# MLE Basics

$$y_i = x_i'\beta + u_i, \quad u_i \sim NID(0, \sigma^2)$$

normal distribution:

$$f(y_i|x_i;\beta,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma^2}\right\}$$

the loglikelihood function:

$$\log L(\beta,\sigma^2) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{N}\frac{(y_i - x_i'\beta)^2}{\sigma^2}$$

the score vector:

$$s(\theta) \equiv \frac{\partial \log L(\theta)}{\partial \theta} = \sum_{i=1}^{N}\frac{\partial \log L_i(\theta)}{\partial \theta} \equiv \sum_{i=1}^{N} s_i(\theta)$$

$$s_i(\beta,\sigma^2) = \begin{pmatrix} \frac{\partial \log L_i(\beta,\sigma^2)}{\partial \beta} \\ \frac{\partial \log L_i(\beta,\sigma^2)}{\partial \sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{(y_i - x_i'\beta)}{\sigma^2} x_i \\ -\frac{1}{2\sigma^2} + \frac{1}{2}\frac{(y_i - x_i'\beta)^2}{\sigma^4} \end{pmatrix}$$

FOC implies:

$$\hat{\beta} = \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1}\sum_{i=1}^{N} x_i y_i \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - x_i'\hat{\beta})^2$$

Information matrix:

$$I(\theta) \equiv E[s(\theta)s(\theta)'] = -E[\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}]$$

$$I_i(\beta,\sigma^2) = E[s_i(\beta,\sigma^2)s_i(\beta,\sigma^2)']$$

calculation:

$$E(s_i s_i') = E[\begin{matrix} \frac{u_i}{\sigma^2}\cdot x_i \\ -\frac{1}{2\sigma^2} + \frac{u_i^2}{2\sigma^4} \end{matrix}] \cdot [\frac{u_i}{\sigma^2}x_i, -\frac{1}{2\sigma^2} + \frac{u_i^2}{2\sigma^4}]$$

$$= E\begin{bmatrix} \left(\frac{u_i}{\sigma^2}x_i\right)^2 & -\frac{u_i x_i}{2\sigma^4} + \frac{u_i^3 x_i}{2\sigma^6} \\ -\frac{u_i x_i}{2\sigma^4} + \frac{u_i^3 x_i}{2\sigma^6} & \left(-\frac{1}{2\sigma^2} + \frac{u_i^2}{2\sigma^4}\right)^2 \end{bmatrix}$$

since $E(u_i) = 0$, $E(u_i^2) = \sigma^2$, $E(u_i^3) = 0$, $E(u_i^4) = 3\sigma^4$, we can finally get

$$I_i(\beta,\sigma^2)=\begin{pmatrix}\frac{1}{\sigma^2}x_ix_i' & 0\\ 0 & \frac{1}{2\sigma^4}\end{pmatrix}$$

The asymptotic covariance matrix:

$$V=I(\beta,\sigma^2)^{-1}=\begin{pmatrix}\sigma^2\Sigma_{xx}^{-1} & 0\\ 0 & 2\sigma^4\end{pmatrix}$$

$$\sqrt{N}(\hat\beta-\beta)\to\mathcal{N}(0,\sigma^2\Sigma_{xx}^{-1})$$

$$\sqrt{N}(\hat\sigma^2-\sigma^2)\to\mathcal{N}(0,2\sigma^4)$$

---

# Hypothesis test

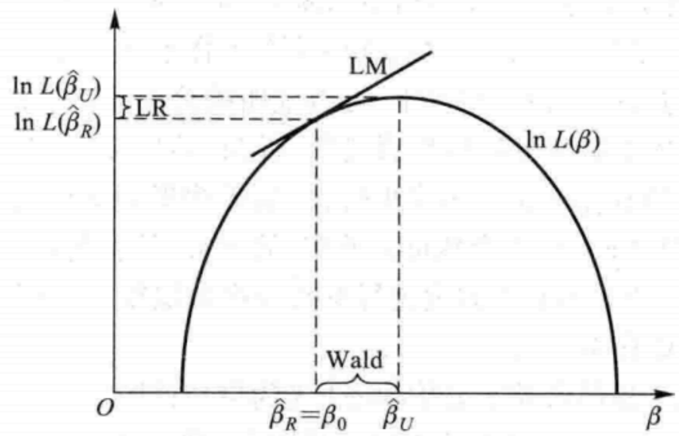


图 6.6　三类渐近等价的统计检验

## LR(Likelihood Ratio Test)

$$\log L=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log(\tilde\sigma^2)-\frac{1}{2\tilde\sigma^2}\tilde u'\tilde u$$

$$SSR=\tilde u'\tilde u,\ \tilde\sigma^2=\frac{SSR}{n}$$

$$\Rightarrow logL=-\frac{n}{2}\log(2\pi)-\frac{n}{2}\log\frac{SSR}{n}-\frac{n}{2}$$

$$\Rightarrow L=(\frac{2\pi}{n})^{-\frac{n}{2}}\cdot e^{-\frac{n}{2}}\cdot(SSR)^{-\frac{n}{2}}$$

$$\lambda\equiv L_U/L_R\equiv(\frac{SSR_U}{SSR_R})^{-n/2}$$

$$F=\frac{(SSR_R-SSR_U)/\#r}{SSR_U/n-k-1}=\frac{n-k-1}{\#r}(\lambda^{\frac{2}{n}}-1)$$

## Wald Test

We have already seen in chap 2:

$$W \equiv (R\tilde{\beta}-r)'\left[s^2 R(\mathbf{X}'\mathbf{X})^{-1}R'\right]^{-1}(R\tilde{\beta}-r) \xrightarrow{d} \chi^2_J$$

$$W = n \cdot h(b_{OLS})'\left[H(b_{OLS})Avar(b_{OLS})H(b_{OLS})'\right]^{-1} h(b_{OLS}) \xrightarrow{d} \chi^2(p)$$

For nonlinear hypotheses:

$$H_0 : h(\theta) = 0$$

*By first-order Taylor expansion:*

$$h(\hat{\theta}) \approx H(\theta)(\hat{\theta}-\theta)$$

$$H(\theta) = \frac{\partial h(\theta)}{\partial \theta'}$$

Wald Test Statistic:

$$W = n \cdot h(\hat{\theta})'[\hat{H}\widehat{Avar}(\hat{\theta})\hat{H}']^{-1}h(\hat{\theta})$$

$\widehat{Avar}(\hat{\theta})$ can be the inverse of information matrix.

Wald Test Examples(see lecture note):
(1)Tests of Exclusion Restrictions
(2)Tests of Statistical Significance
(3)Tests of Nonlinear Restriction

---

## Examples

Verbeek 6.1.1+6.2.3:

(1)

$$L(p) = \binom{N}{N_1} p^{N_1}(1-p)^{N-N_1}$$

$$\log L(p) = \log \binom{N}{N_1} + N_1 \log p + (N-N_1)\log(1-p)$$

(2)

$$\frac{d \log L(p)}{dp} = \frac{N_1}{p} - \frac{N-N_1}{1-p} = 0$$

$$\hat{p} = \frac{N_1}{N}$$

(3)

$$\log L_i(p) = y_i \log p + (1-y_i)\log(1-p)$$

$$\frac{\partial \log L_i(p)}{\partial p} = \frac{y_i}{p} - \frac{1-y_i}{1-p}$$

$$-\frac{\partial^2 \log L_i(p)}{\partial p^2} = \frac{y_i}{p^2} + \frac{1-y_i}{(1-p)^2}$$

$$I = E\left\{-\frac{\partial^2 \log L_i(p)}{\partial p^2}\right\} = \frac{E\{y_i\}}{p^2} + \frac{1-E\{y_i\}}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}$$

note that $E\{y_i\} = p$

$$V = I^{-1}$$

$$\sqrt{n}(\hat{p} - p) \to \mathcal{N}(0, p(1-p))$$

(4)

$$p \pm z_{a/2}\sqrt{\frac{p(1-p)}{n}} = 0.44 \pm 1.96\sqrt{\frac{0.44(1-0.44)}{100}} = (0.3427, 0.5373)$$

(5) $H_0 : \hat{p} = p_0$ for a given value $p_0$
Wald test :

$$W = N(\hat{p} - p_0)[\hat{p}(1-\hat{p})]^{-1}(\hat{p} - p_0)$$

For LR test we need to compare the unrestricted and the restricted model:

$$\log L(\hat{p}) = N_1 \log(N_1/N) + (N - N_1)\log(1 - N_1/N)$$

$$\log L(\tilde{p}) = N_1 \log(p_0) + (N - N_1)\log(1 - p_0)$$

$$LR = 2(\log L(\hat{p}) - \log L(\tilde{p}))$$

Verbeek Exercise 6.2:

a.

$$L(\beta_1, \beta_2) = \prod_{i=1}^{N} \frac{e^{-e^{\beta_1+\beta_2 x_i}} {e^{\beta_1+\beta_2 x_i}}^{y_i}}{y_i!}$$

$$\log L(\beta_1, \beta_2) = \sum_{i=1}^{N} \left[-e^{\beta_1+\beta_2 x_i} + y_i(\beta_1 + \beta_2 x_i) - \log(y_i!)\right]$$

b.

$$\frac{\partial \log L}{\partial \beta_1} = \sum_{i=1}^{N} \left(y_i - e^{\beta_1+\beta_2 x_i}\right) = \sum_{i=1}^{N} (y_i - \lambda_i)$$

$$\frac{\partial \log L}{\partial \beta_2} = \sum_{i=1}^{N} \left(y_i - e^{\beta_1+\beta_2 x_i}\right) x_i = \sum_{i=1}^{N} (y_i - \lambda_i) x_i$$

$$s(\beta) = \begin{pmatrix} \sum_{i=1}^{N} (y_i - \lambda_i) \\ \sum_{i=1}^{N} (y_i - \lambda_i) x_i \end{pmatrix}$$

$$s_i(\beta) = \begin{pmatrix} y_i - \lambda_i \\ (y_i - \lambda_i) x_i \end{pmatrix}$$

given $E[y_i|x_i] = \lambda_i$:

$$E[y_i - \lambda_i|x_i] = 0$$

$$E[(y_i - \lambda_i)x_i|x_i] = x_i E[y_i - \lambda_i|x_i] = 0$$

$$E[s_i(\beta)|x_i] = E[s_i(\beta)] = 0$$

c. $\frac{\partial^2 \log L}{\partial \beta_1^2} = \sum_{i=1}^N (-\lambda_i)$, $\frac{\partial^2 \log L}{\partial \beta_2^2} = \sum_{i=1}^N (-\lambda_i x_i^2)$, $\frac{\partial^2 \log L}{\partial \beta_1 \partial \beta_2} = \sum_{i=1}^N (-\lambda_i x_i)$

$$H(\beta) = -\begin{pmatrix} \sum \lambda_i & \sum \lambda_i x_i \\ \sum \lambda_i x_i & \sum \lambda_i x_i^2 \end{pmatrix}$$

$$I(\beta) = -E[H(\beta)] = \begin{pmatrix} \sum E[\lambda_i] & \sum E[\lambda_i x_i] \\ \sum E[\lambda_i x_i] & \sum E[\lambda_i x_i^2] \end{pmatrix} = \begin{pmatrix} \sum \lambda_i & \sum \lambda_i x_i \\ \sum \lambda_i x_i & \sum \lambda_i x_i^2 \end{pmatrix}$$

the asymptotic covariance matrix: $\mathrm{Var}(\beta) = I(\beta)^{-1}$
consistent estimator: $\widehat{\mathrm{Var}}(\hat{\beta}) = \hat{I}(\hat{\beta})^{-1}$, $\hat{\lambda}_i = \exp(\hat{\beta}_1 + \hat{\beta}_2 x_i)$