$$

且

$$\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$$

θ 未知, 與

$$\Omega_a = \{(\theta_1, \theta_2) : \theta_1 \neq \theta_2\}$$

以及

$$\Omega = \Omega_0 \cup \Omega_a = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$$

接著,

1. 在 Ω_0 下, 最大化

$$L(\Theta) = L(\theta) = \frac{1}{k}\theta^{\sum x_i + \sum y_i} e^{-2n\theta}$$

乃相當於解

$$\begin{aligned}
&\frac{\partial \ln L(\theta)}{\partial \theta} \\
&= \frac{\partial}{\partial \theta}\left[-\ln k + \left(\sum x_i + \sum y_i\right)\ln \theta - 2n\theta\right] \\
&= \frac{\sum x_i + \sum y_i}{\theta} - 2n = 0
\end{aligned}$$

得最大值發生在

$$\widehat{\theta} = \frac{1}{2n}\left(\sum x_i + \sum y_i\right) = \frac{1}{2}(\overline{x} + \overline{y})$$

自行驗證. 代入 $\widehat{\theta}$, 得

$$\begin{aligned}
L(\widehat{\Omega}_0) &= \frac{1}{k}(\widehat{\theta})^{n\overline{x}+n\overline{y}} e^{-2n\widehat{\theta}} \\
&= \frac{1}{k}(\widehat{\theta})^{n\overline{x}+n\overline{y}} e^{-n(\overline{x}+\overline{y})}
\end{aligned}$$

2. 在 Ω 下, 最大化

$$L(\Theta)=L(\theta_1,\theta_2)=\frac{1}{k}\theta_1^{\sum x_i}\theta_2^{\sum y_i}e^{-n(\theta_1+\theta_2)}$$

乃相當於解

$$\begin{aligned}
&\frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_1}\\
&= \frac{\partial}{\partial\theta_1}\Big[-\ln k+\left(\sum x_i\right)\ln\theta_1\\
&\qquad\qquad +\left(\sum y_i\right)\ln\theta_2-n(\theta_1+\theta_2)\Big]\\
&= \frac{\sum x_i}{\theta_1}-n=0
\end{aligned}$$

與

$$\frac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_2}=\frac{\sum y_i}{\theta_2}-n=0$$

並驗證. 得最大值發生在

$$\widehat{\theta}_1=\frac{\sum x_i}{n}=\overline{x}$$

與

$$\widehat{\theta}_2=\frac{\sum y_i}{n}=\overline{y}$$

代入, 得

$$L(\widehat{\Omega})=\frac{1}{k}(\widehat{\theta}_1)^{n\overline{x}}(\widehat{\theta}_2)^{n\overline{y}}e^{-n(\widehat{\theta}_1+\widehat{\theta}_2)}$$

因此,

$$\begin{aligned}\lambda &= \frac{L(\widehat{\Omega}_0)}{L(\widehat{\Omega})} = \frac{(\widehat{\theta})^{n\overline{x}+n\overline{y}}e^{-2n\widehat{\theta}}}{(\widehat{\theta}_1)^{n\overline{x}}(\widehat{\theta}_2)^{n\overline{y}}e^{-n(\widehat{\theta}_1+\widehat{\theta}_2)}} \\ &= \frac{(\widehat{\theta})^{n\overline{x}+n\overline{y}}e^{-n(\overline{x}+\overline{y})}}{(\widehat{\theta}_1)^{n\overline{x}}(\widehat{\theta}_2)^{n\overline{y}}e^{-n(\overline{x}+\overline{y})}} = \frac{(\widehat{\theta})^{n\overline{x}+n\overline{y}}}{(\widehat{\theta}_1)^{n\overline{x}}(\widehat{\theta}_2)^{n\overline{y}}}\end{aligned}$$

且拒絕域

$$\begin{aligned}\text{RR} &= \{\lambda \leq k\},\ k < 1 \\ &= \{-2\ln\lambda \geq -2\ln k = k^*\}\end{aligned}$$

又

$$\Omega = \{(\theta_1, \theta_2) : \theta_1 > 0,\ \theta_2 > 0\}$$

有二個自由參數, 得

$$\Omega_0 = \{(\theta_1, \theta_2) : \theta_1 = \theta_2 = \theta\}$$

有 $r_0 = 1$ 個被確定且 Ω 有 $r = 0$ 個被確定, 故在 $\alpha \approx 0.01$ 且 $n = 100$ 的大樣本下, 由表 6,

$$\begin{aligned}\text{RR} &= \{-2\ln\lambda \geq \chi^2_{0.01}(1)\} \\ &= \{-2\ln\lambda > 6.635\}\end{aligned}$$

或

R: $\chi^2_{0.01}(1) = \texttt{qchisq}(0.99, 1)$

最後, 由資料, 得觀察值

$$\widehat{\theta} = \frac{1}{2}(\overline{x} + \overline{y}) = \frac{1}{2}(20 + 22) = 21$$

與

$$\lambda = \frac{(21)^{100(20)+100(22)}}{(20)^{100(20)}(22)^{100(22)}}$$

以及

$$\begin{aligned} & -2\ln\lambda \\ &= -2(4200\ln 21 - 2000\ln 20 - 2200\ln 22) \\ &= 9.53 \in \mathrm{RR} \end{aligned}$$

因此, 在意義水準 $\alpha \approx 0.01$ 下, 拒絕 H_0, 即二輪班工頭提出的平均申訴案件數確實有差異.