Hypothesis Testing in a Likelihood Framework

I Testing Simple hypotheses

Let $X_1, \ldots, X_n$ be iid with pdf $f(x_i; \theta)$

and assume θ is a scalar.

The hypothesis to be tested is simple & two sided:

$$H_0: \theta = \theta_0 \quad \text{vs.} \quad H_a: \theta \neq \theta_0$$

Statistical Test: Decision rule based on data to either reject H_0 or not reject H_0.

Likelihood Ratio Statistic

Consider the likelihood ratio

$$\lambda = \frac{L(\theta_0; \underset{\sim}{x})}{L(\hat{\theta}_{MLE}; \underset{\sim}{x})}$$

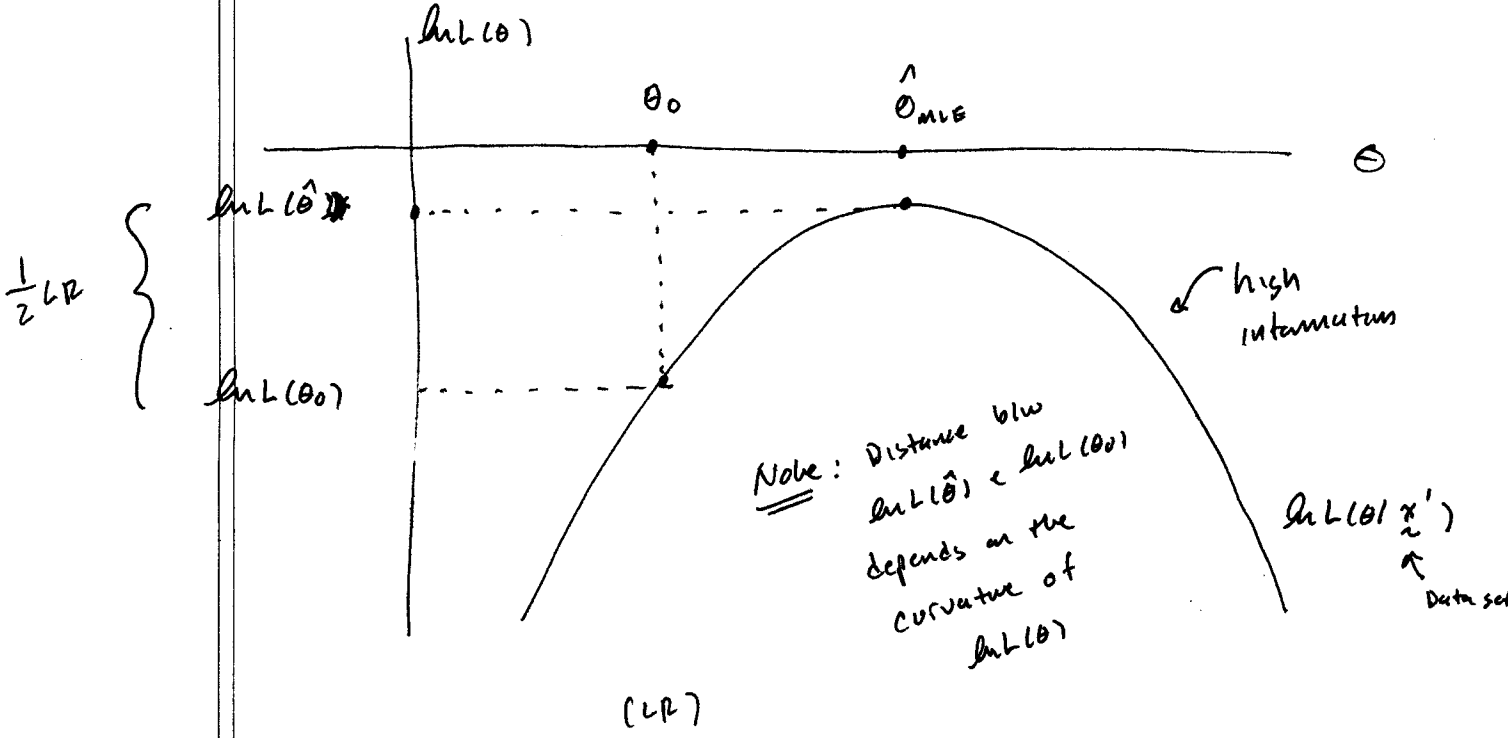
$$\left(= \frac{\max_{\text{s.t. } \theta = \theta_0} L(\theta; \underset{\sim}{x})}{\max_{\theta \in \Omega \text{ unrestricted}} L(\theta; \underset{\sim}{x})} \right.$$

= ratio of "restricted" to unrestricted likelihoods.

By construction $0 < \lambda \leq 1$.

If $H_0: \theta = \theta_0$ is true then $\lambda \approx 1$ and

if $H_0: \theta \neq \theta_0$ is not true then $\lambda < 1$



The LR statistic is a simple transformation of λ

s.t. LR is large if $H_0: \theta = \theta_0$ is not true and

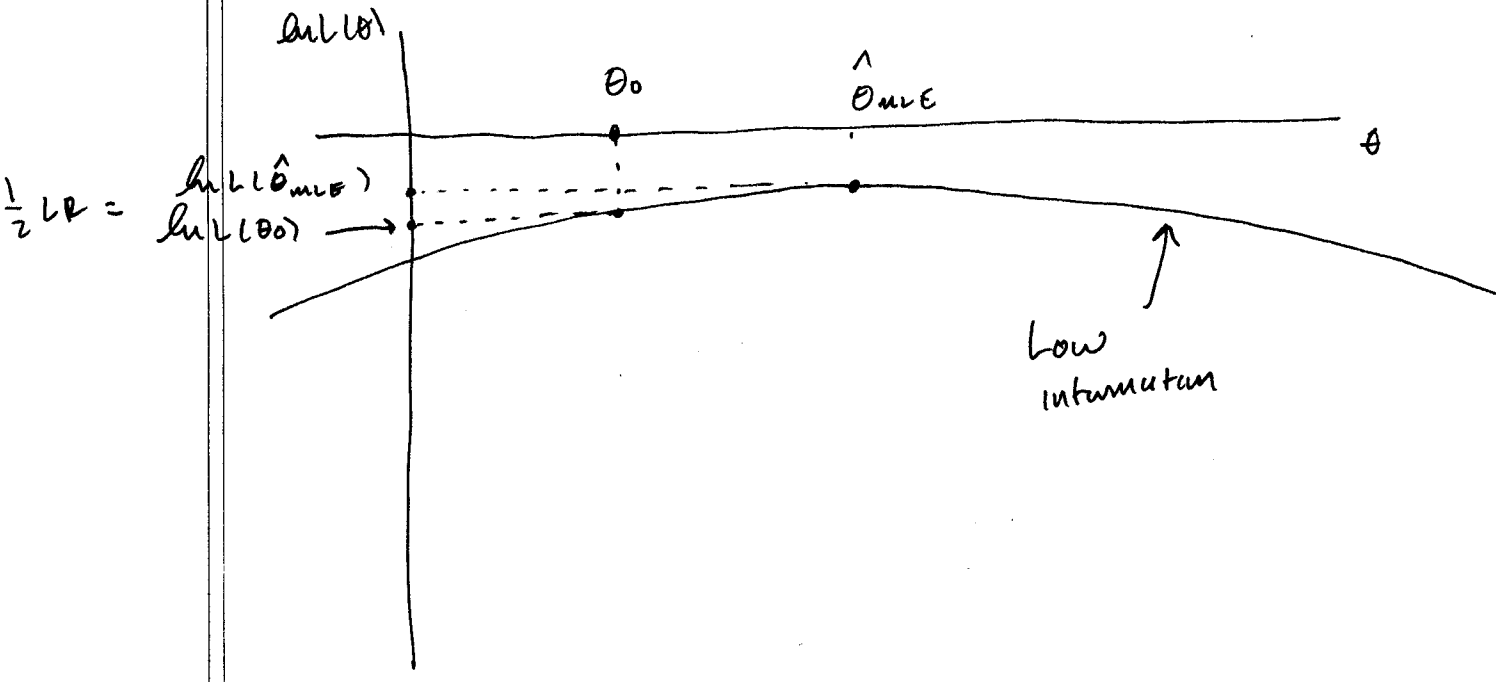
is small if $H_0: \theta = \theta_0$ is true:

$$LR = -2 \ln \lambda$$

$$= -2 \left[\ln L(\theta_0 : \underset{\sim}{x}) - \ln L(\hat{\theta}_{MLE} ; \underset{\sim}{x}) \right]$$

From the diagram, notice that the distance between $\ln L(\hat{\theta}_{MLE})$ and $\ln L(\theta_0)$ depends on the curvature of $\ln L(\theta)$ near $\hat{\theta}_{MLE}$. If the curvature is sharp (i.e. high information) then LR will be large for θ_0 values away from $\hat{\theta}_{MLE}$

If, however, the curvature of $\ln L(\theta)$ is flat (i.e. low information) as depicted below



then LR will be small for θ_0 values away from $\hat{\theta}_{MLE}$.

Result (no proof)

Under very general regularity conditions, if $H_0: \theta = \theta_0$ is true then

$$LR = -2\ln\Lambda \xrightarrow{d} \chi^2(1)$$

χ^2 random variable with 1 d.f.

Note: The # of d.f. is based on the # of restrictions being tested.

We reject $H_0: \theta = \theta_0$ at, say, 5% level if $LR > \chi^2_{0.95}(1)$ = 95% quantile of $\chi^2(1)$ distribution

Wald Statistic

The Wald statistic is based directly on the asymptotic distribution of $\hat{\theta}_{MLE}$:

$$\hat{\theta}_{MLE} \overset{A}{\sim} N\left(\theta, I(\theta : \underset{\sim}{x})^{-1}\right)$$

Sample information matrix

Where $I(\theta; \underset{\sim}{x})$ is estimated using $\hat{\theta}_{MLE}$:

$$\hat{Var}(\hat{\theta}_{MLE}) = I(\hat{\theta}_{MLE}; \underset{\sim}{x})^{-1} = \text{estimated asymptotic variance of } \hat{\theta}_{MLE}$$

Given these results, we know that the t-ratio

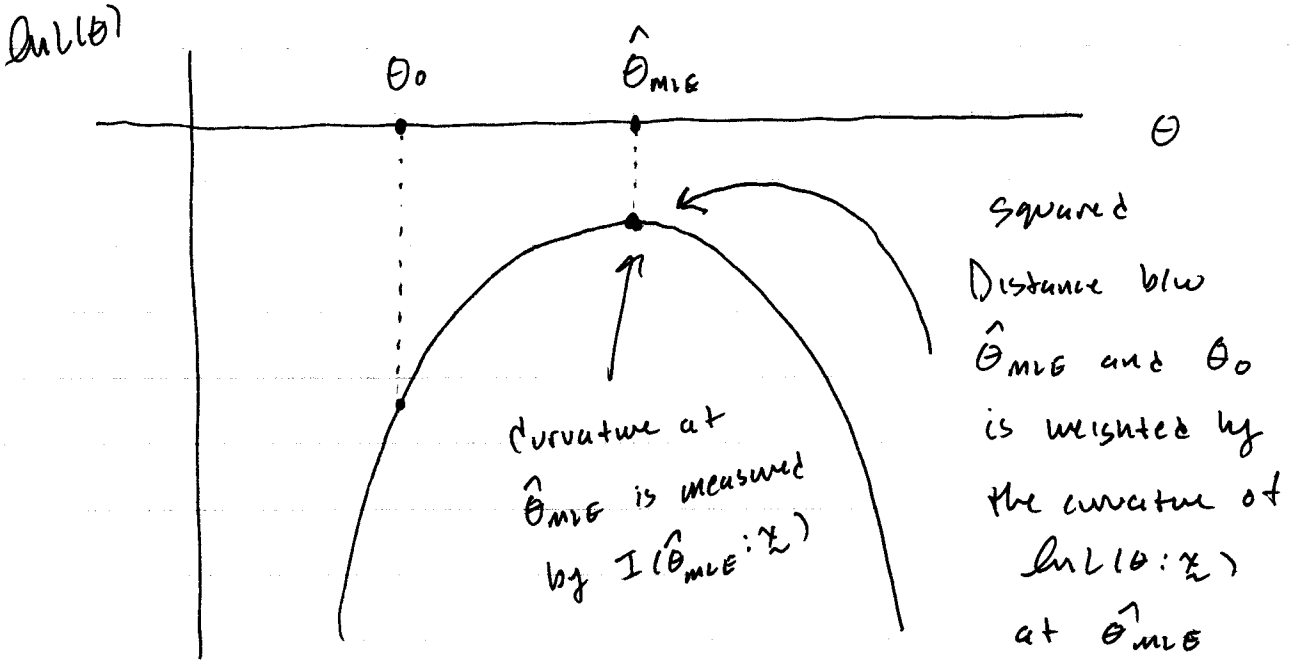
$$\frac{\hat{\theta}_{MLE} - \theta_0}{\sqrt{\hat{Var}(\hat{\theta}_{MLE})}} = (\hat{\theta}_{MLE} - \theta_0)\sqrt{I(\hat{\theta}_{MLE}; \underset{\sim}{x})}$$

is asymptotically $N(0,1)$ assuming $H_0: \theta = \theta_0$ is true. Using the CMT it follows that the square of the t-ratio is asymptotically $\chi^2(1)$. The Wald statistic is defined to be simply the square of this t-ratio:

$$\text{Wald} = \frac{(\hat{\theta}_{MLE} - \theta_0)^2}{\hat{Var}(\hat{\theta}_{MLE})} = (\hat{\theta}_{MLE} - \theta_0)^2 \cdot I(\hat{\theta}_{MLE}; \underset{\sim}{x})$$

Result: Under H_0, $\text{Wald} \xrightarrow{d} \chi^2(1)$

The intuition behind the Wald statistic is illustrated in the following figure



If the curvature at $\hat{\theta}_{MLE}$ is sharp then $I(\hat{\theta}_{MLE} : \underset{\sim}{x})$ is big and the squared distance $(\hat{\theta}_{MLE} - \theta_0)^2$ gets blown up when constructing the Wald statistic.

If the curvature at $\hat{\theta}_{MLE}$ is Low then $I(\hat{\theta}_{MLE} | \underset{\sim}{x})$ is small and the squared distance $(\hat{\theta}_{MLE} - \theta_0)^2$ does not get blown up when constructing the Wald statistic.

LM Statistic

With MLE, $\hat{\theta}_{MLE}$ solves the first order conditions

$$0 = \frac{\partial \ln L(\hat{\theta}_{MLE} | \underset{\sim}{x})}{\partial \theta} = S(\hat{\theta}_{MLE} ; \underset{\sim}{x})$$

If $H_0 : \theta = \theta_0$ is true then we should expect that

$$0 \approx \frac{\partial \ln L(\theta_0 : \underset{\sim}{x})}{\partial \theta} = S(\theta_0 : \underset{\sim}{x})$$

and if $H_0 : \theta = \theta_0$ is not true then

$$0 \neq \frac{\partial \ln L(\theta_0 : \underset{\sim}{x})}{\partial \theta} = S(\theta_0 : \underset{\sim}{x})$$

The LM test is based on how far $S(\theta_0 : \underset{\sim}{x})$ is from zero.

Recall the following properties of $S(\theta : \underset{\sim}{x})$. If $H_0 : \theta = \theta_0$ is true then

(i) $E\left[S(\theta_0 : \underset{\sim}{x}) \right] = 0$

(ii) $\text{var}\left(S(\theta_0 : \underset{\sim}{x}) \right) = I(\theta_0 : \underset{\sim}{x})$

Further, it can be shown that

$$S(\theta_0 : \underset{\sim}{x}) \overset{A}{\sim} N\left(0, I(\theta_0 : \underset{\sim}{x})\right)$$

The LM statistic for testing $H_0 : \theta = \theta_0$ is then

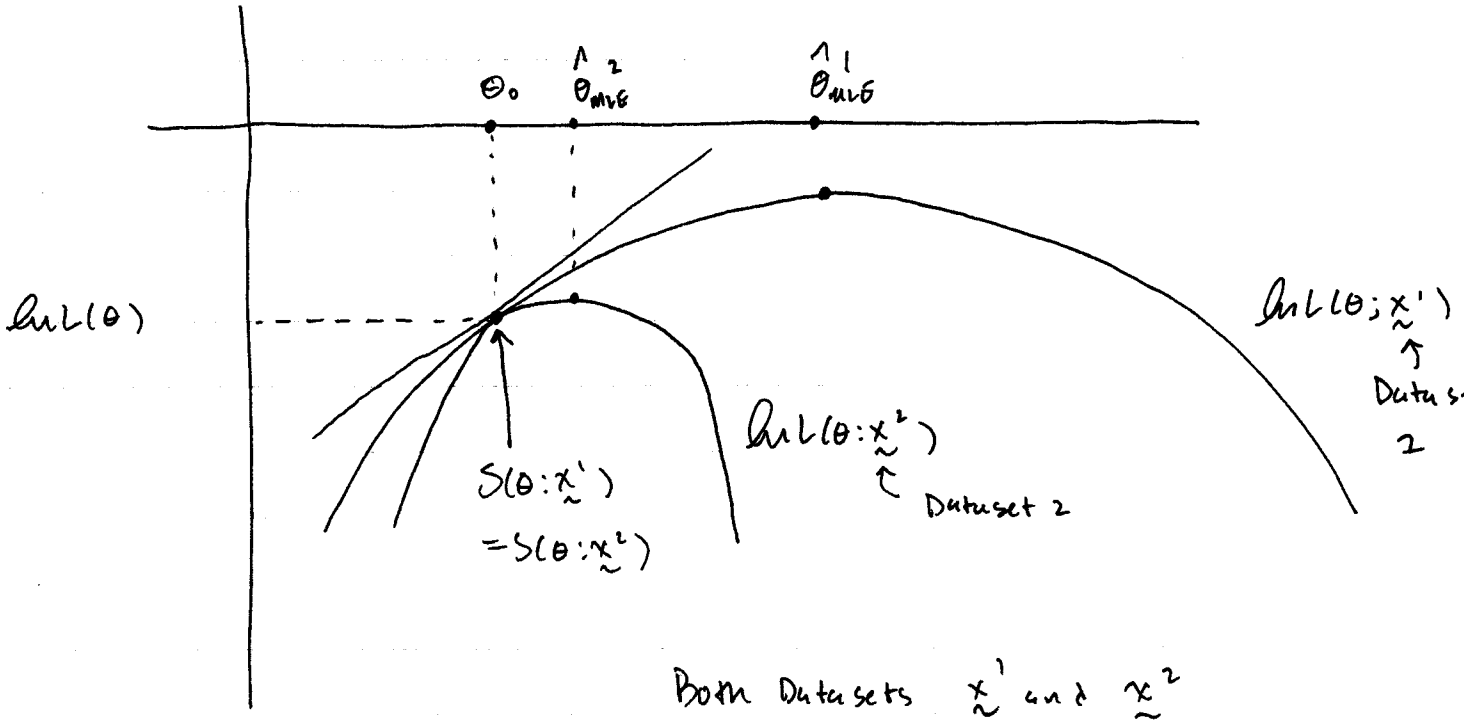
$$LM = \frac{S(\theta_0 : \underset{\sim}{x})^2}{I(\theta_0 : \underset{\sim}{x})} = S(\theta_0 : \underset{\sim}{x})^2 \cdot I(\theta_0 : \underset{\sim}{x})^{-1}$$

Note: The LM statistic is based on the "restricted" estimate of θ

Result: Under $H_0 : \theta = \theta_0$

$$LM \xrightarrow{d} \chi^2(1)$$

The basic intuition behind the LM statistic is given in the following diagram:



Both Datasets $\underset{\sim}{x}^1$ and $\underset{\sim}{x}^2$ have the same Score values at θ_0:

$$S(\theta_0; \underset{\sim}{x}^1) = S(\theta_0; \underset{\sim}{x}^2)$$

$$LM^1 = S(\theta_0; \underset{\sim}{x}^1)^2 * I(\theta_0; \underset{\sim}{x}^1)^{-1}$$

$$LM^2 = S(\theta_0; \underset{\sim}{x}^2)^2 \cdot I(\theta_0; \underset{\sim}{x}^2)^{-1}$$

Here $LM^2 < LM^1$ since $I(\theta_0; \underset{\sim}{x}^2)^{-1} > I(\theta_0; \underset{\sim}{x}^1)^{-1}$

i.e. the Dataset that produces the largest curvature at θ_0 indicates that θ_0 is closer to $\hat{\theta}_{MLE}$.

Test statistics for General Nonlinear hypotheses

Let $X_1, \ldots, X_n$ be iid with pdf $f(x_i; \underset{\sim}{\theta})$

where $\underset{\sim}{\theta} \in \mathbb{R}^K$. Consider testing the general non-linear hypothesis

$$H_0: \underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0} \quad \text{vs.} \quad \underset{\sim}{g}(\underset{\sim}{\theta}) \neq \underset{\sim}{0}$$

$\underset{\sim}{g}$: $J \times 1$

Here

$$\underset{\sim}{g}(\underset{\sim}{\theta}) = \begin{pmatrix} g_1(\underset{\sim}{\theta}) \\ g_2(\underset{\sim}{\theta}) \\ \vdots \\ g_J(\underset{\sim}{\theta}) \end{pmatrix}$$

scalar functions = continuous and 2 times differentiable

and $\underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0}$ places $J \leq K$ total restrictions on $\underset{\sim}{\theta}$

Example $X \sim N(\mu, \sigma^2)$, $\underset{\sim}{\theta} = (\mu, \sigma^2)'$

$H_0: \mu = \mu_0$ vs. $H_1: \mu \neq \mu_0$

$\Leftrightarrow$ $H_0: g(\underset{\sim}{\theta}) = 0$ vs. $H_1: g(\underset{\sim}{\theta}) \neq 0$

with $g(\underset{\sim}{\theta}) = \mu - \mu_0$.

Define

$$G(\underset{\sim}{\theta}) = \underset{J \times K}{\frac{\partial g(\underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}} =$$

$$= \begin{bmatrix} \frac{\partial g_1(\theta)}{\partial \theta_1} & \frac{\partial g_1(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_1(\theta)}{\partial \theta_K} \\ \vdots & & & \\ \frac{\partial g_J(\theta)}{\partial \theta_1} & \frac{\partial g_J(\theta)}{\partial \theta_2} & \cdots & \frac{\partial g_J(\theta)}{\partial \theta_K} \end{bmatrix}$$

and assume $\text{rank}(G) = J < K$.

LR Test

$$\lambda = \frac{\max_{\theta} L(\theta; \underset{\sim}{x}) \quad \text{s.t.} \quad \underset{\sim}{g}(\underset{\sim}{\theta}) = 0}{\max_{\theta} L(\theta : \underset{\sim}{x})}$$

$$= \frac{L(\tilde{\theta}_{MLE}; \underset{\sim}{x})}{L(\hat{\theta}_{MLE}; \underset{\sim}{x})}$$

Requires both restricted & unrestricted estimates!

where $\tilde{\theta}_{MLE}$ = restricted MLE computed subject to $\underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0}$; i.e.

$$g(\tilde{\theta}_{MLE}) = 0$$

and $\hat{\theta}_{MLE}$ = unrestricted MLE. In general, $\underset{\sim}{g}(\hat{\theta}_{MLE}) \neq \underset{\sim}{0}$

Then

$$LR = -2 \ln \lambda = -2 \left[\ln L(\tilde{\theta}_{MLE}; \underset{\sim}{x}) - \ln L(\hat{\theta}_{MLE}; \underset{\sim}{x} \right.$$

and under H_0

$$LR \xrightarrow{d} \chi^2(J)$$

where J = # of restrictions under H_0.

LM Test

Idea: If $H_0: \underset{\sim}{g}(\underset{\sim}{\theta}) = 0$ is true then

$$S(\tilde{\theta}_{MLE}; \underset{\sim}{x}) \approx 0$$

and

$I(\tilde{\theta}_{MLE}; \underset{\sim}{x})$ consistently estimates

$\text{var}\left(S(\tilde{\theta}_{MLE}; \underset{\sim}{x})\right)$

Then

$$LM = \underset{1\times K}{S(\tilde{\theta}_{MLE}; \underset{\sim}{x})'} \underset{K\times K}{I(\tilde{\theta}_{MLE}; \underset{\sim}{x})^{-1}} \underset{K\times 1}{S(\tilde{\theta}_{MLE}; \underset{\sim}{x})}$$

and under H_0

$$LM \xrightarrow{d} \chi^2(J)$$

Wald Test

Here the Wald test is based on the asymptotic normality of $\hat{\theta}_{MLE}$ and the "delta method":

$$\underset{\sim}{\hat{\theta}}_{MLE} \overset{A}{\sim} N\left(\theta, I(\underset{\sim}{\hat{\theta}}_{MLE}; \underset{\sim}{x})^{-1}\right)$$

$$\underset{J\times 1}{\underset{\sim}{g}(\underset{\sim}{\hat{\theta}}_{MLE})} \overset{A}{\sim} N\left(\underset{\sim}{g}(\underset{\sim}{\theta}), \underbrace{\underset{J\times K}{G(\underset{\sim}{\hat{\theta}}_{MLE})} \underset{K\times K}{I(\underset{\sim}{\hat{\theta}}_{MLE}; \underset{\sim}{x})^{-1}} \underset{K\times J}{G(\underset{\sim}{\hat{\theta}}_{MLE})'}}_{J\times J}\right)$$

$$Wald = \underset{1\times J}{\underset{\sim}{g}(\underset{\sim}{\hat{\theta}}_{MLE})'} \left[G(\underset{\sim}{\hat{\theta}}_{MLE}) I(\underset{\sim}{\hat{\theta}}_{MLE}; \underset{\sim}{x})^{-1} G(\underset{\sim}{\hat{\theta}}_{MLE})'\right]^{-1} \underset{\sim}{g}(\underset{\sim}{\hat{\theta}}_{MLE})$$

$$\xrightarrow{d} \chi^2(J) \quad \text{under } H_0 : \underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{0}.$$