Stat 710: Mathematical Statistics
Lecture 28

Jun Shao

Department of Statistics
University of Wisconsin
Madison, WI 53706, USA

Asymptotic distribution of likelihood ratio

An LR test is often equivalent to a test based on a statistic $Y(X)$ whose distribution under H_0 can be used to determine the rejection region of the LR test with size α.

When this technique fails, it is difficult or even impossible to find an LR test with size α, even if the c.d.f. of $\lambda(X)$ is continuous.

In the i.i.d. case we can obtain the asymptotic distribution (under H_0) of the likelihood ratio $\lambda(X)$ so that an LR test having asymptotic significance level α can be obtained.

In many problems Θ_0 is determined by

$$H_0 : \theta = g(\vartheta),$$

where ϑ is a $(k-r)$-vector of unknown parameters and g is a continuously differentiable function from $\mathscr{R}^{k-r}$ to $\mathscr{R}^k$ with a full rank $\partial g(\vartheta)/\partial \vartheta$.

For example, if $\Theta = \mathscr{R}^2$ and $\Theta_0 = \{(\theta_1, \theta_2) \in \Theta : \theta_1 = 0\}$, then $\vartheta = \theta_2$, $g_1(\vartheta) = 0$, and $g_2(\vartheta) = \vartheta$.

Theorem 6.5

Assume the conditions in Theorem 4.16.
Suppose that $H_0 : \theta = g(\vartheta)$, where ϑ is a $(k-r)$-vector of unknown parameters and g is a continuously differentiable function from $\mathscr{R}^{k-r}$ to $\mathscr{R}^k$ with a full rank $\partial g(\vartheta)/\partial\vartheta$.
Under H_0, $-2\log\lambda_n \to_d \chi^2_r$, where $\lambda_n = \lambda(X)$ and χ^2_r is a random variable having the chi-square distribution χ^2_r.
Consequently, the LR test with rejection region $\lambda_n < e^{-\chi^2_{r,\alpha}/2}$ has asymptotic significance level α, where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r.

Proof

Without loss of generality, we assume that there exist an MLE $\widehat{\theta}$ and an MLE $\widehat{\vartheta}$ under H_0 such that

$$\lambda_n = \sup_{\theta\in\Theta_0} \ell(\theta) \Big/ \sup_{\theta\in\Theta} \ell(\theta) = \ell(g(\widehat{\vartheta}))/\ell(\widehat{\theta}).$$

Let $s_n(\theta) = \partial\log\ell(\theta)/\partial\theta$ and $I_1(\theta)$ be the Fisher information about θ contained in X_1.

Theorem 6.5

Assume the conditions in Theorem 4.16.
Suppose that $H_0 : \theta = g(\vartheta)$, where ϑ is a $(k-r)$-vector of unknown parameters and g is a continuously differentiable function from $\mathscr{R}^{k-r}$ to $\mathscr{R}^k$ with a full rank $\partial g(\vartheta)/\partial\vartheta$.
Under H_0, $-2\log\lambda_n \to_d \chi^2_r$, where $\lambda_n = \lambda(X)$ and χ^2_r is a random variable having the chi-square distribution χ^2_r.
Consequently, the LR test with rejection region $\lambda_n < e^{-\chi^2_{r,\alpha}/2}$ has asymptotic significance level α, where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r.

Proof

Without loss of generality, we assume that there exist an MLE $\widehat{\theta}$ and an MLE $\widehat{\vartheta}$ under H_0 such that

$$\lambda_n = \sup_{\theta\in\Theta_0} \ell(\theta) \Big/ \sup_{\theta\in\Theta} \ell(\theta) = \ell(g(\widehat{\vartheta}))/\ell(\widehat{\theta}).$$

Let $s_n(\theta) = \partial \log \ell(\theta)/\partial\theta$ and $I_1(\theta)$ be the Fisher information about θ contained in X_1.

Proof (continued)

Following the proof of Theorem 4.17 in §4.5.2, we can obtain that

$$\sqrt{n}I_1(\theta)(\widehat{\theta}-\theta) = n^{-1/2}s_n(\theta) + o_p(1),$$

and

$$2[\log\ell(\widehat{\theta}) - \log\ell(\theta)] = n(\widehat{\theta}-\theta)^\tau I_1(\theta)(\widehat{\theta}-\theta) + o_p(1).$$

Then

$$2[\log\ell(\widehat{\theta}) - \log\ell(\theta)] = n^{-1}[s_n(\theta)]^\tau [I_1(\theta)]^{-1} s_n(\theta) + o_p(1).$$

Similarly, under H_0,

$$2[\log\ell(g(\widehat{\vartheta})) - \log\ell(g(\vartheta))] = n^{-1}[\tilde{s}_n(\vartheta)]^\tau [\tilde{I}_1(\vartheta)]^{-1} \tilde{s}_n(\vartheta) + o_p(1),$$

where $\tilde{s}_n(\vartheta) = \partial \log\ell(g(\vartheta))/\partial\vartheta = D(\vartheta)s_n(g(\vartheta))$, $D(\vartheta) = \partial g(\vartheta)/\partial\vartheta$, and $\tilde{I}_1(\vartheta)$ is the Fisher information about ϑ (under H_0) contained in X_1.
Combining these results, we obtain that, under H_0,

$$\begin{aligned} -2\log\lambda_n &= 2[\log\ell(\widehat{\theta}) - \log\ell(g(\widehat{\vartheta}))] \\ &= n^{-1}[s_n(g(\vartheta))]^\tau B(\vartheta) s_n(g(\vartheta)) + o_p(1) \end{aligned}$$

where $B(\vartheta) = [I_1(g(\vartheta))]^{-1} - [D(\vartheta)]^\tau [\tilde{I}_1(\vartheta)]^{-1} D(\vartheta)$.

Proof (continued)

By the CLT, $n^{-1/2}[I_1(\theta)]^{-1/2}s_n(\theta) \to_d Z$, where $Z = N_k(0, I_k)$.
Then, it follows from Theorem 1.10(iii) that, under H_0,

$$-2\log\lambda_n \to_d Z^\tau[I_1(g(\vartheta))]^{1/2}B(\vartheta)[I_1(g(\vartheta))]^{1/2}Z.$$

Let $D = D(\vartheta)$, $B = B(\vartheta)$, $A = I_1(g(\vartheta))$, and $C = \tilde{I}_1(\vartheta)$.
Then

$$\begin{aligned}
(A^{1/2}BA^{1/2})^2 &= A^{1/2}BABA^{1/2} \\
&= A^{1/2}(A^{-1} - D^\tau C^{-1}D)A(A^{-1} - D^\tau C^{-1}D)A^{1/2} \\
&= (I_k - A^{1/2}D^\tau C^{-1}DA^{1/2})(I_k - A^{1/2}D^\tau C^{-1}DA^{1/2}) \\
&= I_k - 2A^{1/2}D^\tau C^{-1}DA^{1/2} + A^{1/2}D^\tau C^{-1}DAD^\tau C^{-1}DA^{1/2} \\
&= I_k - A^{1/2}D^\tau C^{-1}DA^{1/2} \\
&= A^{1/2}BA^{1/2},
\end{aligned}$$

where the fourth equality follows from the fact that $C = DAD^\tau$.
This shows that $A^{1/2}BA^{1/2}$ is a projection matrix.

Proof (continued)

The rank of $A^{1/2}BA^{1/2}$ is

$$\begin{aligned} \mathrm{tr}(A^{1/2}BA^{1/2}) &= \mathrm{tr}(I_k - D^\tau C^{-1}DA) \\ &= k - \mathrm{tr}(C^{-1}DAD^\tau) \\ &= k - \mathrm{tr}(C^{-1}C) \\ &= k - (k-r) \\ &= r. \end{aligned}$$

Thus, by Exercise 51 in §1.6, $Z^\tau[I_1(g(\vartheta))]^{1/2}B(\vartheta)[I_1(g(\vartheta))]^{1/2}Z = \chi_r^2$.

Asymptotic tests

Tests whose rejection regions are constructed using asymptotic theory (so that these tests have asymptotic significance level α) are called *asymptotic tests*, which are useful when a test of exact size α is difficult to find.

The LR test in Theorem 6.5 is one example of an asymptotic test.

Proof (continued)

The rank of $A^{1/2}BA^{1/2}$ is

$$\begin{aligned} \mathrm{tr}(A^{1/2}BA^{1/2}) &= \mathrm{tr}(I_k - D^\tau C^{-1}DA) \\ &= k - \mathrm{tr}(C^{-1}DAD^\tau) \\ &= k - \mathrm{tr}(C^{-1}C) \\ &= k - (k-r) \\ &= r. \end{aligned}$$

Thus, by Exercise 51 in §1.6, $Z^\tau[I_1(g(\vartheta))]^{1/2}B(\vartheta)[I_1(g(\vartheta))]^{1/2}Z = \chi^2_r$.

Asymptotic tests

Tests whose rejection regions are constructed using asymptotic theory (so that these tests have asymptotic significance level α) are called *asymptotic tests*, which are useful when a test of exact size α is difficult to find.

The LR test in Theorem 6.5 is one example of an asymptotic test.

Asymptotic tests

There are two popular asymptotic tests based on likelihoods that are asymptotically equivalent to LR tests.

The hypothesis $H_0 : \theta = g(\vartheta)$ is equivalent to a set of $r \le k$ equations:

$$H_0 : R(\theta) = 0,$$

where $R(\theta)$ is a continuously differentiable function from $\mathscr{R}^k$ to $\mathscr{R}^r$.

Wald (1943) introduced a test that rejects H_0 when the value of

$$W_n = [R(\widehat{\theta})]^\tau \{[C(\widehat{\theta})]^\tau [I_n(\widehat{\theta})]^{-1} C(\widehat{\theta})\}^{-1} R(\widehat{\theta})$$

is large, where $C(\theta) = \partial R(\theta)/\partial \theta$, $I_n(\theta)$ is the Fisher information matrix based on $X_1, ..., X_n$, and $\widehat{\theta}$ is an MLE or RLE of θ.

For testing $H_0 : \theta = \theta_0$ with a known θ_0, $R(\theta) = \theta - \theta_0$ and

$$W_n = (\widehat{\theta} - \theta_0)^\tau I_n(\widehat{\theta})(\widehat{\theta} - \theta_0).$$

Rao (1947) introduced a *score* test that rejects H_0 when the value of

$$R_n = [s_n(\tilde{\theta})]^\tau [I_n(\tilde{\theta})]^{-1} s_n(\tilde{\theta})$$

is large, where $s_n(\theta) = \partial \log \ell(\theta)/\partial \theta$ is the score function and $\tilde{\theta}$ is an MLE or RLE of θ under $H_0 : R(\theta) = 0$.

Theorem 6.6

Assume the conditions in Theorem 4.16.
(i) Under $H_0 : R(\theta) = 0$, where $R(\theta)$ is a continuously differentiable function from $\mathscr{R}^k$ to $\mathscr{R}^r$, $W_n \to_d \chi^2_r$ and, therefore, the test rejects H_0 if and only if $W_n > \chi^2_{r,\alpha}$ has asymptotic significance level α, where $\chi^2_{r,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_r.
(ii) The result in (i) still holds if W_n is replaced by R_n.

Remarks

- Wald's test, Rao's score test, and the LR test are asymptotically equivalent.
- Wald's test requires computing $\widehat{\theta}$, not $\tilde{\theta} = g(\widehat{\vartheta})$.
- Rao's score test requires computing $\tilde{\theta}$, not $\widehat{\theta}$.
- The LR test requires computing both $\widehat{\theta}$ and $\tilde{\theta}$ (or solving two maximization problems), but it may be more efficient.
- Hence, one may choose one of these tests in terms of computation and efficiency in a particular application.

Theorem 6.6

Assume the conditions in Theorem 4.16.
(i) Under $H_0 : R(\theta) = 0$, where $R(\theta)$ is a continuously differentiable function from $\mathscr{R}^k$ to $\mathscr{R}^r$, $W_n \to_d \chi_r^2$ and, therefore, the test rejects H_0 if and only if $W_n > \chi_{r,\alpha}^2$ has asymptotic significance level α, where $\chi_{r,\alpha}^2$ is the $(1-\alpha)$th quantile of the chi-square distribution χ_r^2.
(ii) The result in (i) still holds if W_n is replaced by R_n.

Remarks

- Wald's test, Rao's score test, and the LR test are asymptotically equivalent.
- Wald's test requires computing $\widehat{\theta}$, not $\tilde{\theta} = g(\widehat{\vartheta})$.
- Rao's score test requires computing $\tilde{\theta}$, not $\widehat{\theta}$.
- The LR test requires computing both $\widehat{\theta}$ and $\tilde{\theta}$ (or solving two maximization problems), but it may be more efficient.
- Hence, one may choose one of these tests in terms of computation and efficiency in a particular application.

Proof

(i) Using Theorems 1.12 and 4.17,

$$\sqrt{n}[R(\widehat{\theta}) - R(\theta)] \to_d N_r\left(0, [C(\theta)]^\tau [I_1(\theta)]^{-1} C(\theta)\right),$$

where $I_1(\theta)$ is the Fisher information about θ contained in X_1.
Under H_0, $R(\theta) = 0$ and, therefore (by Theorem 1.10),

$$n[R(\widehat{\theta})]^\tau \{[C(\theta)]^\tau [I_1(\theta)]^{-1} C(\theta)\}^{-1} R(\widehat{\theta}) \to_d \chi_r^2$$

Then the result follows from Slutsky's theorem (Theorem 1.11) and the fact that $\widehat{\theta} \to_p \theta$ and $I_1(\theta)$ and $C(\theta)$ are continuous at θ.
(ii) From the Lagrange multiplier, $\tilde{\theta}$ satisfies

$$s_n(\tilde{\theta}) + C(\tilde{\theta})\lambda_n = 0 \qquad \text{and} \qquad R(\tilde{\theta}) = 0.$$

Using Taylor's expansion, one can show (exercise) that under H_0,

$$[C(\theta)]^\tau (\tilde{\theta} - \theta) = o_p(n^{-1/2}) \qquad (1)$$

and

$$s_n(\theta) - I_n(\theta)(\tilde{\theta} - \theta) + C(\theta)\lambda_n = o_p(n^{1/2}), \qquad (2)$$

where $I_n(\theta) = nI_1(\theta)$.

Proof (continued)

Multiplying $[C(\theta)]^\tau [I_n(\theta)]^{-1}$ to the left-hand side of (2) and using (1), we obtain that

$$[C(\theta)]^\tau [I_n(\theta)]^{-1} C(\theta)\lambda_n = -[C(\theta)]^\tau [I_n(\theta)]^{-1} s_n(\theta) + o_p(n^{-1/2}),$$

which implies

$$\lambda_n^\tau [C(\theta)]^\tau [I_n(\theta)]^{-1} C(\theta)\lambda_n \to_d \chi_r^2 \tag{3}$$

(exercise).

Then the result follows from (3) and the fact that $C(\tilde\theta)\lambda_n = -s_n(\tilde\theta)$, $I_n(\theta) = nI_1(\theta)$, and $I_1(\theta)$ is continuous at θ.