# Likelihood Ratio Test of a General Linear Hypothesis

Consider the Likelihood Ratio Test of

$$H_0 : \boldsymbol{C\beta} = \boldsymbol{d} \quad \text{vs} \quad H_A : \boldsymbol{C\beta} \neq \boldsymbol{d}.$$

Suppose

$$\boldsymbol{y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}).$$

The likelihood function is

$$L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{X\beta})'(\boldsymbol{y}-\boldsymbol{X\beta})} \quad \text{for } \boldsymbol{\beta} \in \mathbb{R}^p \text{ and } \sigma^2 > 0.$$

Note that

$$L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) = (2\pi\sigma^2)^{-n/2} e^{-\frac{1}{2\sigma^2} Q(\boldsymbol{\beta})},$$

where

$$\begin{aligned} Q(\boldsymbol{\beta}) &= (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2. \end{aligned}$$

The parameter space under the null hypothesis $H_0 : \boldsymbol{C\beta} = \boldsymbol{d}$ is

$$\Omega_0 = \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{C\beta} = \boldsymbol{d}, \sigma^2 > 0\}.$$

The parameter space corresponding to the union of the null and alternative parameter spaces is

$$\Omega = \{(\boldsymbol{\beta}, \sigma^2) : \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0\}.$$

The likelihood ratio test rejects $H_0$ iff

$$\Lambda(\boldsymbol{y}) = \frac{\sup_{\Omega_0} L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y})}{\sup_{\Omega} L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y})}$$

is sufficiently small.

To conduct a significance level $\alpha$ likelihood ratio test, we reject $H_0$ iff

$$\Lambda(\boldsymbol{y}) \leq c_\alpha,$$

where $c_\alpha$ satisfies

$$\sup\{\mathbb{P}(\Lambda(\boldsymbol{y}) \leq c_\alpha | \boldsymbol{\beta}, \sigma^2) : (\boldsymbol{\beta}, \sigma^2) \in \Omega_0\} \leq \alpha.$$

To find $\Lambda(\boldsymbol{y})$, we must maximize the likelihood over $\Omega_0$ and $\Omega$.

For any fixed $\boldsymbol{\beta} \in \mathbb{R}^p$, we can find the value of $\sigma^2 > 0$ that maximizes $L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y})$ as follows.

Because log is a strictly increasing function, the value of $\sigma^2$ that maximizes $L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y})$ is the same as the value of $\sigma^2$ that maximizes

$$l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) \equiv \log L(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}).$$

$$l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}Q(\boldsymbol{\beta})$$
$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y})}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{Q(\boldsymbol{\beta})}{2\sigma^4}.$$

Equating to 0 and solving for $\sigma^2$ yields

$$\hat{\sigma}^2(\boldsymbol{\beta}) = \frac{Q(\boldsymbol{\beta})}{n}.$$

Furthermore, examination $\frac{\partial l(\boldsymbol{\beta},\sigma^2|\boldsymbol{y})}{\partial \sigma^2}$ as a function of $\sigma^2$ shows that $l(\boldsymbol{\beta},\sigma^2|\boldsymbol{y})$ as a function of $\sigma^2$ is increasing to the left of $Q(\boldsymbol{\beta})/n$ and decreasing to the right of $Q(\boldsymbol{\beta})/n$.

Thus, for any fixed $\boldsymbol{\beta} \in \mathbb{R}^p$, the likelihood is maximized over $\sigma^2 > 0$ at $Q(\boldsymbol{\beta})/n$.

Now note that

$$L(\boldsymbol{\beta}, \hat{\sigma}^2(\boldsymbol{\beta})|\boldsymbol{y}) = (2\pi Q(\boldsymbol{\beta})/n)^{-n/2} e^{-\frac{n}{2Q(\boldsymbol{\beta})}Q(\boldsymbol{\beta})}$$
$$= (2\pi e Q(\boldsymbol{\beta})/n)^{-n/2},$$

which is clearly maximized over $\boldsymbol{\beta}$ by minimizing $Q(\boldsymbol{\beta})$ over $\boldsymbol{\beta}$.

Let $\hat{\boldsymbol{\beta}}$ be any minimizer of $Q(\boldsymbol{\beta})$ over $\boldsymbol{\beta} \in \mathbb{R}^p$.

We know from previous results that $\hat{\boldsymbol{\beta}}$ minimizes $Q(\boldsymbol{\beta})$ over $\boldsymbol{\beta} \in \mathbb{R}^p$ iff $\hat{\boldsymbol{\beta}}$ is a solution to NE. $(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{y}$ is one solution.

Thus, the maximum likelihood estimator (MLE) of $\sigma^2$ is

$$\hat{\sigma}^2_{\text{MLE}} = Q(\hat{\boldsymbol{\beta}})/n = \|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2/n,$$

where $\hat{\boldsymbol{\beta}}$ is any solution to the NE.

Recall that $\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{P_X y}$ is the same for all solution to NE.

Thus, $\hat{\sigma}^2_{\text{MLE}}$ is the same for any $\hat{\boldsymbol{\beta}}$ that solves the NE.

Note that the MLE of $\sigma^2$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{Q(\hat{\boldsymbol{\beta}})}{n} = \frac{Q(\hat{\boldsymbol{\beta}})}{n - rank(\boldsymbol{X})} \frac{n - rank(\boldsymbol{X})}{n}$$
$$= \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}}{n - rank(\boldsymbol{X})} \frac{n - rank(\boldsymbol{X})}{n}$$
$$= \hat{\sigma}^2 \frac{n - rank(\boldsymbol{X})}{n}.$$

Recall

$$E(\hat{\sigma}^2) = \sigma^2.$$

Thus,

$$E(\hat{\sigma}^2_{\text{MLE}}) = \frac{n - rank(\boldsymbol{X})}{n} \sigma^2 < \sigma^2.$$

If $\boldsymbol{X}$ is of full-column rank, then

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

is the maximum likelihood estimator (MLE) of $\boldsymbol{\beta}$.

We have

$$\sup_{\Omega} L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) = L(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2_{\text{MLE}} | \boldsymbol{y})$$

$$= (2\pi e Q(\hat{\boldsymbol{\beta}})/n)^{-n/2}.$$

If we let $\tilde{\boldsymbol{\beta}}$ denote the minimizer of $Q(\boldsymbol{\beta})$ over $\boldsymbol{\beta} \in \mathbb{R}^p$ satisfying $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$, then

$$\begin{aligned}
&\sup_{\Omega_0} L(\boldsymbol{\beta}, \sigma^2 | \boldsymbol{y}) \\
&= L(\tilde{\boldsymbol{\beta}}, Q(\tilde{\boldsymbol{\beta}})/n | \boldsymbol{y}) \\
&= (2\pi e Q(\tilde{\boldsymbol{\beta}})/n)^{-n/2}.
\end{aligned}$$

Thus

$$\Lambda(\boldsymbol{y}) = \frac{(2\pi e Q(\tilde{\boldsymbol{\beta}})/n)^{-n/2}}{(2\pi e Q(\hat{\boldsymbol{\beta}})/n)^{-n/2}} = \left[\frac{Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})}\right]^{-n/2}.$$

Now note that

$$\begin{aligned}
\Lambda(\boldsymbol{y}) \leq c_\alpha \quad &\iff \quad \left[\frac{Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})}\right]^{-n/2} \leq c_\alpha \\
&\iff \quad \frac{Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})} \geq c_\alpha^{-2/n} \\
&\iff \quad \frac{Q(\tilde{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})} - 1 \geq c_\alpha^{-2/n} - 1 \\
&\iff \quad \frac{Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})}{Q(\hat{\boldsymbol{\beta}})} \geq c_\alpha^{-2/n} - 1.
\end{aligned}$$

$$\iff \frac{[Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})]/q}{Q(\hat{\boldsymbol{\beta}})/(n-r)} \geq \frac{n-r}{q}(c_{\alpha}^{-2/n} - 1),$$

where

$$q = rank(\boldsymbol{C}) \quad \text{and} \quad n - r = n - rank(\boldsymbol{X}).$$

# Example:

Suppose

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where

$$\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I}).$$

Furthermore, suppose $rank(\underset{n\times p}{\boldsymbol{X}}) = p$. Partition

$$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2] \quad \text{and} \quad \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix},$$

where $\boldsymbol{X}_1$ is $n \times p_1$ and $\boldsymbol{\beta}_1$ is $p_1 \times 1$.

Suppose we wish to test $H_0 : \boldsymbol{\beta}_2 = \boldsymbol{0}$.

$H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ is a GLH $H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ with

$$\boldsymbol{C} = [\underset{q \times p_1}{\mathbf{0}}, \underset{q \times q}{\boldsymbol{I}}] \quad \text{and} \quad \boldsymbol{d} = \underset{q \times 1}{\mathbf{0}},$$

where

$$q = p - p_1.$$

This GLH is testable because $\underset{q \times p}{\boldsymbol{C}}$ has rank $q$ and $\boldsymbol{C}\boldsymbol{\beta}$ is estimable due to full-column rank of $\boldsymbol{X}$.

$$
\begin{aligned}
Q(\tilde{\boldsymbol{\beta}}) &= \min\{Q(\boldsymbol{\beta}) : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}\} \\
&= \min\{\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 : \boldsymbol{\beta} \in \mathbb{R}^p \ni \boldsymbol{\beta}_2 = \boldsymbol{0}\} \\
&= \min\{\|\boldsymbol{y} - \boldsymbol{X}_1\boldsymbol{\beta}_1\|^2 : \boldsymbol{\beta}_1 \in \mathbb{R}^{p_1}\} \\
&= \|\boldsymbol{y} - \boldsymbol{X}_1\hat{\boldsymbol{\beta}}_1\|^2 = \|\boldsymbol{y} - \boldsymbol{P}_{X_1}\boldsymbol{y}\|^2 \\
&= \|(\boldsymbol{I} - \boldsymbol{P}_{X_1})\boldsymbol{y}\|^2 = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_{X_1})\boldsymbol{y} \\
&= \text{SSE}_{\text{Reduced}}.
\end{aligned}
$$

$$
\begin{aligned}
Q(\hat{\boldsymbol{\beta}}) &= \|\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 \\
&= \|\boldsymbol{y} - \boldsymbol{P_X y}\|^2 \\
&= \|(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y}\|^2 \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P_X})\boldsymbol{y} \\
&= \text{SSE}_{\text{Full}}.
\end{aligned}
$$

The DF for $\mathrm{SSE}_{\mathrm{Full}}$ is

$$DF_F = rank(\boldsymbol{I} - \boldsymbol{P_X}) = n - p.$$

The DF for $\mathrm{SSE}_{\mathrm{Reduced}}$ is

$$DF_R = rank(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_1}) = n - p_1.$$

$$DF_R - DF_F = p - p_1 = q.$$

We have shown that, in the general case, the likelihood ratio test (LRT) of

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$$

rejects for sufficiently large

$$\frac{[Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})]/q}{Q(\hat{\boldsymbol{\beta}})/(n-r)}.$$

In this example,

$$\frac{[Q(\tilde{\beta}) - Q(\hat{\beta})]/q}{Q(\hat{\beta})/(n-r)} = \frac{[\text{SSE}_{\text{Reduced}} - \text{SSE}_{\text{Full}}]/(DF_R - DF_F)}{\text{SSE}_{\text{Full}}/DF_F},$$

which should look familiar.

We now show that for a testable GLH

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d},$$

the GLT statistic

$$\begin{aligned} F &= \frac{(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})/q}{\hat{\sigma}^2} \\ &= \frac{[Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}})]/q}{Q(\hat{\boldsymbol{\beta}})/(n-r)}. \end{aligned}$$

Thus, the GLT is equivalent to the LRT.

Note that

$$\begin{aligned} Q(\hat{\boldsymbol{\beta}})/(n-r) &= \boldsymbol{y}'\left(\frac{\boldsymbol{I}-\boldsymbol{P_X}}{n-r}\right)\boldsymbol{y} \\ &= \hat{\sigma}^2. \end{aligned}$$

Thus, it remains to show

$$Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}}) = (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d}).$$

From 3.10, we know that $\tilde{\boldsymbol{\beta}}$ is leading subvector of solution to RNE.

# Theorem 6.1:

If $\boldsymbol{C}\boldsymbol{\beta} = \boldsymbol{d}$ is testable and $\tilde{\boldsymbol{\beta}}$ is the leading subvecctor of a solution to the RNE

$$\begin{bmatrix} \boldsymbol{X}'\boldsymbol{X} & \boldsymbol{C}' \\ \boldsymbol{C} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}'\boldsymbol{y} \\ \boldsymbol{d} \end{bmatrix},$$

then

$$\begin{aligned} Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}}) &= (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) \\ &= (\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d}). \end{aligned}$$

# Proof of Result 6.1:

First show that

$$Q(\tilde{\boldsymbol{\beta}}) - Q(\hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}).$$

Now let $\tilde{\boldsymbol{\lambda}}$ denote the trailing subvector of the solution to RNE whose leading subvector is $\tilde{\boldsymbol{\beta}}$; i.e., suppose

$$\begin{bmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\lambda}} \end{bmatrix}$$

is solution to RNE.

Show $\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \boldsymbol{C}'\tilde{\boldsymbol{\lambda}}$.

Now show

$$\tilde{\boldsymbol{\lambda}} = (\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d}).$$

We have established

(1) $Q(\hat{\boldsymbol{\beta}}) - Q(\tilde{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}})$

(2) $\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}) = \boldsymbol{C}'\tilde{\boldsymbol{\lambda}}$

(3) $\tilde{\boldsymbol{\lambda}} = (\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}\hat{\boldsymbol{\beta}} - \boldsymbol{d})$.

Use these results to finish the proof.

# Corollary 6.4:

Suppose $\boldsymbol{C\beta} = \boldsymbol{d}$ is testable, and suppose $\hat{\boldsymbol{\beta}}$ is a solution to the NE.

Then the leading subvector of a solution to the RNE with constraint

$$\boldsymbol{C\beta} = \boldsymbol{d}$$

can be found by solving for $\boldsymbol{b}$ in the equations

$$\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y} - \boldsymbol{C}'(\boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}')^{-1}(\boldsymbol{C}'\hat{\boldsymbol{\beta}} - \boldsymbol{d}).$$