Likelihood ratio

When both H_0 and H_1 are simple (i.e., $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta_1\}$), Theorem 6.1 applies and a UMP test rejects H_0 when

$$\frac{f_{\theta_1}(X)}{f_{\theta_0}(X)} > c_0$$

for some $c_0 > 0$.
The following definition is a natural extension of this idea.

Definition 6.2

Let $\ell(\theta) = f_\theta(X)$ be the likelihood function. For testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$, a *likelihood ratio* (LR) test is any test that rejects H_0 if and only if $\lambda(X) < c$, where $c \in [0,1]$ and $\lambda(X)$ is the likelihood ratio defined by

$$\lambda(X) = \sup_{\theta \in \Theta_0} \ell(\theta) \Big/ \sup_{\theta \in \Theta} \ell(\theta).$$

Discussions

If $\lambda(X)$ is well defined, then $\lambda(X) \leq 1$.

The rationale behind LR tests is that when H_0 is true, $\lambda(X)$ tends to be close to 1, whereas when H_1 is true, $\lambda(X)$ tends to be away from 1.

If there is a sufficient statistic, then $\lambda(X)$ depends only on the sufficient statistic.

LR tests are as widely applicable as MLE's in §4.4 and, in fact, they are closely related to MLE's.

If $\widehat{\theta}$ is an MLE of θ and $\widehat{\theta}_0$ is an MLE of θ subject to $\theta \in \Theta_0$ (i.e., Θ_0 is treated as the parameter space), then

$$\lambda(X) = \ell(\widehat{\theta}_0)/\ell(\widehat{\theta}).$$

For a given $\alpha \in (0,1)$, if there exists a $c_\alpha \in [0,1]$ such that

$$\sup_{\theta \in \Theta_0} P_\theta(\lambda(X) < c_\alpha) = \alpha,$$

then an LR test of size α can be obtained.

Even when the c.d.f. of $\lambda(X)$ is continuous or randomized LR tests are introduced, it is still possible that such a c_α does not exist.

Optimality

When a UMP or UMPU test exists, an LR test is often the same as this optimal test.

Proposition 6.5

Suppose that X has a p.d.f. in a one-parameter exponential family:

$$f_\theta(x) = \exp\{\eta(\theta)Y(x) - \xi(\theta)\}h(x)$$

w.r.t. a σ-finite measure ν, where η is a strictly increasing and differentaible function of θ.

(i) For testing $H_0 : \theta \le \theta_0$ versus $H_1 : \theta > \theta_0$, there is an LR test whose rejection region is the same as that of the UMP test T_* given in Theorem 6.2.

(ii) For testing $H_0 : \theta \le \theta_1$ or $\theta \ge \theta_2$ versus $H_1 : \theta_1 < \theta < \theta_2$, there is an LR test whose rejection region is the same as that of the UMP test T_* given in Theorem 6.3.

(iii) For testing the other two-sided hypotheses, there is an LR test whose rejection region is equivalent to $Y(X) < c_1$ or $Y(X) > c_2$ for some constants c_1 and c_2.

Proof

We prove (i) only.

Let $\widehat{\theta}$ be the MLE of θ.

Note that $\ell(\theta)$ is increasing when $\theta \le \widehat{\theta}$ and decreasing when $\theta > \widehat{\theta}$.

Thus,

$$\lambda(X) = \begin{cases} 1 & \widehat{\theta} \le \theta_0 \\ \frac{\ell(\theta_0)}{\ell(\widehat{\theta})} & \widehat{\theta} > \theta_0. \end{cases}$$

Then $\lambda(X) < c$ is the same as $\widehat{\theta} > \theta_0$ and $\ell(\theta_0)/\ell(\widehat{\theta}) < c$.

From the property of exponential families, $\widehat{\theta}$ is a solution of the likelihood equation

$$\frac{\partial \log \ell(\theta)}{\partial \theta} = \eta'(\theta)Y(X) - \xi'(\theta) = 0$$

and $\psi(\theta) = \xi'(\theta)/\eta'(\theta)$ has a positive derivative $\psi'(\theta)$.

Since $\eta'(\widehat{\theta})Y - \xi'(\widehat{\theta}) = 0$, $\widehat{\theta}$ is an increasing function of Y and $\frac{d\widehat{\theta}}{dY} > 0$.
Consequently, for any $\theta_0 \in \Theta$,

$$\begin{aligned}\frac{d}{dY}\left[\log \ell(\widehat{\theta})-\log \ell(\theta_0)\right] &= \frac{d}{dY}\left[\eta(\widehat{\theta})Y-\xi(\widehat{\theta})-\eta(\theta_0)Y+\xi(\theta_0)\right] \\ &= \frac{d\widehat{\theta}}{dY}\eta'(\widehat{\theta})Y+\eta(\widehat{\theta})-\frac{d\widehat{\theta}}{dY}\xi'(\widehat{\theta})-\eta(\theta_0) \\ &= \frac{d\widehat{\theta}}{dY}[\eta'(\widehat{\theta})Y-\xi'(\widehat{\theta})]+\eta(\widehat{\theta})-\eta(\theta_0) \\ &= \eta(\widehat{\theta})-\eta(\theta_0),\end{aligned}$$

which is positive (or negative) if $\widehat{\theta} > \theta_0$ (or $\widehat{\theta} < \theta_0$), i.e., $\log \ell(\widehat{\theta}) - \log \ell(\theta_0)$ is strictly increasing in Y when $\widehat{\theta} > \theta_0$ and strictly decreasing in Y when $\widehat{\theta} < \theta_0$.

Hence, for any $d \in \mathscr{R}$, $\widehat{\theta} > \theta_0$ and $\ell(\theta_0)/\ell(\widehat{\theta}) < c$ is equivalent to $Y > d$ for some $c \in (0,1)$.

Example 6.20

Consider the testing problem $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ based on i.i.d. $X_1,...,X_n$ from the uniform distribution $U(0,\theta)$.

We now show that the UMP test with rejection region $X_{(n)} > \theta_0$ or $X_{(n)} \leq \theta_0 \alpha^{1/n}$ given in Exercise 19(c) is an LR test.

Note that $\ell(\theta) = \theta^{-n} I_{(X_{(n)},\infty)}(\theta)$.
Hence

$$\lambda(X) = \begin{cases} (X_{(n)}/\theta_0)^n & X_{(n)} \le \theta_0 \\ 0 & X_{(n)} > \theta_0 \end{cases}$$

and $\lambda(X) < c$ is equivalent to $X_{(n)} > \theta_0$ or $X_{(n)}/\theta_0 < c^{1/n}$.
Taking $c = \alpha$ ensures that the LR test has size α.

Example 6.21

Consider normal linear model $X = N_n(Z\beta, \sigma^2 I_n)$ and the hypotheses

$$H_0 : L\beta = 0 \qquad \text{versus} \qquad H_1 : L\beta \neq 0,$$

where L is an $s \times p$ matrix of rank $s \le r$ and all rows of L are in $\mathscr{R}(Z)$.
The likelihood function in this problem is

$$\ell(\theta) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\|X - Z\beta\|^2\right\}, \quad \theta = (\beta, \sigma^2).$$

Since $\|X - Z\beta\|^2 \ge \|X - Z\widehat{\beta}\|^2$ for any β and the LSE $\widehat{\beta}$,

$$\ell(\theta) \le \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma^2}\|X - Z\widehat{\beta}\|^2\right\}.$$

Treating the right-hand side of this expression as a function of σ^2, it is easy to show that it has a maximum at $\sigma^2 = \widehat{\sigma}^2 = \|X - Z\widehat{\beta}\|^2/n$ and

$$\sup_{\theta\in\Theta} \ell(\theta) = (2\pi\widehat{\sigma}^2)^{-n/2} e^{-n/2}.$$

Similarly, let $\widehat{\beta}_{H_0}$ be the LSE under H_0 and $\widehat{\sigma}^2_{H_0} = \|X - Z\widehat{\beta}_{H_0}\|^2/n$:

$$\sup_{\theta\in\Theta_0} \ell(\theta) = (2\pi\widehat{\sigma}^2_{H_0})^{-n/2} e^{-n/2}.$$

Thus,

$$\lambda(X) = (\widehat{\sigma}^2/\widehat{\sigma}^2_{H_0})^{n/2} = \left(\frac{\|X - Z\widehat{\beta}\|^2}{\|X - Z\widehat{\beta}_{H_0}\|^2}\right)^{n/2}.$$

For a two-sample problem, we let $n = n_1 + n_2$, $\beta = (\mu_1, \mu_2)$, and

$$Z = \begin{pmatrix} J_{n_1} & 0 \\ 0 & J_{n_2} \end{pmatrix}.$$

Testing $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ is the same as testing $H_0 : L\beta = 0$ versus $H_1 : L\beta \neq 0$ with $L = \begin{pmatrix} 1 & -1 \end{pmatrix}$.
The LR test is the same as the two-sample two-sided t-tests in §6.2.3.

Example: Exercise 6.84

Let F and G be two known cumulative distribution functions on $\mathscr{R}$ and X be a single observation from the cumulative distribution function $\theta F(x) + (1-\theta)G(x)$, where $\theta \in [0,1]$ is unknown.
We first derive the likelihood ratio $\lambda(X)$ for testing

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$$

where $\theta_0 \in [0,1)$ is a known constant.
Let f and g be the probability densities of F and G, respectively, with respect to the measure corresponding to $F + G$.
Then, the likelihood function is

$$\ell(\theta) = \theta[f(X) - g(X)] + g(X)$$

and

$$\sup_{0\leq\theta\leq 1} \ell(\theta) = \begin{cases} f(X) & f(X) \geq g(X) \\ g(X) & f(X) < g(X). \end{cases}$$

For $\theta_0 \in [0,1)$,

$$\sup_{0\leq\theta\leq\theta_0} \ell(\theta) = \begin{cases} \theta_0[f(X) - g(X)] + g(X) & f(X) \geq g(X) \\ g(X) & f(X) < g(X). \end{cases}$$

Hence,

$$\lambda(X) = \begin{cases} \frac{\theta_0[f(X)-g(X)]+g(X)}{f(X)} & f(X) \geq g(X) \\ 1 & f(X) < g(X). \end{cases}$$

Choose a constant c with $\theta_0 \leq c < 1$.
Then $\lambda(X) \leq c$ is the same as

$$\frac{g(X)}{f(X)} \leq \frac{c-\theta_0}{1-\theta_0}$$

We may find a c with $P(\lambda(X) \leq c) = \alpha$ when $\theta = \theta_0$.
Consider next

$$H_0 : \theta_1 \leq \theta \leq \theta_2 \quad \text{versus} \quad H_1 : \theta < \theta_1 \;\text{ or }\; \theta > \theta_2$$

where $0 \leq \theta_1 < \theta_2 \leq 1$ are known constants.
For $0 \leq \theta_1 \leq \theta_2 \leq 1$,

$$\sup_{0\leq\theta_1\leq\theta\leq\theta_2\leq1} \ell(\theta) = \begin{cases} \theta_2[f(X)-g(X)]+g(X) & f(X) \geq g(X) \\ \theta_1[f(X)-g(X)]+g(X) & f(X) < g(X) \end{cases}$$

Hence,

$$\lambda(X) = \begin{cases} \frac{\theta_2[f(X)-g(X)]+g(X)}{f(X)} & f(X) \geq g(X) \\ \frac{\theta_1[f(X)-g(X)]+g(X)}{g(X)} & f(X) < g(X). \end{cases}$$

Choose a constant c with $\max\{1-\theta_1, \theta_2\} \le c < 1$.

Then $\lambda(X) \le c$ is the same as

$$\frac{g(X)}{f(X)} \le \frac{c-\theta_0}{1-\theta_0} \quad \text{or} \quad \frac{g(X)}{f(X)} \ge \frac{\theta_1}{c-(1-\theta_1)}.$$

How to find a c with $\sup_{\theta_1 \le \theta \le \theta_2} P(\lambda(X) \le c) = \alpha$?

Finally, consider

$$H_0 : \theta \le \theta_1 \ \text{ or } \ \theta \ge \theta_2 \quad \text{versus} \quad \theta_1 < \theta < \theta_2$$

where $0 \le \theta_1 \le \theta_2 \le 1$ are known constants.

Note that

$$\sup_{0 \le \theta \le \theta_1, \theta_2 \le \theta \le 1} \ell(\theta) = \sup_{0 \le \theta \le 1} \ell(\theta).$$

Hence,

$$\lambda(X) = 1$$

This means that, unless we consider randomizing, we cannot find a c such that $\sup_{\theta \le \theta_1 \text{ or } \theta \ge \theta_2} P(\lambda(X) \le c) = \alpha$.

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

Definition 2.13 (asymptotic tests)

Let $X = (X_1, ..., X_n)$ be a sample from $P \in \mathscr{P}$ and $T_n(X)$ be a test for $H_0 : P \in \mathscr{P}_0$ versus $H_1 : P \in \mathscr{P}_1$.

(i) If $\limsup_n \alpha_{T_n}(P) \leq \alpha$ for any $P \in \mathscr{P}_0$, then α is an *asymptotic significance level* of T_n.

(ii) If $\lim_{n\to\infty} \sup_{P\in\mathscr{P}_0} \alpha_{T_n}(P)$ exists, it is called the *limiting size* of T_n.

(iii) T_n is *consistent* iff the type II error probability converges to 0.

- If $\mathscr{P}_0$ is not a parametric family, the limiting size of T_n may be 1. This is the reason why we consider the weaker requirement in (i).
- If $\alpha \in (0, 1)$ is a pre-assigned level of significance for the problem, then a consistent test T_n having asymptotic significance level α is called *asymptotically correct*, and a consistent test having limiting size α is called *strongly* asymptotically correct.

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.

Theorem 6.5 (asymptotic distribution of likelihood ratio)

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.

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

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

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

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.
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 based on likelihoods

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.

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

- Wald's test requires computing $\widehat{\theta}$, not $\tilde{\theta} = g(\widehat{\vartheta})$.
- Rao's score test requires computing $\tilde{\theta}$, not $\widehat{\theta}$.

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.

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),

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

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) See the textbook.