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

$$\lambda(X) = \sup_{\theta \in \Theta_0} \ell(\theta) \div \sup_{\theta \in \Theta} \ell(\theta).$$
Lecture 26: Likelihood ratio tests

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

$$\lambda(X) = \sup_{\theta \in \Theta_0} \ell(\theta) / \sup_{\theta \in \Theta} \ell(\theta).$$
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 $\hat{\theta}$ is an MLE of $\theta$ and $\hat{\theta}_0$ is an MLE of $\theta$ subject to $\theta \in \Theta_0$ (i.e., $\Theta_0$ is treated as the parameter space), then

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

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 $\alpha$ 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 $c_\alpha$ 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 $\sigma$-finite measure $\nu$, where $\eta$ is a strictly increasing and differentiable function of $\theta$.

(i) For testing $H_0 : \theta \leq \theta_0$ 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 $H_0 : \theta \leq \theta_1$ or $\theta \geq \theta_2$ 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 $c_2$. 
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 $\sigma$-finite measure $\nu$, where $\eta$ is a strictly increasing and differentiable function of $\theta$.

(i) For testing $H_0: \theta \leq \theta_0$ 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 $H_0: \theta \leq \theta_1$ or $\theta \geq \theta_2$ 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 $\hat{\theta}$ be the MLE of $\theta$.

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

Thus,

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

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

From the property of exponential families, $\hat{\theta}$ 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 $\eta'(\hat{\theta}) Y - \xi'(\hat{\theta}) = 0$, $\hat{\theta}$ is an increasing function of $Y$ and $\frac{d\hat{\theta}}{dY} > 0$. 
Consequently, for any $\theta_0 \in \Theta$,
\[
\frac{d}{dY} \left[ \log \ell(\hat{\theta}) - \log \ell(\theta_0) \right] = \frac{d}{dY} \left[ \eta(\hat{\theta}) Y - \xi(\hat{\theta}) - \eta(\theta_0) Y + \xi(\theta_0) \right]
= \frac{d\hat{\theta}}{dY} \eta'(\hat{\theta}) Y + \eta(\hat{\theta}) - \frac{d\hat{\theta}}{dY} \xi'(\hat{\theta}) - \eta(\theta_0)
= \frac{d\hat{\theta}}{dY} [\eta'(\hat{\theta}) Y - \xi'(\hat{\theta})] + \eta(\hat{\theta}) - \eta(\theta_0)
= \eta(\hat{\theta}) - \eta(\theta_0),
\]
which is positive (or negative) if $\hat{\theta} > \theta_0$ (or $\hat{\theta} < \theta_0$), i.e.,
\[
\log \ell(\hat{\theta}) - \log \ell(\theta_0)
\]
is strictly increasing in $Y$ when $\hat{\theta} > \theta_0$ and strictly decreasing in $Y$ when $\hat{\theta} < \theta_0$.

Hence, for any $d \in \mathbb{R}$, $\hat{\theta} > \theta_0$ and $\ell(\theta_0)/\ell(\hat{\theta}) < c$ 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. \( X_1, \ldots, X_n \) 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
\[
\lambda(X) = \begin{cases} 
(X_{(n)}/\theta_0)^n & X_{(n)} \leq \theta_0 \\
0 & X_{(n)} > \theta_0 
\end{cases}
\]
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 \( \alpha \).

Example 6.21
Consider normal linear model \( X = N_n(Z\beta, \sigma^2 I_n) \) and the hypotheses
\[
H_0 : L\beta = 0 \quad \text{versus} \quad H_1 : L\beta \neq 0,
\]
where \( L \) is an \( s \times p \) matrix of rank \( s \leq r \) and all rows of \( L \) are in \( \mathcal{R}(Z) \).
Example 6.20
Consider the testing problem \( H_0 : \theta = \theta_0 \) versus \( H_1 : \theta \neq \theta_0 \) based on i.i.d. \( X_1, \ldots, X_n \) 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 \( \ell(\theta) = \theta^{-n} l_{(X_{(n)}, \infty)}(\theta) \).
Hence
\[
\lambda(X) = \begin{cases} 
(X_{(n)}/\theta_0)^n & X_{(n)} \leq \theta_0 \\
0 & X_{(n)} > \theta_0 
\end{cases}
\]
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 \( \alpha \).

Example 6.21
Consider normal linear model \( X = N_n(Z\beta, \sigma^2 I_n) \) and the hypotheses
\[
H_0 : L\beta = 0 \quad \text{versus} \quad H_1 : L\beta \neq 0,
\]
where \( L \) is an \( s \times p \) matrix of rank \( s \leq r \) and all rows of \( L \) are in \( \mathcal{R}(Z) \).
Example 6.21 (continued)

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 $\|X - Z\beta\|^2 \geq \|X - Z\hat{\beta}\|^2$ for any $\beta$ and the LSE $\hat{\beta}$,

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

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

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

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

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

Thus,

$$\lambda(X) = \left(\frac{\hat{\sigma}^2}{\hat{\sigma}^2_{H_0}}\right)^{n/2} = \left(\frac{\|X - Z\hat{\beta}\|^2}{\|X - Z\hat{\beta}_{H_0}\|^2} \right)^{n/2}.$$
Example 6.21 (continued)

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 $L = (1 \hphantom{1} -1\hphantom{1})$.

Since $\hat{\beta}_{H_0} = \bar{X}$ and $\hat{\beta} = (\bar{X}_1, \bar{X}_2)$, where $\bar{X}_1$ and $\bar{X}_2$ are the sample means based on $X_1, \ldots, X_{n_1}$ and $X_{n_1+1}, \ldots, X_n$, respectively, we have

$$n\hat{\sigma}^2 = \sum_{i=1}^{n_1} (X_i - \bar{X}_1)^2 + \sum_{i=n_1+1}^{n} (X_i - \bar{X}_2)^2 = (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2$$

and

$$n\hat{\sigma}^2_{H_0} = (n - 1)S^2 = n^{-1}n_1n_2(\bar{X}_1 - \bar{X}_2)^2 + (n_1 - 1)S_1^2 + (n_2 - 1)S_2^2.$$
Example 6.21 (continued)

Therefore, $\lambda(X) < c$ is equivalent to $|t(X)| > c_0$, where

$$t(X) = \frac{(\bar{X}_2 - \bar{X}_1)/\sqrt{\frac{n_1^{-1} + n_2^{-1}}{\sqrt{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)}}}},$$

and LR tests are the same as the two-sample two-sided t-tests in §6.2.3.

Asymptotic tests

It is often difficult to construct tests (such as LR tests) with exactly size $\alpha$ or level $\alpha$.

Asymptotic approximation can be used.

Statistical inference based on asymptotic criteria and approximations is called asymptotic statistical inference or simply asymptotic inference.

We now focus on asymptotic hypothesis tests.
Example 6.21 (continued)

Therefore, $\lambda(X) < c$ is equivalent to $|t(X)| > c_0$, where

$$t(X) = \frac{(\bar{X}_2 - \bar{X}_1)/\sqrt{n_1^{-1} + n_2^{-1}}}{\sqrt{[(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2]/(n_1 + n_2 - 2)}}$$

and LR tests are the same as the two-sample two-sided t-tests in §6.2.3.

Asymptotic tests

It is often difficult to construct tests (such as LR tests) with exactly size $\alpha$ or level $\alpha$.

Asymptotic approximation can be used

Statistical inference based on asymptotic criteria and approximations is called *asymptotic statistical inference* or simply *asymptotic inference*.

We now focus on asymptotic hypothesis tests.
Definition 2.13

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

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

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

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

Discussion

- If $\mathcal{P}_0$ is not a parametric family, it is likely that the limiting size of $T_n$ is 1 (see, e.g., Example 2.37). This is the reason why we consider the weaker requirement in Definition 2.13(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 $\alpha$ is called asymptotically correct, and a consistent test having limiting size $\alpha$ is called strongly asymptotically correct.
Definition 2.13

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

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

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

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

Discussion

- If \( \mathcal{P}_0 \) is not a parametric family, it is likely that the limiting size of \( T_n \) is 1 (see, e.g., Example 2.37). This is the reason why we consider the weaker requirement in Definition 2.13(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 \( \alpha \) is called asymptotically correct, and a consistent test having limiting size \( \alpha \) is called strongly asymptotically correct.